



Configurazione dell'host

Enterprise applications

NetApp
May 09, 2024

Sommario

- Configurazione dell'host 1
 - Database Oracle con IBM AIX 1
 - Database Oracle con HP-UX 2
 - Database Oracle con Linux 4
 - Database Oracle con ASMSLib/AFD (driver filtro ASM). 8
 - Database Oracle con Microsoft Windows 10
 - Database Oracle con Solaris 10

Configurazione dell'host

Database Oracle con IBM AIX

Argomenti di configurazione per database Oracle su IBM AIX con ONTAP.

I/o simultanei

Per ottenere prestazioni ottimali su IBM AIX è necessario utilizzare l'i/o simultaneo. Senza i/o simultaneo, è probabile che le limitazioni delle prestazioni siano dovute al fatto che AIX esegue i/o atomico serializzato, che comporta un overhead significativo.

In origine, NetApp ha consigliato di utilizzare `cio` Opzione di montaggio per forzare l'uso di i/o simultanei sul file system, ma questo processo ha avuto degli inconvenienti e non è più necessario. Dall'introduzione di AIX 5.2 e Oracle 10gR1, Oracle su AIX può aprire singoli file per i/o simultanei, anziché forzare i/o simultanei sull'intero file system.

Il metodo migliore per abilitare l'i/o simultaneo è impostare `init.ora` parametro `filesystemio_options` a `setall`. In questo modo, Oracle può aprire file specifici da utilizzare con i/o simultanei

Utilizzo di `cio` Come opzione di montaggio, l'utilizzo di i/o simultanei può avere conseguenze negative. Ad esempio, forzando i/o simultanei si disabilita la lettura dei file system, che può danneggiare le prestazioni dell'i/o al di fuori del software del database Oracle, come la copia dei file e l'esecuzione di backup su nastro. Inoltre, prodotti come Oracle GoldenGate e SAP BR*Tools non sono compatibili con l'uso di `cio`. Montare l'opzione con alcune versioni di Oracle.

NetApp consiglia quanto segue:



- Non utilizzare `cio` opzione di montaggio a livello di file system. Abilitare invece l'i/o simultaneo tramite l'utilizzo di `filesystemio_options=setall`.
- Utilizzare solo l' `cio` l'opzione di montaggio dovrebbe essere impostata se non è possibile `filesystemio_options=setall`.

Opzioni di montaggio NFS AIX

Nella tabella seguente sono elencate le opzioni di montaggio NFS AIX per i database Oracle a istanza singola.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144</code>
File di controllo File di dati Registri di ripristino	<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>

Nella tabella seguente sono elencate le opzioni di montaggio NFS AIX per RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
File di controllo File di dati Registri di ripristino	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac</code>
CRS/Voting	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac</code>
Dedicato ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Condiviso ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr</code>

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `noac` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare la cache del sistema operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati.

Anche se si utilizza il `cio` montare l'opzione `e. init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `noac`. `noac` è obbligatorio per condiviso ORACLE_HOME Implementazioni per facilitare la coerenza di file quali file di password Oracle e `spfile` file di parametri. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata ORACLE_HOME, questo parametro non è necessario.

Opzioni di montaggio di AIX jfs/JFS2

Nella tabella seguente sono elencate le opzioni di montaggio di AIX jfs/JFS2.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	Valori predefiniti
File di controllo File di dati Registri di ripristino	Valori predefiniti
ORACLE_HOME	Valori predefiniti

Prima di utilizzare AIX `hdisk` i dispositivi in qualsiasi ambiente, inclusi i database, controllano il parametro `queue_depth`. Questo parametro non è la profondità della coda HBA, bensì la profondità della coda SCSI dell'individuo `hdisk` device. Depending on how the LUNs are configured, the value for `queue_depth` potrebbe essere troppo basso per garantire buone prestazioni. I test hanno dimostrato che il valore ottimale è 64.

Database Oracle con HP-UX

Argomenti di configurazione per database Oracle su HP-UX con ONTAP.

Opzioni di montaggio NFS HP-UX

Nella tabella seguente sono elencate le opzioni di montaggio NFS HP-UX per una singola istanza.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>
File di controllo File di dati Registri di ripristino	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,forcedirectio, nointr,suid</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>

Nella tabella seguente sono elencate le opzioni di montaggio NFS HP-UX per RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,noac,suid</code>
File di controllo File di dati Registri di ripristino	<code>rw, bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,forcedirectio,suid</code>
CRS/votazione	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac, forcedirectio,suid</code>
Dedicato ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>
Condiviso ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,suid</code>

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `noac` e `forcedirectio` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare il caching del sistema operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati. Anche se si utilizza il `init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `noac` e `forcedirectio`.

