

ORACLE 11G DIRECT NFS

İçindekiler

1. Network File System - NFS.....	3
2. Oracle 11G öncesi NFS kullanımı	3
3. Oracle 11G ve Direct NFS	5
3.1. NFS Sunucu olarak hizmet verecek ortamda yapılacaklar.....	5
3.2. NFS İstemci (Client) tarafında yapılacaklar.....	6
3.3. Oracle tarafında yapılacaklar.....	6
4. Performans ve Failover Testleri.....	8
5. Direct NFS Değerlendirmesi	10

1. Network File System - NFS

NFS (Network File System), 1984 yılında Sun Microsystems tarafından geliştirilmiş bir protokoldür. Uzaktaki makine üzerinde bulunan dosya sistem(lerin)i, farklı bir Unix sisteme bağlayabilmeniz (mount) için geliştirilmiştir. Ve bunu yaparken, kullanıcının sanki yerel bir dosya sistemi üzerindeymiş gibi çalışmasını sağlar.

NFS'in şu zamana kadar dört temel sürümü (NFS, NFSv2, NFSv3 ve NFSv4) çıkmıştır. Alt versiyon olarak çıkan NFSv4.1 sürümüyle birlikte paralellik özelliği (pNFS) gelmiştir. Böylece uzaktaki veriye paralel erişim imkanı getirilmiştir.

Hemen hemen bütün Unix tabanlı işletim sistemlerinde NFS'i bulabilirsiniz. Çalıştığı platforma göre NFS araçları yeniden yazılmıştır ve kullandığı parametreler/komutlar farklılık gösterebilir. Ancak farklı mimariye sahip işletim sistemleri kullansanız dahi, NFS ile veri transferini sorunsuzca yapabilirsiniz. Örneğin Linux ile AIX ya da FreeBSD ile Solaris birbirleriyle NFS üzerinden dosya sistemlerini paylaşabilir, veri transferi gerçekleştirebilir. Karşı tarafta hangi dosya sistemi (XFS, ReiserFS, ext3/4 vs...) olduğu ya da kaç bitlik bir platform olduğu sizi etkilemez. Ayrıca sağlam (tutarlı) olması dışında, FTP'ye yakın bir performansa sahiptir; doğru yapılandırma ile network bandwidth'ini efektif kullanır.

2. Oracle 11G öncesi NFS kullanımı

Oracle veritabanını NFS üzerinden kullanmak mümkün. Ancak bu çok rahat olmuyor. Çünkü Oracle dosya sistemlerinden birisinin NFS olduğunu görürse çalışmayı reddediyor. Örneğin NFS dosya sisteminin üzerinde bir şeyler yapmayı (örneğin dosya yaratmak) denerseniz, aşağıdaki gibi bir mesajla işleminiz kesiliyor:

```
ORA-01119: error in creating database file
'/data2/ccebi_test/test8_fs/deneme_ts_01.dbf'
ORA-27054: NFS file system where the file is created or resides
is not mounted with correct options
Additional information: 6
```

Gördüğünüz gibi "***Doğru seçeneklerle bağlanmayan bir dosya sisteminde dosya yaratmaya çalışıyorsunuz.***" şeklinde bir uyarı gelmekte... Doğru seçeneklerin ne olduğunu metalink'ten ya da web sayfalarından bulabilirsiniz. (Fikir vermesi açısından <http://www.dbaexpert.com/blog/?p=40> adresinden aldığım tabloyu aşağıya koyuyorum.)

