ZFS STORAGE APPLIANCE

Oracleテクニカル・ホワイト・ペーパー
2013年10月

Oracle Linuxと Oracle ZFS Storage Applianceの マルチパスの構成

概要 2
はじめに 2
DM-MultipathとOracle Linux 3
DM-Multipathのインストール 4
Oracle LinuxでのDM-Multipathの構成 7
DM-Multipathのデバイス・ノードとシンボリック・リンク 12
DM-Multipath構成の検証 12
DM Multipathでのパス障害検出の検証 13
デバイスの所有権 18
Oracle Linux Release 5 18
Oracle Linux Release 6 19
結論 21
参考資料 22

概要

可用性に優れたサービス構成では、データ・アクセスの耐障害性と継続性を確保するために、データ・パス内にさまざまなコンポーネントが重複して配置されます。このような重複する冗長なコンポーネントは、シングル・ポイント障害を回避するために使用されます。この冗長なセットアップでは、これらの複数のコンポーネントやパスを認識して効率的かつ効果的に使用できるように、サーバーを構成する必要があります。この効果的な冗長化を実現するには、一貫した名前と所有権が重要となります。

Oracle Linuxは、Linux DM-Multipath（Device-Mapper-Multipath）に基づいて構成可能なマルチパス・ソリューションを提供します。DM-Multipathでは、土台からの構成が必要になります。このホワイト・ペーパーでは、一貫した方法でボリュームを表示できるように、iSCSI環境のOracle ZFS Storage Applianceファミリー用にDM-Multipathを構成する方法について説明します。

標準のLinuxデフォルトを使用する場合を除き、Oracle ZFS Storage Applianceで提供されるLUNがOracle Automatic Storage Manager（Oracle ASM）で使用される場合など、特定の状況では、これらのボリュームの所有権を確立する必要があります。このような場合、管理者の介入なしにデータベースを自動的に起動できるように、Oracle LinuxのデバイスマネージャUDEVを使用して、起動時にボリュームの所有権を適切な所有者に自動的に変更できます。

はじめに

Oracle ZFS Storage Applianceファミリーは、数多くのさまざまなオペレーティング・システムに耐障害性のあるストレージを提供しますが、このストレージへの耐障害性のあるアクセスをインフラストラクチャで実現できるかどうかは、Oracle ZFS Storage Applianceおよびアクセスが許可されているホストの構成で決まります。

このホワイト・ペーパーでは、この耐障害性のあるアクセスをOracle LinuxのDM-Multipath機能を使用して実装する方法、およびDM-MultipathとOracle LinuxのUDEVで提供される他の機能を使用して、Oracle ZFS Storage Applianceで提供されているLUNの一貫したネーミングと所有権を実現する方法について説明します。

このホワイト・ペーパーでは、次の図に示す構成例を前提としています。2台のOracle Linuxサーバー（片方がRelease 5、もう片方がRelease 6）が3つのデータ・ネットワークと1つの管理ネットワークに接続されています。