Il motivo `noac` è obbligatorio per condiviso ORACLE_HOME Le distribuzioni consentono di semplificare la coerenza di file quali file di password Oracle e file `sfile`. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata ORACLE_HOME, questo parametro non è richiesto.

Opzioni di montaggio VxFS HP-UX

Utilizzare le seguenti opzioni di montaggio per i file system che ospitano file binari Oracle:

```
delaylog,nodatainlog
```

Utilizzare le seguenti opzioni di montaggio per i file system contenenti file di dati, log di ripristino, log di archivio e file di controllo in cui la versione di HP-UX non supporta i/o simultanei:

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

Quando l'i/o simultaneo è supportato (VxFS 5.0.1 e versioni successive o con ServiceGuard Storage Management Suite), utilizzare queste opzioni di montaggio per i file system contenenti file di dati, log di ripristino, log di archivio e file di controllo:

```
delaylog,cio
```



Il parametro `db_file_multiblock_read_count` È particolarmente critico negli ambienti VxFS. Oracle consiglia di non impostare questo parametro in Oracle 10g R1 e versioni successive, a meno che non sia diversamente specificato. L'impostazione predefinita con dimensioni blocco Oracle 8KB è 128 KB. Se il valore di questo parametro è forzato a 16 o inferiore, rimuovere l' `convosync=direct` Montare l'opzione perché può danneggiare le prestazioni i/o sequenziali. Questa operazione danneggia altri aspetti delle prestazioni e deve essere eseguita solo se il valore di `db_file_multiblock_read_count` deve essere modificato dal valore predefinito.

Database Oracle con Linux

Argomenti di configurazione specifici del sistema operativo Linux.

Tablelle degli slot TCP per Linux NFSv3

Le tabelle degli slot TCP sono l'equivalente di NFSv3 della profondità della coda degli HBA (host Bus Adapter). Queste tabelle controllano il numero di operazioni NFS che possono essere in sospeso in qualsiasi momento. Il valore predefinito è di solito 16, che è troppo basso per ottenere prestazioni ottimali. Il problema opposto si verifica sui kernel Linux più recenti, che possono aumentare automaticamente il limite della tabella degli slot TCP a un livello che satura il server NFS con le richieste.

Per prestazioni ottimali e per evitare problemi di prestazioni, regolare i parametri del kernel che controllano le tabelle degli slot TCP.

Eseguire `sysctl -a | grep tcp.*.slot_table` e osservare i seguenti parametri:

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

Tutti i sistemi Linux dovrebbero includere `sunrpc.tcp_slot_table_entries`, ma solo alcuni includono `sunrpc.tcp_max_slot_table_entries`. Entrambi devono essere impostati su 128.

Attenzione

La mancata impostazione di questi parametri può avere effetti significativi sulle prestazioni. In alcuni casi, le prestazioni sono limitate poiché il sistema operativo linux non fornisce i/o sufficienti. In altri casi, le latenze i/o aumentano quando il sistema operativo linux tenta di emettere più i/o di quanto possa essere gestito.

Opzioni di montaggio NFS Linux

Nella tabella seguente sono elencate le opzioni di montaggio NFS Linux per una singola istanza.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
File di controllo File di dati Registri di ripristino	<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>

Nella tabella seguente sono elencate le opzioni di montaggio NFS Linux per RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,actimeo=0</code>
File di controllo File di dati Registri di ripristino	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,actimeo=0</code>
CRS/votazione	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,actimeo=0</code>
Dedicato ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Condiviso ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,actimeo=0</code>

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `actimeo=0` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare il caching del sistema operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati. Anche se si utilizza il `init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `actimeo=0`.

Il motivo `actimeo=0` è obbligatorio per condiviso `ORACLE_HOME` Le distribuzioni consentono di semplificare la coerenza di file quali file di password e file `spfile` di Oracle. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata `ORACLE_HOME`, questo parametro non è necessario.

In genere, i file non di database devono essere montati con le stesse opzioni utilizzate per i file di dati a singola istanza, sebbene applicazioni specifiche possano avere requisiti diversi. Evitare le opzioni di montaggio `noac` e `actimeo=0` se possibile perché queste opzioni disabilitano la lettura e il buffering a livello di file system. Ciò può causare gravi problemi di prestazioni per processi quali l'estrazione, la traduzione e il caricamento.