Mount Options for Oracle files when used with NAS devices

Operating System	Mount options for Binaries	Mount options for Oracle Datafiles	Mount options for CRS Voting Disk and OCR
Sun Solaris	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,noac,vers=3,suid	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,noac,forcedirectio, vers=3,suid	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,vers=3,noac,forcedirectio
AIX (5L)	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,vers=3,timeo=600	cio,rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,noac,vers=3,timeo=600	cio,rw,bg,hard,intr,rsize=32768,wsiz=32768,tcp,noac,vers=3,timeo=600
HPUX 11.23	rw,bg,vers=3,proto=tcp,noac,hard,nointr,timeo=600,rsiz=32768,wsiz=32768,suid	rw,bg,vers=3,proto=tcp,noac,forcedirectio,hard,nointr,timeo=600,rsiz=32768,wsiz=32768,suid	rw,bg,vers=3,proto=tcp,noac,forcedirectio,hard,nointr,timeo=600,rsiz=32768,wsiz=32768,suid
Windows	Not Supported	Not Supported	Not Supported
Linux x86	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,vers=3,timeo=600, actime=0	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,actimeo=0,vers=3,timeo=600	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,noac,vers=3,timeo=600
Linux x86-64	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,vers=3,timeo=600, actime=0	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,actimeo=0,vers=3,timeo=600	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,vers=3,timeo=600, noac
Linux - Itanium	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,vers=3,timeo=600, actime=0	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,actimeo=0,vers=3,timeo=600	rw,bg,hard,nointr,rsize=32768,wsiz=32768,tcp,noac,vers=3,timeo=600

Ancak en doğru parametreleri kullanırsanız bile, uyarı almaya devam ediyorsunuz. Çünkü parametreler n'olursa olsun, Oracle NFS dosya sistemini gördüğü anda işlemi iptal ediyor. Aslında koruma amaçlı, mantıklı bir karar. Çünkü deneyimsiz bir kullanıcının mount edeceği file system'ler, Oracle'ın çalışmasını etkileyebilir; daha da kötüsü tutarlılığını bozup, veritabanını kaybetmenize neden olabilir.)

Riskleri göz önüne alıp, illâki NFS kullanmak istiyorsanız, Oracle'ın NFS kontrolü yapmasını engellemek gerekiyor. Bunun için aşağıdaki ifadenin çalıştırılıp, veritabanının restart edilmesi gerekli:

```
SQL> alter system set event='10298 trace name context forever, level 32' scope = spfile;
```

Veritabanını yeniden başlattıktan sonra, yaptığınız değişikliği `show parameter event;` ile ya da aşağıdaki sorguyla kontrol edebilirsiniz:

```
SQL> select name, value from v$parameter where name = 'event';
```

NAME	VALUE
event	10298 trace name context forever, level 32

Yaptığınız değişikliğin kayda değer bir yan etkisi olmayacaktır¹. Tek yaptığı Oracle'ın NFS kontrolünü engellemek. Bu noktadan sonra NFS ile mount edilmiş dosya sistemlerini dilediğiniz gibi kullanabiliyorsunuz.

3. Oracle 11G ve Direct NFS

11G öncesi NFS dosya sistemleriyle çalışmak, biraz meşakkatli bir uğraştı. Fakat yeni sürümle birlikte Direct NFS isimli bir teknoloji ortaya çıktı. (Bu teknolojinin geliştirilmesinde Oracle, NetApp ile birlikte çalışmış.²) Direct NFS ile birlikte yanlış parametrelerden dolayı sorun çıkmaması ve yönetim kolaylığı sağlanıyor. Ayrıca NFS kullanım maliyetleri düşüyor.

Sistemin yapılandırması hâyli basit. Önce mount edilmiş bir NFS dosya sistemi gerekiyor. Ardından Oracle tarafında bir iki ufak değişiklik yapıyorsunuz ve Direct NFS kullanıma hazır hâle geliyor. Adım adım yapılandırma ile bunu nasıl yapabileceğinizi, aşağıda bulabilirsiniz. (Uygulamada kullanılan ortamlar IBM AIX5L'dir; fakat Linux ya da Solaris için de benzer işlemler yapılacaktır.)