Oracle ZFS Storage Applianceが展開されており、3つの各データ・ネットワークと1つの管理ネットワークに接続されています。

3つのデータ・ネットワークは、高可用性システムに必要となる複数の冗長パスを提供するタスク専用となっています。

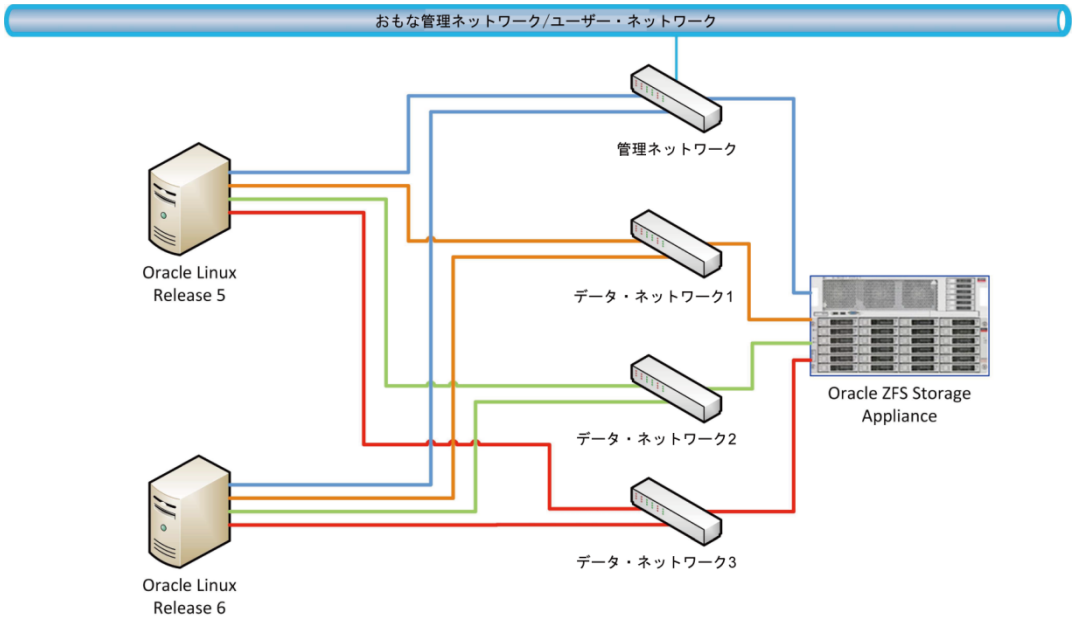

図1：マルチパス化されたOracle LinuxインスタンスとOracle ZFS Storage Applianceのサンプル・ネットワーク図

DM-MultipathとOracle Linux

DM-MultipathはOracle Linuxで提供されている機能で、障害の発生時と解決時にストレージへの複数のパスを自動的に管理および維持できるため、管理者の介入はまったく必要なくなるか最小限に抑えられます。

図2のブロック・モデルは、データ・パス・スタックの機能を表した概要図です。

DM-Multipathは、ボリューム・マネージャ（Linux論理ボリューム・マネージャ（LVM）やOracle Automatic Storage Managementなど）とストレージ・プロバイダ（この場合、SCSIレイヤー）の間に配置されます。DM-Multipathは、ストレージ・プロバイダの抽象化レベルを提供するため、耐障害性のある構成に必要となる冗長性は示しません。

ここに示す例では、Oracle ZFS Storage Applianceによって提供されるLUNは、SCSIレイヤーの3つのデバイス・ノードで表され、Linuxサーバーの/devに3つのSCSIディスク・デバイスとして表示されます。

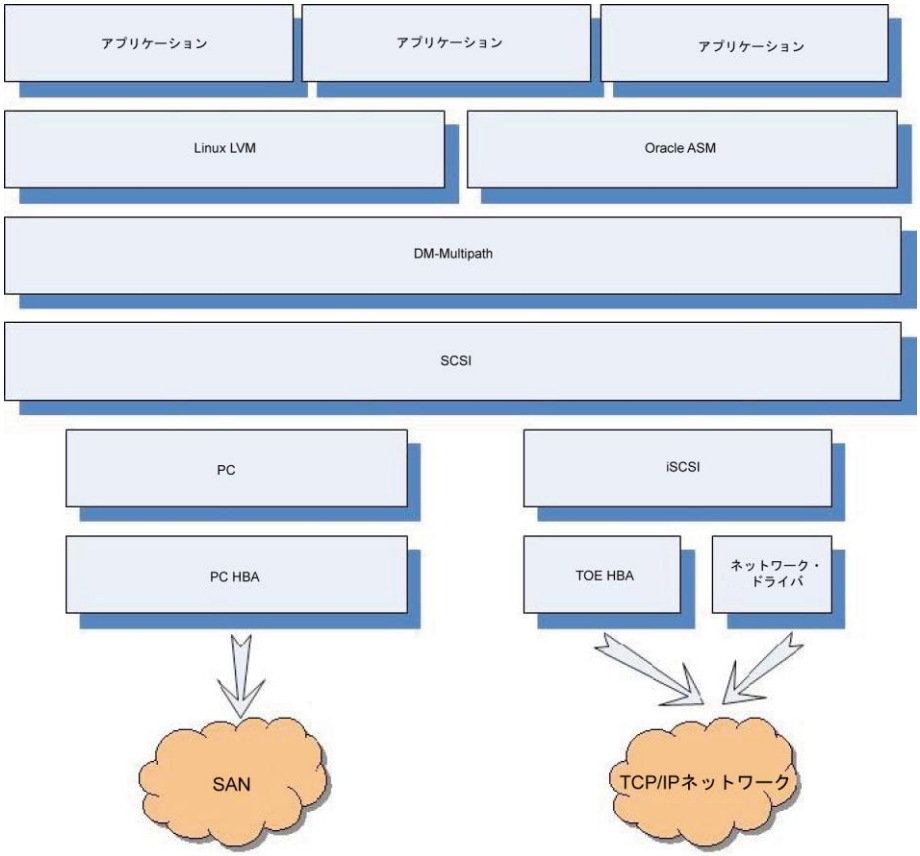

図2：DM Multipathの機能的概要

DM-Multipathのインストール

DM-Multipathのインストールが、最初のシステム・インストール時に含まれるかどうかは、どのインストール・パッケージを選択したのかによって決まります。インストールが必要な場合は、インストール・メディアがあるか、オラクルの一般向けのYellowdog Updater Modified（YUM）サーバーにアクセスできるようにOracle Linuxサーバーが構成されていれば、DM-Multipathをインストールするのは難しいことではありません。オラクルの一般向けのYUMサーバーをサーバーのインストール後に使用して、パッケージの更新と修正を行うことができます。