ACCESSO e GETATTR

Alcuni clienti hanno notato che un livello estremamente elevato di altri IOPS, come ACCESSO e GETATTR, può dominare i propri workload. In casi estremi, operazioni come letture e scritture possono arrivare fino al 10% del totale. Si tratta di un comportamento normale con qualsiasi database che include l'uso di `actimeo=0` e/o `noac` Su Linux perché queste opzioni fanno sì che il sistema operativo Linux ricarichi costantemente i metadati dei file dal sistema di archiviazione. Operazioni quali ACCESSO e GETATTR sono operazioni a basso impatto gestite dalla cache ONTAP in un ambiente di database. Non dovrebbero essere considerati IOPS autentici, come le letture e le scritture, che creano una vera domanda sui sistemi storage. Tuttavia, questi altri IOPS creano un certo carico, specialmente negli ambienti RAC. Per risolvere questo problema, abilitare DNFS, che ignora la cache buffer del sistema operativo ed evita queste operazioni non necessarie relative ai metadati.

Linux Direct NFS

Un'opzione di montaggio aggiuntiva, denominata `nosharecache`, È necessario quando (a) DNFS è abilitato e (b) un volume sorgente è montato più di una volta su un singolo server (c) con un mount NFS nidificato. Questa configurazione si osserva principalmente in ambienti che supportano applicazioni SAP. Ad esempio, un singolo volume di un sistema NetApp può avere una directory situata in `/vol/oracle/base` e un secondo a `/vol/oracle/home`. Se `/vol/oracle/base` è montato su `/oracle` e `/vol/oracle/home` è montato su `/oracle/home`, Il risultato sono montaggi NFS nidificati che hanno origine sulla stessa fonte.

Il sistema operativo è in grado di rilevare il fatto che `/oracle` e `/oracle/home` risiedono sullo stesso volume, che è lo stesso file system di origine. Il sistema operativo utilizza quindi lo stesso handle di dispositivo per l'accesso ai dati. In questo modo si migliora l'uso della cache del sistema operativo e di alcune altre operazioni, ma interferisce con DNFS. Se DNFS deve accedere a un file, ad esempio `spfile`, attivato `/oracle/home`, potrebbe erroneamente tentare di utilizzare il percorso errato per i dati. Il risultato è un'operazione i/o non riuscita. In queste configurazioni, aggiungere `nosharecache` Opzione di montaggio su qualsiasi file system NFS che condivide un volume FlexVol di origine con un altro file system NFS su quell'host. In questo modo, il sistema operativo Linux assegna un handle di dispositivo indipendente al file system.

Linux Direct NFS e Oracle RAC

L'uso di DNFS offre speciali vantaggi in termini di prestazioni per Oracle RAC sul sistema operativo Linux, poiché Linux non dispone di un metodo per forzare l'i/o diretto, necessario con RAC per la coerenza tra i nodi. Come soluzione, Linux richiede l'uso di `actimeo=0` Opzione di montaggio, che fa sì che i dati dei file scadano immediatamente dalla cache del sistema operativo. Questa opzione a sua volta obbliga il client NFS Linux a rileggere costantemente i dati degli attributi, danneggiando la latenza e aumentando il carico sullo storage

controller.

Abilitando DNFS si ignora il client NFS dell'host ed evita questo danno. Diversi clienti hanno segnalato significativi miglioramenti delle performance sui cluster RAC e una significativa riduzione del carico ONTAP (soprattutto in relazione ad altri IOPS) quando si attiva DNFS.

Linux Direct NFS e file orafstab

Quando si utilizza DNFS su Linux con l'opzione multipathing, è necessario utilizzare più sottoreti. Su altri sistemi operativi, è possibile stabilire più canali DNFS utilizzando LOCAL e. DONTRROUTE Opzioni per configurare più canali DNFS su una singola subnet. Tuttavia, questo non funziona correttamente su Linux e possono verificarsi problemi di prestazioni imprevisti. Con Linux, ogni NIC utilizzata per il traffico DNFS deve trovarsi su una subnet diversa.

Utilità di pianificazione i/O.

Il kernel Linux permette un controllo di basso livello sul modo in cui l'i/o blocca i dispositivi è programmato. Le impostazioni predefinite su varie distribuzioni di Linux variano notevolmente. I test dimostrano che la scadenza di solito offre i migliori risultati, ma a volte NOOP è stato leggermente migliore. La differenza di prestazioni è minima, ma è necessario verificare entrambe le opzioni se è necessario estrarre le massime prestazioni possibili da una configurazione di database. CFQ è l'impostazione predefinita in molte configurazioni e ha dimostrato di avere problemi significativi di prestazioni con i carichi di lavoro del database.