3.1. NFS Sunucu olarak hizmet verecek ortamda yapılacaklar

Server olarak çalışacağınız ortamda NFS Daemon (nfsd) çalışır. Client ise nfs mount araçlarıyla sunucunun dosya sistemlerini kendine bağlar. Öncelikle sisteminizde **nfsd**'nin çalıştığından emin olun:

```
$ ps -ef|grep nfs
   root   868358 1093736    0   Mar 13   -   0:00 /usr/sbin/nfsd 10
```

Başlamadıysa, başlatın:

```
$ startsrc -s nfsd
0513-059 The nfsd Subsystem has been started. Subsystem PID is 315568.
```

Şimdi paylaşım açacağımız, klasörleri (dosya sistemlerini) ayarlayacağız. Bunun için /etc/exportfs dosyasının düzenlenmesi gerekiyor.

```
$ cat /etc/exports
/data2/share_folder/ rw=14.13.1.5 14.13.1.32
```

¹ Metalink: 356199.1

² Direct NFS on EMC NAS : <http://oraclestorageguy.typepad.com/oraclestorageguy/2007/10/direct-nfs-on-e.html>

Yukarıdaki yapılandırmayla /data2/shared_folder/ 14.13.1.5 ve 14.13.1.32 numaralı IP'ye okuma yazma (read/write) açılmış oluyor. (Eğer bu klasörü herkese açmak isteseydik, klasör adından sonra sadece -rw belirtirdik.) exportfs dosyasını değiştirdikten sonra, bunu geçerli hâle getirip, kontrol ediyoruz:

```
$ exportfs -a
$ showmount -e localhost
export list for localhost:
/data2/shared_folder/ rw=14.13.1.5,14.13.1.32
```

3.2. NFS İstemci (Client) tarafında yapılacaklar

Sunucu tarafında işlemleri doğru yaptıysak, dosya sistemini client'a bağlayabiliriz. İstemci tarafında yapacaklarımız hâyli basit:

```
$ mount 14.13.1.27:/data2/shared_folder/ /data2/nfs_fs/
```

14.13.1.27 numaralı NFS sunucumuzun (nfs_server) /data2/shared_folder/ dosya sistemini (klasörünü), istemci olarak kullandığımız makinenin /data2/nfs_fs/ klasörüne bağladık. Verdiğim örneklerde, IP kullandım; ama elbette böyle bir zorunluluk yok, hostname de kullanabilirsiniz. Şimdi gerçekten bağlanmış mı diye kontrol yapıyoruz:

```
$ mount |grep nfs
10.3.1.22 /data2/shared_folder /data2/nfs_fs/ nfs3 Apr 29 10:30
```

3.3. Oracle tarafında yapılacaklar

Bahsedeceğimiz işlemleri veritabanı kapalıyken yapmak gerekiyor. Öncelikle Oracle Standard Disk Manager (ODM) kütüphanesini (*libodm11.so*), NFS desteği olan kütüphaneyle (*libnfsodm11.so*) değiştirmek gerekiyor. Bunun için aşağıdaki adımları takip edebilirsiniz:

```
$ cd $ORACLE_HOME/lib/
$ cp libodm11.so libodm11.so_stub
$ ln -sf libnfsodm11.so libodm11.so
$ ls libodm11.so
lrwxrwxrwx 1 orayat dba 14 Apr 29 14:03 libodm11.so -> libnfsodm11.so
```

Eğer NFS kullanımını iptal etmek isterseniz, libodm11.so_stub dosyasını kullanabilir ya da aşağıdaki gibi symbolic link'i tekrar yaratırsınız:

```
$ ln -sf libodm11.so libodm11.so
```

Şimdi sıra Oracle'ın mount point'leri göreceği dosyayı hazırlamaya geldi. Oracle mount edilen dosya sistemlerini üç yere göre belirler:

- i. \$ORACLE_HOME/dbs/oranfstab** : Oracle'ın mount point'ler için bakacağı ilk dosyadır. Sadece veritabanına özgüdür. Diğer veritabanları görmez.
- ii. /etc/oranfstab** : Bütün veritabanları tarafından ortak kullanılan dosyadır. Buraya kaydettiğiniz bir alan, sistemdeki bütün veritabanları tarafından kullanılabilir.
- iii. /etc/mtab** : Eğer ilk iki dosya da kayıt yoksa, işletim sisteminden NFS mount noktaları çekilir. Linux için bu /etc/mtab'dir.