サーバーからオラクルの一般向けのYUMサーバーにアクセスできるようにするには、Oracle Linuxサーバーに正しい構成ファイルが配置されていることを確認します。

ここに示す最初の例は、Oracle Linux Release 6の場合のものです。

```
[root@ol-r6 ~]# cd /etc/yum.repos.d
[root@ol-r6 yum.repos.d]# ls -l
total 0
[root@ol-r6 yum.repos.d]# wget http://public-yum.oracle.com/public-yum-
ol6.repo
--2013-02-25 12:50:32--  http://public-yum.oracle.com/public-yum-ol6.repo
Resolving public-yum.oracle.com... 141.146.44.34
Connecting to public-yum.oracle.com|141.146.44.34|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2411 (2.4K) [text/plain]
Saving to: “public-yum-ol6.repo”

100%[=====================================>] 2,411       --.-K/s   in 0.001s

2013-02-25 12:50:32 (3.80 MB/s) - “public-yum-ol6.repo” saved [2411/2411]

[root@ol-r6 yum.repos.d]# vi public-yum-ol6.repo
//
// Ensure that the first entry, [ol6_latest] ends with the line enabled=1
//
```

2番目の例は、Oracle Linux Release 5の場合のものです。

```
[root@ol-r5 /]# cd /etc/yum.repos.d
[root@ol-r5 yum.repos.d]# ls -l
total 0
[root@ol-r5 yum.repos.d]# wget http://public-yum.oracle.com/public-yum-
el5.repo
--2013-02-25 13:04:47--  http://public-yum.oracle.com/public-yum-el5.repo
Resolving public-yum.oracle.com... 141.146.44.34
Connecting to public-yum.oracle.com|141.146.44.34|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4220 (4.1K) [text/plain]
Saving to: `public-yum-el5.repo'

100%[=====================================>] 4,220       --.-K/s   in 0.004s

2013-02-25 13:04:47 (1.05 MB/s) - `public-yum-el5.repo' saved [4220/4220]

[root@ol-r5 yum.repos.d]# vi public-yum-el5.repo
//
```

```
// Ensure that the first entry, [el5_latest] ends with the line enabled=1
//
```

Webページ（http://public-yum.oracle.com）に、各Oracle Linux Releaseの手順が記載されています。

この時点では、Oracle Linux Release 5とOracle Linux Release 6の手順は同じです。まず、正しいパッケージ名を検索します。

```
[root@ol-r6 yum.repos.d]# yum search multipath
Loaded plugins: refresh-packagekit, security
============================ N/S Matched: multipath
============================
device-mapper-multipath.x86_64 : Tools to manage multipath devices using
                               : device-mapper
device-mapper-multipath-libs.x86_64 : The device-mapper-multipath modules and
                                    : shared library

  Name and summary matches only, use "search all" for everything.
```

次に、YUMに組み込まれている依存性解決により、ツール・パッケージ（device-mapper-multipath）を指定するだけで済みます。YUMにより、実行される操作を承認するようユーザーの確認を求めるプロンプトが表示され、必要となる他のパッケージ（この場合、device-mapper-multipath-libsパッケージ）が管理されます。

```
[root@ol-r6 yum.repos.d]# yum install device-mapper-multipath
Loaded plugins: refresh-packagekit, security
Setting up Install Process
Resolving Dependencies
--> Running transaction check
---> Package device-mapper-multipath.x86_64 0:0.4.9-56.el6_3.1 will be
installed
--> Processing Dependency: device-mapper-multipath-libs = 0.4.9-56.el6_3.1
for package: device-mapper-multipath-0.4.9-56.el6_3.1.x86_64
--> Processing Dependency: libmultipath.so()(64bit) for package: device-
mapper-multipath-0.4.9-56.el6_3.1.x86_64
--> Running transaction check
---> Package device-mapper-multipath-libs.x86_64 0:0.4.9-56.el6_3.1 will be
installed
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================
 Package                        Arch     Version             Repository    Size
```

```
================================================================================
Installing:
 device-mapper-multipath        x86_64   0.4.9-56.el6_3.1    ol6_latest    96 k
Installing for dependencies:
 device-mapper-multipath-libs   x86_64   0.4.9-56.el6_3.1    ol6_latest   158 k

Transaction Summary
================================================================================
Install       2 Package(s)

Total download size: 254 k
Installed size: 576 k
Is this ok [y/N]: y
Downloading Packages:
(1/2): device-mapper-multipath-0.4.9-56.el6_3.1.x86_64.r |  96 kB     00:00
(2/2): device-mapper-multipath-libs-0.4.9-56.el6_3.1.x86 | 158 kB     00:00
--------------------------------------------------------------------------------
Total                                            104 kB/s | 254 kB     00:02
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing : device-mapper-multipath-libs-0.4.9-56.el6_3.1.x86_64         1/2
  Installing : device-mapper-multipath-0.4.9-56.el6_3.1.x86_64              2/2
  Verifying  : device-mapper-multipath-0.4.9-56.el6_3.1.x86_64              1/2
  Verifying  : device-mapper-multipath-libs-0.4.9-56.el6_3.1.x86_64         2/2

Installed:
  device-mapper-multipath.x86_64 0:0.4.9-56.el6_3.1

Dependency Installed:
  device-mapper-multipath-libs.x86_64 0:0.4.9-56.el6_3.1

Complete!
[root@ol-r6 yum.repos.d]# chkconfig multipathd on
```