Per istruzioni sulla configurazione dello scheduler i/o, consultare la documentazione del fornitore di Linux pertinente.

Multipathing

Alcuni clienti hanno riscontrato arresti anomali durante l'interruzione della rete perché il daemon multipath non era in esecuzione sul proprio sistema. Nelle versioni recenti di Linux, il processo di installazione del sistema operativo e del demone multipathing potrebbero lasciare questi sistemi operativi vulnerabili a questo problema. I pacchetti sono installati correttamente, ma non sono configurati per l'avvio automatico dopo un riavvio.

Ad esempio, il valore predefinito per il daemon multipath su RHEL5,5 potrebbe essere il seguente:

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

Questo può essere corretto con i seguenti comandi:

```
[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
```

Mirroring ASM

Il mirroring ASM potrebbe richiedere modifiche alle impostazioni di multipath Linux per consentire ad ASM di riconoscere un problema e passare a un gruppo di errori alternativo. La maggior parte delle configurazioni ASM su ONTAP utilizza la ridondanza esterna, il che significa che la protezione dei dati è fornita dall'array esterno e ASM non esegue il mirroring dei dati. Alcuni siti utilizzano ASM con ridondanza normale per fornire il

mirroring bidirezionale, in genere su siti diversi.

Le impostazioni di Linux visualizzate nella "[Documentazione delle utilità host NetApp](#)" Includi parametri multipath che determinano indefinite code di i/O. Ciò significa che un i/o su un dispositivo LUN senza percorsi attivi attende finché l'i/o non viene completato. Questo è solitamente consigliabile perché gli host Linux attendono il tempo necessario per il completamento delle modifiche al percorso SAN, per il riavvio degli switch FC o per il completamento di un failover da parte di un sistema di storage.

Questo comportamento di accodamento illimitato causa un problema con il mirroring ASM perché ASM deve ricevere un errore di i/o per consentire al reparto IT di riprovare l'i/o su un LUN alternativo.

Impostare i seguenti parametri in Linux `multipath.conf` File per i LUN ASM utilizzati con il mirroring ASM:

```
polling_interval 5
no_path_retry 24
```

Queste impostazioni creano un timeout di 120 secondi per i dispositivi ASM. Il timeout viene calcolato come `polling_interval * no_path_retry` in pochi secondi. In alcuni casi potrebbe essere necessario regolare il valore esatto, ma per la maggior parte degli utilizzi dovrebbe essere sufficiente un timeout di 120 secondi. In particolare, 120 secondi devono consentire il takeover o il giveback del controller senza produrre un errore di i/o che porterebbe il gruppo guasto a diventare offline.

Un più basso `no_path_retry` Il valore può ridurre il tempo richiesto per ASM per passare a un gruppo di errori alternativo, ma aumenta anche il rischio di un failover indesiderato durante attività di manutenzione come il takeover di un controller. Il rischio può essere mitigato tramite un attento monitoraggio dello stato di mirroring ASM. Se si verifica un failover indesiderato, è possibile risincronizzare rapidamente i mirror se la risincronizzazione viene eseguita in modo relativamente rapido. Per ulteriori informazioni, consultare la documentazione Oracle su ASM Fast Mirror Resync per la versione del software Oracle in uso.

Linux xfs, ext3, e ext4 opzioni di mount



NetApp recommended usando le opzioni di mount predefinite.

Database Oracle con ASMSLib/AFD (driver filtro ASM)

Argomenti di configurazione specifici per il sistema operativo Linux utilizzando AFD e ASMLib

Dimensioni dei blocchi ASMLib

ASMLib è una libreria di gestione ASM opzionale e le utilità associate. Il suo valore principale è la capacità di contrassegnare un LUN o un file basato su NFS come una risorsa ASM con un'etichetta leggibile da un utente.

Le versioni recenti di ASMLib rilevano un parametro LUN chiamato Logical Blocks per Physical Block Exponent (LBPPBE). Questo valore non è stato segnalato dal target SCSI ONTAP fino a poco tempo fa. Ora restituisce un valore che indica che è preferibile una dimensione blocco 4KB. Questa non è una definizione della dimensione del blocco, ma è un suggerimento per qualsiasi applicazione che utilizza LBPPBE che i/o di una certa dimensione potrebbero essere gestiti in modo più efficiente. ASMLib, tuttavia, interpreta LBPPBE come dimensione del blocco e contrassegna in modo permanente l'intestazione ASM quando viene creato il dispositivo ASM.