Aynı dosya sistemini işaret eden birden fazla kayıt varsa, ilk okuduğu dosya dikkate alınır. Biz çalışmamız için oranfstab dosyasını aşağıdaki gibi oluşturalım:

```
$ cat $ORACLE_HOME/dbs/oranfstab
server: nfs_server
path: 14.13.1.27
path: 14.13.1.28
export: /data2/shared_folder/ mount:/data2/nfs_fs/
```

ALAN AÇIKLAMALARI:

- server** : Bağlandığımız NFS sunucu
- path** : NFS sunucusuna bağlanacağınız Network arayüz kartının adresidir. Dört adede kadar path tanımlayabilirsiniz. Path'lerin yük dağılımı Oracle tarafından otomatik yapılır. Çöken veya erişilemeyen path olursa, yine Oracle tarafından devre dışına alınır.
- export** : NFS ile uzaktan eriştiğimiz klasörü temsil eder.
- mount** : NFS dosya sistemini bağladığımız yerel klasörü gösterir.

Bütün işlemler tamam; artık Oracle'ı NFS üzerinden kullanabiliriz. Veritabanını açın ve bu sırada alertlog dosyasını takip edin. Aşağıdaki gibi bir ifadeyi alertlog dosyasında göreceksiniz:

```
Oracle instance running with ODM: Oracle Direct NFS ODM Library
Version 2.0
```

Performans ya da durum takibi yapmak isterseniz, aşağıdaki view'lar işinize yarayacaktır:

- v\$dnfs_servers** : NFS üzerinden eriştiğimiz sunuculara (ve dosya sistemlerine) dair bilgi verir.
- v\$dnfs_files** : DirectNFS kullanan veritabanı dosyalarına dair bilgi verir.
- v\$dnfs_channels** : Hangi path'lerden hangi sunuculara eriştiğimiz gösterilir.
- v\$dnfs_stats** : Direct NFS ile ilgili performans bilgisi sunar.

4. Performans ve Failover Testleri

Direct NFS ilk duyurulduğunda 32bit'lik Oracle veritabanında yapılan bazı testler hâyâlkırıklığı yaratmıştı³. Zamanla bu testleri yapıp, olumsuz görüş bildirenler fikirlerini değiştirdi. Oracle yayınlamış olduğu dökümanlarda da yüksek performans gösteren sonuçlar çıktı.⁴ Bu sonuçların güvenilir olduğunu düşünebiliriz. Bu yüzden yapacağımız testleri daha basit tutacağız.

Her denemede nologging mode'da boş bir tablespace yaratacağız. Ardından bu tablespace altında yine nologging mode'da bir tablo oluşturup, tekrar tekrar okuma ve yazma işlemleri gerçekleştireceğiz. (Testlerde nologging mode'da çalışmamızın en büyük nedeni, redolog'lar yerel disklerde tutulduğu için etkisini azaltmak.)

```
DROP TABLESPACE DENEME_TS INCLUDING CONTENTS AND DATAFILES;  
CREATE TABLESPACE DENEME_TS DATAFILE  
  '/data2/nfs_fs/deneme_ts_01.dbf' SIZE 1M AUTOEXTEND ON NEXT 8K MAXSIZE  
UNLIMITED,  
  '/data2/nfs_fs/deneme_ts_02.dbf' SIZE 1M AUTOEXTEND ON NEXT 8K MAXSIZE  
UNLIMITED  
NOLOGGING  
BLOCKSIZE 8K;  
  
CREATE TABLE D_CCEBI.DENEME  
TABLESPACE DENEME_TS  
NOLOGGING  
PARALLEL (DEGREE 4)  
AS SELECT * FROM DBA_OBJECTS;
```

Paralelliği iyi kullanabilmek için tablonun paralellik derecesini 4 olarak ayarlıyoruz. (Tablespace dosyası da bu yüzden birden fazla olarak yaratıldı.)