これで、DM-Multipathがインストールされました。次の手順では、特定のセットアップの構成を行います。

Oracle LinuxでのDM-Multipathの構成

注：次の手順は、Oracle Linux Release 5およびOracle Linux Release 6の両方にこの形式で適用可能です。

DM-Multipathでは、メインの構成ファイル/etc/multipath.confが使用されます。このファイルは、コンテキストが非常に主観的であるため、インストール時には作成されません。構成の最初のステップで、このファイルを作成します。これを半自動的に行うためのツールがあります。ただし、作成された構成ファイルは、容易に展開できますが、実装前にさらに編集する必要があるため、このファイルをエディタで直接作成することも選択できます。

/etc/multipath.confファイルのコンテンツは、少なくとも次のようになっている必要があります。

```
devices {
        device {
                vendor              "SUN"
        }
}
```

マルチパス対応デバイスがOracle Linuxサーバーにすでに構成されている場合、これらのデバイスをそのままにし、上記の行をdevices宣言に追加する必要があります。他のOracle ZFS Storage Appliance製品が定義されている場合、上記のvendor宣言に加えて、product宣言を追加することが必要になる場合があります。詳細については、multipath.conf(5)のマニュアル・ページを参照してください。

Oracle ZFS Storage Applianceによって提供される特定のLUNの情報が判明したら、/etc/multipath.confファイルをさらに編集して構成をカスタマイズする必要があります。

スケルトン構成ファイルが配置されている場合、ここで、iSCSIサブシステムを起動できるようになります。起動すると、特定のLUNの構成情報が表示されます。

```
[root@ol-r5 ~]# service iscsid start
Starting iSCSI daemon:                                [  OK  ]
root@ol-r5 ~]# iscsiadm -m discovery -t sendtargets -p 192.168.1.13
192.168.2.13:3260,2 iqn.1986-03.com.sun:02:bfe9431c-3bba-c1f2-d72f-
a0829cc2d637
192.168.3.13:3260,3 iqn.1986-03.com.sun:02:bfe9431c-3bba-c1f2-d72f-
a0829cc2d637
192.168.4.13:3260,4 iqn.1986-03.com.sun:02:bfe9431c-3bba-c1f2-d72f-
a0829cc2d637
[root@ol-r5 ~]# iscsiadm -m node --login
Logging in to [iface: default, target: iqn.1986-03.com.sun:02:bfe9431c-3bba-
c1f2-d72f-a0829cc2d637, portal: 192.168.3.13,3260] (multiple)
Logging in to [iface: default, target: iqn.1986-03.com.sun:02:bfe9431c-3bba-
c1f2-d72f-a0829cc2d637, portal: 192.168.2.13,3260] (multiple)
Logging in to [iface: default, target: iqn.1986-03.com.sun:02:bfe9431c-3bba-
c1f2-d72f-a0829cc2d637, portal: 192.168.4.13,3260] (multiple)
Login to [iface: default, target: iqn.1986-03.com.sun:02:bfe9431c-3bba-c1f2-
d72f-a0829cc2d637, portal: 192.168.3.13,3260] successful.
Login to [iface: default, target: iqn.1986-03.com.sun:02:bfe9431c-3bba-c1f2-
d72f-a0829cc2d637, portal: 192.168.2.13,3260] successful.
Login to [iface: default, target: iqn.1986-03.com.sun:02:bfe9431c-3bba-c1f2-
d72f-a0829cc2d637, portal: 192.168.4.13,3260] successful.
[root@ol-r5 ~]# service multipathd start
Starting multipathd daemon:                                [  OK  ]
```

```
[root@ol-r5 ~]# multipath -ll
3600144f0c532bda00000512a754e0001 dm-2 SUN,ZFS Storage 7420
size=4.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 4:0:0:0 sdb 8:16 active ready running
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 5:0:0:0 sdc 8:32 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 3:0:0:0 sdd 8:48 active ready running
3600144f0c532bda00000512a75600002 dm-3 SUN,ZFS Storage 7420
size=8.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=active
| `- 5:0:0:1 sdf 8:80 active ready running
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 3:0:0:1 sdg 8:96 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 4:0:0:1 sde 8:64 active ready running
```