Questo processo può causare problemi di aggiornamento e migrazione in vari modi, tutti basati sull'impossibilità di combinare dispositivi ASMLib con dimensioni dei blocchi diverse nello stesso gruppo di dischi ASM.

Ad esempio, gli array meno recenti generalmente riportavano un valore LBPPBE pari a 0 o non riportavano affatto questo valore. ASMLib lo interpreta come una dimensione di blocco di 512 byte. Gli array più recenti dovrebbero essere interpretati come aventi una dimensione del blocco di 4KB. Non è possibile combinare dispositivi a 512 byte e 4KB nello stesso gruppo di dischi ASM. In questo modo, si impedirebbe a un utente di aumentare le dimensioni del gruppo di dischi ASM utilizzando LUN di due array o sfruttando ASM come strumento di migrazione. In altri casi, RMAN potrebbe non consentire la copia dei file tra un gruppo di dischi ASM con dimensioni del blocco di 512 byte e un gruppo di dischi ASM con dimensioni del blocco di 4KB.

La soluzione preferita è quella di tamponare ASMLib. L'ID del bug di Oracle è 13999609 e la patch è presente in `oracleasm-support-2,1.8-1` e versioni successive. Questo patch consente all'utente di impostare il parametro `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` a `true` in `/etc/sysconfig/oracleasm` file di configurazione. In questo modo, ASMLib non utilizza il parametro LBPPBE, il che significa che i LUN del nuovo array sono ora riconosciuti come dispositivi a blocchi da 512 byte.



L'opzione non modifica le dimensioni del blocco sui LUN precedentemente contrassegnati da ASMLib. Ad esempio, se un gruppo di dischi ASM con blocchi da 512 byte deve essere migrato in un nuovo sistema di storage che riporta un blocco da 4KB, è possibile scegliere questa opzione `ORACLEASM_USE_LOGICAL_BLOCK_SIZE`. Deve essere impostato prima che i nuovi LUN siano contrassegnati con ASMLib. Se i dispositivi sono già stati contrassegnati da `oracleasm`, è necessario riformattarli prima di essere contrassegnati con una nuova dimensione del blocco. Innanzitutto, deconfigurare il dispositivo con `oracleasm deletedisk`, E quindi cancellare i primi 1GB del dispositivo con `dd if=/dev/zero of=/dev/mapper/device bs=1048576 count=1024`. Infine, se il dispositivo era stato precedentemente partizionato, utilizzare `kpartx`. Per rimuovere le partizioni obsolete o semplicemente riavviare il sistema operativo.

Se ASMLib non può essere aggiornato, ASMLib può essere rimosso dalla configurazione. Questa modifica comporta un'interruzione e richiede la rimozione dello stampaggio dei dischi ASM e la verifica che `asm_diskstring` parametro impostato correttamente. Questa modifica, tuttavia, non richiede la migrazione dei dati.

Dimensioni blocco comando filtro ASM (AFD)

AFD è una libreria di gestione ASM opzionale che sta diventando il sostituto di ASMLib. Dal punto di vista dello storage, è molto simile ad ASMLib, ma include funzionalità aggiuntive come la capacità di bloccare i/o non Oracle per ridurre le possibilità di errori di utenti o applicazioni che potrebbero danneggiare i dati.

Dimensioni dei blocchi dei dispositivi

Come ASMLib, anche AFD legge il parametro LUN Logical Blocks per Physical Block Exponent (LBPPBE) e per impostazione predefinita utilizza la dimensione fisica del blocco, non la dimensione logica del blocco.

Ciò potrebbe creare un problema se l'AFD viene aggiunto a una configurazione esistente in cui i dispositivi ASM sono già formattati come dispositivi a blocchi da 512 byte. Il driver AFD riconosce il LUN come un dispositivo 4K e la mancata corrispondenza tra l'etichetta ASM e il dispositivo fisico impedirebbe l'accesso. Allo stesso modo, le migrazioni sarebbero influenzate dal fatto che non è possibile combinare dispositivi a 512 byte e 4KB nello stesso gruppo di dischi ASM. In questo modo, si impedirebbe a un utente di aumentare le dimensioni del gruppo di dischi ASM utilizzando LUN di due array o sfruttando ASM come strumento di migrazione. In altri casi, RMAN potrebbe non consentire la copia dei file tra un gruppo di dischi ASM con dimensioni del blocco di 512 byte e un gruppo di dischi ASM con dimensioni del blocco di 4KB.