³ Direct NFS on EMC NAS: <http://oraclestorageguy.typepad.com/oraclestorageguy/2007/10/direct-nfs-on-e.html>

⁴ Oracle Database 11g Direct NFS Client, Temmuz 2007

```
SET TIMING ON
BEGIN
  FOR i IN 1..5 LOOP
    insert /*+ APPEND */ into d_ccebi.deneme
      select * from d_ccebi.deneme;
    COMMIT;
  END LOOP;
END;
ALTER SYSTEM CHECKPOINT;
```

TEST – I : *İki Path Tanımlı NFS Bağlantı*

İşlem sonunda toplam boyutu 292 MB'lık iki adet dosya oluştu. Döngü içindeki yazma 2dk. 33sn. sürdü.

TEST – II : *Tek Path Tanımlı NFS Bağlantı*

İşlem sonunda 300 MB'lık iki adet dosya oluştu. Döngü içi süre 2dk. 16 sn. sürdü. Tek path verilmesine rağmen daha kısa çıkması elbette şaşırtıcı bir durum. Fakat network trafiğini zorlamadığımız düşünülürse, 2 path tanımının hız bakımından fayda sağlamaması normal karşılanabilir. Kaldı ki iki path denetlenmesinin maliyetini de hesaba katmak gerekir. Burada tek anormal durum boyutun fazla çıkması...

TEST – III : *İki path tanımlı NFS Bağlantıda NIC'lerden birinin bozulması*

Bu test Direct NFS'in failover becerisini ölçmek için yapıldı. İşlem esnasında network interface'lerden birini kapatılarak gerçekleştirildi. Interface'ı kapattıktan sonra, alert log'da aşağıdaki uyarılar verildi:

```
Direct NFS: channel id 1 path 192.168.1.22 to filer nfs_server PING
timeout
```

Veritabanı işleme devam etti. Bunun üzerine işlem devam ederken network interface'i 4 defa daha kapatıp açtım. İşlem yine devam etti. Yapı çalışıyor. Ancak önemli bir sıkıntı, işlemin çok yavaş devam etmesi. Bu denemede 9 dk. 51 sn sürdü. Süre %77 daha fazla. Yani kartlardan biri giderse, Oracle toparlanıyor ama bu işi çok yavaş yapıyor.

TEST – IV : *Yanlış yapılandırma*

Kasıtlı olarak yanlış path tanımları yapıldı. Ardından alert log'da aşağıdaki uyarılar alındı, tablespace ile ilgili işlem yapılamadı:

```
Direct NFS: warn could not mount /data2/nfs_fs on svr test2 via dNFS
Direct NFS: warn could not get mount handle from svr test2 nfSPORT
2049 mntport 32774
```

TEST – V : *Eski NFS Yöntemi*

Eski biçimde NFS'i kullanmayı denediğimizde, 2dk. 33 sn. sürdü. Bu Direct NFS testleriyle tamamen aynı.

TEST – VI : *İşlemlerin lokalde gerçekleştirilmesi*

Aynı işlemler lokal disklerin üstünde gerçekleştiğinde 24 saniye sürüyor.

5. Direct NFS Değerlendirmesi

Yaptığım testlerde hız bakımından şaşırtıcı sonuçlara ulaşamadım. Eski NFS'e göre bazı noktalarda daha hızlı çalıştı. Ancak hissedilir bir fark görmedim. Çalışmada network bağlantısını sonuna kadar kullanmadığımı hissettim. Belki düz bir tablo yerine farklı farklı birçok datafile'a erişen bir yapı çalışsa, sonuç daha farklı olurdu.

Performansı geçerse, Direct NFS'in benim için en büyük özelliği failover becerisi oldu. Ethernet'lerden biri gittiğinde işlem çok yavaşlıyor ama Oracle zarar görmüyor. (Hem de defalarca ethernet'i açıp kapatmama rağmen...) Böyle bir yapı, EtherChannel, Teaming gibi bütün yapıları gereksiz kılabilir. Üstelik EtherChannel/Teaming vb. teknolojilerin sahip olamayacağı kadar esnek.

Teknoloji olarak Direct NFS'in sunduğu imkanlar ve kullanım kolaylığı büyük avantaj. NFS vb. teknolojileri çok bilmeden de bu yapıyı kurabilmek bile büyük artı. Performans konusunun ise 11gR2'de daha da iyileştirileceğini umuyorum.