この最後のコマンド`multipath -ll`で、シンボリック名をボリュームに指定するために必要な情報が表示されます。`dm-<n>`名の前に表示されている識別子は、Oracle ZFS Storage Applianceで提供されているLUNのWWNであり、バス識別子が先頭に追加されています。この例では、バス識別子は'3'です。

WWNは、図3に示すように、Oracle ZFS Storage Applianceのブラウザ・ユーザー・インタフェース（BUI）のGUIDの下に表示されます。

図3：提供されているLUNおよびLUNのシンボリック名

この図から、各Oracle Linuxサーバーに表示されている、2つのLUN（1つの4GB LUNと1つの8GB LUN）が存在することが分かります。

Oracle Linux R6サーバーの`ol-r5`では、Oracle ZFS Storage Applianceによって表示されている2つのWWNは、`600144F0C532BDA00000512A754E0001`と`600144F0C532BDA00000512A75600002`です。

これは、前述の`multipath -ll`コマンドの出力で確認されています。Oracle ZFS Storage Applianceによってレポートされるiscsi修飾名（IQN）と、DM-Multipathで必要となるIQNでは、大文字と小文字が異なることに注意してください。

この例では、4GBのボリュームはOracle ASMに使用されるため、シンボリック名`ora_redo`が付与され、8GBのボリュームはアプリケーションに使用されるため、`app_space`が付与されています。

特定のLUNにこれらの名前を指定するには、`/etc/multipath.conf`ファイルを再度編集する必要があります。`multipaths`宣言を追加すると、LUNのWWNが指定されます。

次の例では、Oracleボリュームが表示されます。

```
multipaths {
        multipath {
                wwid            3600144f0c532bda00000512a754e0001
                alias           ora_redo
        }

        multipath {
                wwid            3600144f0c532bda00000512a75600002
                alias           app_space
        }
}
```

device宣言と同様に、複数の`multipath {…}`宣言を`multipaths {…}`文に配置できます。

`/etc/multipath.conf`ファイルに、次の画面出力と同様のものが含まれるようになりました。言うまでもなく、WWNは個々の状況によって異なります。

```
#
# We don't want user_friendly_names as this results in just WWNs being used
# Our configuration has symbolic names for the LUNs to allow automatic
processing
# by UDEV later...
#
defaults {
        user_friendly_names     no
}

devices {
        device  {
                vendor                  "SUN"
                path_checker            readsector0
                hardware_handler        0
                failback                15
                rr_weight               priorities
```

```
                no_path_retry           queue
        }
}

multipaths {
        multipath {
                wwid            3600144f0c532bda00000512a754e0001
                alias           ora_redo
        }

        multipath {
                wwid            3600144f0c532bda00000512a75600002
                alias           app_space
        }
}
```

最後に、選択した名前を指定するには、DM-Multipathで構成を再度読み込む必要があります。

```
[root@ol-r5 ~]# multipath -r
rename: ora_redo (3600144f0c532bda00000512a754e0001) undef SUN,ZFS Storage 7420
size=4.0G features='0' hwhandler='0' wp=undef
|-+- policy='round-robin 0' prio=1 status=undef
| `- 4:0:0:0 sdb 8:16 active ready running
|-+- policy='round-robin 0' prio=1 status=undef
| `- 5:0:0:0 sdc 8:32 active ready running
`-+- policy='round-robin 0' prio=1 status=undef
  `- 3:0:0:0 sdd 8:48 active ready running
Jun 11 17:55:48 | 3600144f0c532bda00000512a75600002: rename
3600144f0c532bda00000512a75600002 to app_space
rename: app_space (3600144f0c532bda00000512a75600002) undef SUN,ZFS Storage 7420
size=8.0G features='0' hwhandler='0' wp=undef
|-+- policy='round-robin 0' prio=1 status=undef
| `- 4:0:0:1 sde 8:64 active ready running
|-+- policy='round-robin 0' prio=1 status=undef
| `- 5:0:0:1 sdf 8:80 active ready running
`-+- policy='round-robin 0' prio=1 status=undef
  `- 3:0:0:1 sdg 8:96 active ready running
[root@ol-r5 ~]# cd /dev/mapper
[root@ol-r5 mapper]# ls -l
total 0
brw-rw---- 1 root disk 253,  3 Jun 11 17:51 app_space
crw------- 1 root root  10, 62 Jun 11 17:51 control
brw-rw---- 1 root disk 253,  2 Jun 11 17:51 ora_redo
brw-rw---- 1 root disk 253,  0 Jun 11 17:51 VolGroup00-LogVol00
brw-rw---- 1 root disk 253,  1 Jun 11 17:51 VolGroup00-LogVol01
[root@ol-r5 mapper]#
```