La soluzione è semplice: AFD include un parametro per controllare se utilizza le dimensioni del blocco logico o fisico. Si tratta di un parametro globale che interessa tutti i dispositivi del sistema. Per forzare AFD a utilizzare le dimensioni del blocco logico, impostare `options oracleafd oracleafd_use_logical_block_size=1` in `/etc/modprobe.d/oracleafd.conf` file.

Dimensioni di trasferimento multipath

Le recenti modifiche al kernel linux impongono restrizioni delle dimensioni di i/o inviate ai dispositivi multipath e AFD non rispetta queste restrizioni. Gli i/o vengono quindi rifiutati, il che causa la disconnessione del percorso LUN. Il risultato è un'impossibilità di installare Oracle Grid, configurare ASM o creare un database.

La soluzione consiste nel specificare manualmente la lunghezza massima di trasferimento nel file `multipath.conf` per i LUN ONTAP:

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



Anche se attualmente non esistono problemi, questo parametro deve essere impostato se si utilizza AFD per garantire che un futuro aggiornamento linux non causi inaspettatamente problemi.

Database Oracle con Microsoft Windows

Argomenti di configurazione per database Oracle su Microsoft Windows con ONTAP.

NFS

Oracle supporta l'utilizzo di Microsoft Windows con il client NFS diretto. Questa funzionalità offre un percorso per i vantaggi di gestione di NFS, tra cui la possibilità di visualizzare i file tra più ambienti, ridimensionare dinamicamente i volumi e sfruttare un protocollo IP meno costoso. Consultare la documentazione ufficiale di Oracle per informazioni sull'installazione e la configurazione di un database in Microsoft Windows utilizzando DNFS. Non esistono Best practice speciali.

SAN

Per un'efficienza di compressione ottimale, assicurarsi che il file system NTFS utilizzi un'unità di allocazione di 8K GB o superiore. L'utilizzo di un'unità di allocazione 4K, generalmente predefinita, influisce negativamente sull'efficienza della compressione.

Database Oracle con Solaris

Argomenti di configurazione specifici di Solaris.

Opzioni di montaggio NFS Solaris

Nella tabella seguente sono elencate le opzioni di montaggio di Solaris NFS per una singola istanza.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],roto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
File di controllo File di dati Registri di ripristino	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,llock,suid</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>

L'utilizzo di `llock` è stato dimostrato di migliorare drasticamente le performance negli ambienti dei clienti rimuovendo la latenza associata all'acquisizione e al rilascio di blocchi sul sistema storage. Utilizzare questa opzione con attenzione negli ambienti in cui sono configurati numerosi server per montare gli stessi file system e Oracle è configurato per montare questi database. Sebbene si tratti di una configurazione molto insolita, viene utilizzata da un numero limitato di clienti. Se un'istanza viene avviata accidentalmente una seconda volta, i dati potrebbero danneggiarsi perché Oracle non è in grado di rilevare i file di blocco sul server esterno. I blocchi NFS non offrono altrimenti protezione; come nella versione 3 di NFS, sono solo di natura consultiva.

Perché il `llock` e `forcedirectio` i parametri si escludono a vicenda, è importante che `filesystemio_options=setall` è presente in `init.ora` file in modo che `directio` viene utilizzato. Senza questo parametro, viene utilizzato il caching del buffer del sistema operativo host e le prestazioni possono essere compromesse.

Nella tabella seguente sono elencate le opzioni di montaggio Solaris NFS RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,noac</code>
File di controllo File di dati Registri di ripristino	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,forcedirectio</code>
CRS/votazione	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,forcedirectio</code>
Dedicato ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>
Condiviso ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,suid</code>

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `noac` e `forcedirectio` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare il caching del sistema

operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati. Anche se si utilizza il `init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `noac` e `forcedirectio`.

Il motivo `actimeo=0` è obbligatorio per condiviso `ORACLE_HOME`. Le distribuzioni consentono di semplificare la coerenza di file quali file di password Oracle e file `sfile`. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata `ORACLE_HOME`, questo parametro non è richiesto.

Opzioni di montaggio UFS di Solaris

NetApp consiglia vivamente di utilizzare l'opzione di montaggio della registrazione in modo che l'integrità dei dati venga preservata in caso di arresto anomalo dell'host Solaris o di interruzione della connettività FC. L'opzione di montaggio della registrazione preserva anche l'usabilità dei backup Snapshot.