DM-Multipathでは、再起動のたびにデバイス・ノードにこれらの名前が使用されます。検出順序に応じて再起動のたびに異なる可能性がある、非確定的な`dm-<n>`名は適用されません

DM-Multipathのデバイス・ノードとシンボリック・リンク

DM-Multipathでは、内部使用と外部使用のために、追加のリンクとデバイス・ノードが自動的に作成されて保持されます。次の表に、これらのリンクとデバイス・ノードのネーミングと用途を示します。

デバイス/シンボリック・リンク	用途
/dev/sd*	サーバーからアクセスできる各LUNに、そのLUNへのすべてのパスの`/dev/sd`エントリが作成されます。これらのエントリがアプリケーションによって直接使用されないようにしてください。
/dev/dm-*	DM-Multipath内部名。アプリケーションによって直接使用されないようにしてください。
/dev/mapper/* /dev/mpath/*	`/dev/mapper/*`エントリは、起動プロセスの早い段階で作成され、このプロセス中に必要となるボリュームに使用されます。`/dev/mpath/*`エントリは後で作成され、起動プロセスの完了後に開始されるアプリケーションによって使用されます。 `/dev/mpath`ディレクトリは、アクセス可能な同じ場所にすべての`/dev/mpath/*`エントリを簡単に保持するために作成されます。これらのデバイス（エントリ）は、起動プロセス中は使用できません。シンボリック・リンクは、さまざまな方法でストレージにアクセスできるようにするために、`/dev/disk`ディレクトリに作成されます。作成されるリンクの名前は、非常に長くなる傾向があるため、アプリケーション構成に使用できないことがあります。
/dev/disk/by-id/x /dev/disk/by-label/x /dev/disk/by-path/x /dev/disk/by-uuid/x	シンボリック・リンクは、さまざまな方法でストレージにアクセスできるようにするために、`/dev/disk`ディレクトリに作成されます。作成されるリンクの名前は、非常に長くなる傾向があるため、アプリケーション構成に使用できないことがあります。

DM-Multipath構成の検証

`multipath -ll`の出力から、各ボリュームに3つのアクティブ・パスが存在することがわかります。

```
[root@ol-r5 mapper]# multipath -ll
app_space (3600144f0c532bda00000512a75600002) dm-3 SUN,ZFS Storage 7420
size=8.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 4:0:0:1 sde 8:64 active ready running
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 5:0:0:1 sdf 8:80 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 3:0:0:1 sdg 8:96 active ready running
ora_redo (3600144f0c532bda00000512a754e0001) dm-2 SUN,ZFS Storage 7420
```

```
size=4.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 4:0:0:0 sdb 8:16 active ready running
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 5:0:0:0 sdc 8:32 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 3:0:0:0 sdd 8:48 active ready running
[root@ol-r5 mapper]#
```

このコード例では、パスは3:0:0、4:0:0、および5:0:0に存在し、すべてがactive ready runningステータスになっています。

DM Multipathでのパス障害検出の検証

複数パスの使用は、障害発生時にDM-Multipathが既存のパスを効果的に機能させることができる場合のみ役立ちます。これを実際にテストできます。次の例に、適切なエラー・リカバリが実行されて、障害にかかわらずI/O処理が続行されることを確認するためにシミュレートした、ボリュームの1つで生成されるワークロードおよび障害を示します。

ターミナル・セッションで、次のコマンドを実行します。

```
[root@ol-r5 ~]# while /bin/true
> do
> dd if=/dev/mapper/app_space of=/dev/null
> done
```

別のターミナル・セッションで、ファイル/var/log/messagesに対し常にtailを実行させて、障害時のDM-Multipathの状態を監視します。

```
[root@ol-r5 ~]# tail -f /var/log/messages &
[root@ol-r5 ~]# ifconfig eth3 down
[root@ol-r5 ~]# Jun 11 17:59:19 ol-r5 avahi-daemon[3452]: Interface eth3.IPv6
no longer relevant for mDNS.
Jun 11 17:59:19 ol-r5 avahi-daemon[3452]: Leaving mDNS multicast group on
interface eth3.IPv6 with address fe80::a00:27ff:fe96:2c9c.
Jun 11 17:59:19 ol-r5 avahi-daemon[3452]: Interface eth3.IPv4 no longer
relevant for mDNS.
Jun 11 17:59:19 ol-r5 avahi-daemon[3452]: Leaving mDNS multicast group on
interface eth3.IPv4 with address 192.168.4.82.
Jun 11 17:59:19 ol-r5 avahi-daemon[3452]: Withdrawing address record for
fe80::a00:27ff:fe96:2c9c on eth3.
Jun 11 17:59:19 ol-r5 avahi-daemon[3452]: Withdrawing address record for
192.168.4.82 on eth3.
```

```
Jun 11 17:59:24 ol-r5 kernel:  connection3:0: ping timeout of 5 secs expired,
recv timeout 5, last rx 4295174128, last ping 4295179128, now 4295184128
Jun 11 17:59:24 ol-r5 kernel:  connection3:0: detected conn error (1011)
Jun 11 17:59:25 ol-r5 iscsid: Kernel reported iSCSI connection 3:0 error
(1011) state (3)
Jun 11 17:59:25 ol-r5 iscsid: Kernel reported iSCSI connection 3:0 error
(1011) state (3)
Jun 11 18:01:24 ol-r5 kernel:  session3: session recovery timed out after 120
secs
Jun 11 18:01:24 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is down
Jun 11 18:01:24 ol-r5 multipathd: checker failed path 8:32 in map ora_redo
Jun 11 18:01:24 ol-r5 kernel: device-mapper: multipath: Failing path 8:32.
Jun 11 18:01:24 ol-r5 kernel: device-mapper: multipath: Failing path 8:80.
Jun 11 18:01:24 ol-r5 multipathd: ora_redo: remaining active paths: 2
Jun 11 18:01:24 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is down
Jun 11 18:01:24 ol-r5 multipathd: checker failed path 8:80 in map app_space
Jun 11 18:01:24 ol-r5 multipathd: app_space: remaining active paths: 2
Jun 11 18:01:24 ol-r5 multipathd: dm-3: add map (uevent)
Jun 11 18:01:24 ol-r5 multipathd: dm-3: devmap already registered
Jun 11 18:01:24 ol-r5 multipathd: dm-2: add map (uevent)
Jun 11 18:01:24 ol-r5 multipathd: dm-2: devmap already registered
Jun 11 18:01:26 ol-r5 iscsid: connect to 192.168.4.13:3260 failed (No route
to host)
Jun 11 18:01:29 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is down
Jun 11 18:01:29 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is down
Jun 11 18:01:32 ol-r5 iscsid: connect to 192.168.4.13:3260 failed (No route
to host)
Jun 11 18:01:34 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is down
Jun 11 18:01:34 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is down
Jun 11 18:01:38 ol-r5 iscsid: connect to 192.168.4.13:3260 failed (No route
to host)
Jun 11 18:01:39 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is down
Jun 11 18:01:39 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is down
Jun 11 18:01:44 ol-r5 iscsid: connect to 192.168.4.13:3260 failed (No route
to host)
Jun 11 18:01:44 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is down
Jun 11 18:01:44 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is down
Jun 11 18:01:49 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is down
Jun 11 18:01:49 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is down
```

ログ・ファイルの緑で強調表示された部分は、該当する時間にパス障害が発生したことが検出された、iSCSIサブシステムを示しています（/etc/iscsi/iscsid.confに構成）。

DM-Multipathにより、構成されているパスの数が3つから2つに減ったことも検出されています。このことが、ワークロードが発生しているボリュームだけでなく、アイドル状態のボリュームでも検出されていることに注意してください。

DM-Multipathでは、次のmultipath -llコマンド出力で示されているように、障害の発生したパスをエラーとしてマーク付け（緑で強調表示）します。

```
[root@ol-r5 ~]# multipath -ll
app_space (3600144f0c532bda00000512a75600002) dm-3 SUN,ZFS Storage 7420
size=8.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=active
| `- 4:0:0:1 sde 8:64 active ready  running
|-+- policy='round-robin 0' prio=0 status=enabled
| `- 5:0:0:1 sdf 8:80 failed faulty running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 3:0:0:1 sdg 8:96 active ready  running
ora_redo (3600144f0c532bda00000512a754e0001) dm-2 SUN,ZFS Storage 7420
size=4.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 4:0:0:0 sdb 8:16 active ready  running
|-+- policy='round-robin 0' prio=0 status=enabled
| `- 5:0:0:0 sdc 8:32 failed faulty running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 3:0:0:0 sdd 8:48 active ready  running
```