Solaris ZFS

Solaris ZFS deve essere installato e configurato con attenzione per garantire prestazioni ottimali.

mvector

Solaris 11 ha introdotto una modifica nel modo in cui elabora operazioni i/o di grandi dimensioni, che può causare gravi problemi di prestazioni sugli array di storage SAN. Il problema è documentato in dettaglio nel bug report di NetApp 630173, "riduzione delle prestazioni di Solaris 11 ZFS." La soluzione è modificare un parametro OS chiamato `zfs_mvector_max_size`.

Eseguire il seguente comando come root:

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

Se da questa modifica emergono problemi imprevisti, è possibile annullarli facilmente eseguendo il seguente comando come root:

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

Kernel

Prestazioni ZFS affidabili richiedono un kernel Solaris con patch contro i problemi di allineamento LUN. La correzione è stata introdotta con la patch 147440-19 in Solaris 10 e con SRU 10,5 per Solaris 11. Utilizzare solo Solaris 10 e versioni successive con ZFS.

Configurazione del LUN

Per configurare un LUN, attenersi alla seguente procedura:

1. Creare un LUN di tipo `solaris`.
2. Installare l'host Utility Kit (HUK) appropriato specificato da ["Tool di matrice di interoperabilità NetApp \(IMT\)"](#).
3. Seguire esattamente le istruzioni nell'HUK come descritto. I passaggi di base sono descritti di seguito, ma fare riferimento a ["documentazione più recente"](#) per la procedura corretta.

- a. Eseguire `host_config` utilità per aggiornare `sd.conf/sdd.conf` file. Questo consente alle unità SCSI di rilevare correttamente i LUN ONTAP.
 - b. Seguire le istruzioni fornite da `host_config` Utility per abilitare l'input/output multipath (MPIO).
 - c. Reboot (Riavvia). Questa fase è necessaria per consentire il riconoscimento di eventuali modifiche nel sistema.
4. Partizionare i LUN e verificare che siano allineati correttamente. Vedere "Appendice B: Verifica dell'allineamento WAFL" per istruzioni su come eseguire direttamente il test e confermare l'allineamento.

zpool

Uno zpool deve essere creato solo dopo i passaggi nella "[Configurazione LUN](#)" vengono eseguite. Se la procedura non viene eseguita correttamente, le prestazioni potrebbero peggiorare notevolmente a causa dell'allineamento i/O. Per ottenere prestazioni ottimali con ONTAP è necessario allineare l'i/o a un confine di 4K su un'unità. I file system creati su uno zpool utilizzano una dimensione di blocco effettiva controllata tramite un parametro chiamato `ashift`, che può essere visualizzato eseguendo il comando `zdb -C`.

Il valore di `ashift` il valore predefinito è 9, ovvero 2^9 o 512 byte. Per prestazioni ottimali, la `ashift` Il valore deve essere 12 ($2^{12}=4K$). Questo valore viene impostato al momento della creazione di zpool e non può essere modificato, il che significa che i dati in zpool con `ashift` oltre a 12 deve essere eseguita la migrazione copiando i dati in uno zpool appena creato.

Dopo aver creato uno zpool, verificare il valore di `ashift` prima di procedere. Se il valore non è 12, i LUN non sono stati rilevati correttamente. Distruggere lo zpool, verificare che tutti i passaggi indicati nella relativa documentazione delle utilità host siano stati eseguiti correttamente e ricreare lo zpool.

Zpool e LDOM Solaris

Gli LDOM di Solaris creano un requisito aggiuntivo per assicurarsi che l'allineamento i/o sia corretto. Sebbene un LUN possa essere rilevato correttamente come un dispositivo 4K, un dispositivo `vdsk` virtuale su un LDOM non eredita la configurazione dal dominio i/O. `Vdsk` basato su tale LUN torna per impostazione predefinita a un blocco da 512 byte.

È necessario un file di configurazione aggiuntivo. In primo luogo, i singoli LDOM devono essere aggiornati per Oracle bug 15824910 per abilitare le opzioni di configurazione aggiuntive. Questa patch è stata trasferita in tutte le versioni attualmente utilizzate di Solaris. Una volta installato il software LDOM, è pronto per la configurazione dei nuovi LUN correttamente allineati come segue:

1. Identificare il LUN o i LUN da utilizzare nel nuovo zpool. In questo esempio, si tratta del dispositivo `c2d1`.

```
[root@LDM1 ~]# 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. Recuperare l'istanza `vdc` dei dispositivi da utilizzare per un pool 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. Modifica /platform/sun4v/kernel/drv/vdc.conf:

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

Ciò significa che all'istanza di dispositivo 1 viene assegnata una dimensione di blocco di 4096.

Come ulteriore esempio, si supponga che le istanze vdisk da 1 a 6 debbano essere configurate per una dimensione di blocco di 4K e. /etc/path_to_inst recita:

```
"/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 finale vdc.conf il file deve contenere quanto segue:

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


Attenzione

L'LDOM deve essere riavviato dopo la configurazione di `vdc.conf` e la creazione di `vdisk`. Questa fase non può essere evitata. La modifica delle dimensioni del blocco ha effetto solo dopo un riavvio. Procedere con la configurazione di `zpool` e accertarsi che `ashift` sia impostato correttamente su 12 come descritto in precedenza.

ZFS Intent Log (ZIL)

In genere, non esiste alcun motivo per individuare ZFS Intent Log (ZIL) su un dispositivo diverso. Il registro può condividere lo spazio con il pool principale. L'uso principale di una ZIL separata è quando si utilizzano unità fisiche che non dispongono delle funzionalità di cache di scrittura nei moderni array di storage.

logbias

Impostare `logbias` Parametro sui file system ZFS che ospitano dati Oracle.

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

L'utilizzo di questo parametro riduce i livelli di scrittura complessivi. Per impostazione predefinita, i dati scritti vengono salvati prima nella ZIL e quindi nel pool di storage principale. Questo approccio è appropriato per una configurazione che utilizza una configurazione a disco normale, che include un dispositivo ZIL basato su SSD e supporti rotanti per il pool di storage principale. Questo perché consente l'esecuzione di un commit in una singola transazione i/o sul supporto con latenza più bassa disponibile.

Quando si utilizza un moderno storage array che include funzionalità di caching autonome, questo approccio generalmente non è necessario. In rare circostanze, potrebbe essere opportuno assegnare una scrittura con una singola transazione al registro, ad esempio un carico di lavoro costituito da scritture casuali altamente concentrate e sensibili alla latenza. Vi sono conseguenze sotto forma di amplificazione in scrittura poiché i dati registrati vengono infine scritti nel pool di archiviazione principale, con il risultato di raddoppiare l'attività di scrittura.

I/o diretto

Molte applicazioni, inclusi i prodotti Oracle, possono bypassare la cache del buffer host attivando l'i/o diretto. Questa strategia non funziona come previsto con i file system ZFS. Anche se la cache del buffer host viene ignorata, ZFS continua a memorizzare i dati nella cache. Questa azione può produrre risultati fuorvianti quando si utilizzano strumenti come `fiio` o `sio` per eseguire test delle prestazioni perché è difficile prevedere se l'i/o raggiunge il sistema di storage o se viene memorizzato nella cache locale del sistema operativo. Questa azione rende inoltre molto difficile l'utilizzo di tali test sintetici per confrontare le prestazioni di ZFS con altri file system. In pratica, le performance del file system differiscono da poco a nulla per i carichi di lavoro degli utenti reali.

Diversi zpool

Backup basati su snapshot, ripristini, cloni e archiviazione dei dati basati su ZFS devono essere eseguiti al livello di `zpool` e in genere richiedono più `zpool`. Uno `zpool` è analogo a un gruppo di dischi LVM e deve essere configurato utilizzando le stesse regole. Ad esempio, è probabilmente meglio disporre un database con i file di dati residenti su `zpool1` e i log di archivio, i file di controllo e i log di ripristino che risiedono su `zpool2`. Questo approccio consente un backup a caldo standard in cui il database viene posto in modalità hot backup, seguito da uno snapshot di `zpool1`. Il database viene quindi rimosso dalla modalità di backup a caldo, l'archivio di log viene forzato e viene creata una snapshot di `zpool2` viene creato. Un'operazione di ripristino

richiede lo smontaggio dei file system zfs e l'offlining completo di zpool, in seguito a un'operazione di ripristino di SnapRestore. Lo zpool può quindi essere portato nuovamente online e il database recuperato.

filesystemio_options

Parametro Oracle `filesystemio_options` Funziona in modo diverso con ZFS. Se `setall` oppure `directio` Viene utilizzato, le operazioni di scrittura sono sincrone e ignorano la cache del buffer del sistema operativo, ma le letture sono bufferizzate da ZFS. Questa azione causa difficoltà nell'analisi delle performance perché talvolta l'i/o viene intercettato e gestito dalla cache ZFS, rendendo la latenza dello storage e l'i/o totale inferiori a quanto pare.

Informazioni sul copyright

Copyright © 2024 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIALIZZABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEGUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina <http://www.netapp.com/TM> sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.