重要なのは、DM-MultipathによってOracle ZFS Storage ApplianceとOracle Linuxホスト間のパスの障害が検出され、このパスのリカバリも実行されることです。次のコード例に示すように、パスが再度使用できるようになると、DM-Multipathによってパスが自動的に再度有効化されます。

```
[root@ol-r5 ~]# ifconfig eth3 up
Jun 11 18:06:49 ol-r5 kernel: e1000: eth3 NIC Link is Up 1000 Mbps Full
Duplex, Flow Control: RX
Jun 11 18:06:49 ol-r5 avahi-daemon[3452]: New relevant interface eth3.IPv4
for mDNS.
Jun 11 18:06:49 ol-r5 avahi-daemon[3452]: Joining mDNS multicast group on
interface eth3.IPv4 with address 192.168.4.82.
Jun 11 18:06:49 ol-r5 avahi-daemon[3452]: Registering new address record for
192.168.4.82 on eth3.
```

```
Jun 11 18:06:50 ol-r5 avahi-daemon[3452]: New relevant interface eth3.IPv6
for mDNS.
Jun 11 18:06:50 ol-r5 avahi-daemon[3452]: Joining mDNS multicast group on
interface eth3.IPv6 with address fe80::a00:27ff:fe96:2c9c.
Jun 11 18:06:50 ol-r5 avahi-daemon[3452]: Registering new address record for
fe80::a00:27ff:fe96:2c9c on eth3.
Jun 11 18:06:50 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is down
Jun 11 18:06:50 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is down
Jun 11 18:06:52 ol-r5 iscsid: connection3:0 is operational after recovery (32
attempts)
Jun 11 18:06:55 ol-r5 multipathd: ora_redo: sdc - readsector0 checker reports
path is up
Jun 11 18:06:55 ol-r5 multipathd: 8:32: reinstated
Jun 11 18:06:55 ol-r5 multipathd: ora_redo: remaining active paths: 3
Jun 11 18:06:55 ol-r5 multipathd: app_space: sdf - readsector0 checker
reports path is up
Jun 11 18:06:55 ol-r5 multipathd: 8:80: reinstated
Jun 11 18:06:55 ol-r5 multipathd: app_space: remaining active paths: 3
Jun 11 18:06:55 ol-r5 multipathd: dm-2: add map (uevent)
Jun 11 18:06:55 ol-r5 multipathd: dm-2: devmap already registered
Jun 11 18:06:55 ol-r5 multipathd: dm-3: add map (uevent)
Jun 11 18:06:55 ol-r5 multipathd: dm-3: devmap already registered
[root@ol-r5 ~]# multipath -ll
app_space (3600144f0c532bda00000512a75600002) dm-3 SUN,ZFS Storage 7420
size=8.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=active
| `- 4:0:0:1 sde 8:64 active ready running
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 5:0:0:1 sdf 8:80 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 3:0:0:1 sdg 8:96 active ready running
ora_redo (3600144f0c532bda00000512a754e0001) dm-2 SUN,ZFS Storage 7420
size=4.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 4:0:0:0 sdb 8:16 active ready running
|-+- policy='round-robin 0' prio=1 status=enabled
| `- 5:0:0:0 sdc 8:32 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
  `- 3:0:0:0 sdd 8:48 active ready running
```

Oracle ZFS Storage Appliance側でパスの障害が発生した場合、Avahiネットワークなどのエラーは Oracle Linuxサーバーではレポートされないため、エラー・メッセージはやや異なります。ただし、これは障害の追跡で貴重な手掛かりとなる可能性があります。次に、障害がこのように発生した場合にOracle Linuxサーバーに表示される、エラー・メッセージを示します。障害は黄色の強調表示の部分で検出されており、この問題の解決は緑の強調表示の行で通知されています。

```
Jun 11 18:11:15 ol-r5 kernel:  connection1:0: ping timeout of 5 secs expired,
recv timeout 5, last rx 4295885428, last ping 4295890428, now 4295895428
Jun 11 18:11:15 ol-r5 kernel:  connection1:0: detected conn error (1011)
Jun 11 18:11:15 ol-r5 iscsid: Kernel reported iSCSI connection 1:0 error
(1011) state (3)
Jun 11 18:11:39 ol-r5 iscsid: connect to 192.168.3.13:3260 failed (No route
to host)
Jun 11 18:12:16 ol-r5 last message repeated 6 times
[root@ol-r5 ~]#  Jun 11 18:13:11 ol-r5 last message repeated 9 times
Jun 11 18:13:15 ol-r5 kernel:  session1: session recovery timed out after 120
secs
Jun 11 18:13:16 ol-r5 multipathd: ora_redo: sdd - readsector0 checker reports
path is down
Jun 11 18:13:16 ol-r5 multipathd: checker failed path 8:48 in map ora_redo
Jun 11 18:13:16 ol-r5 multipathd: ora_redo: remaining active paths: 2
Jun 11 18:13:16 ol-r5 kernel: device-mapper: multipath: Failing path 8:48.
Jun 11 18:13:16 ol-r5 kernel: device-mapper: multipath: Failing path 8:96.
Jun 11 18:13:16 ol-r5 multipathd: app_space: sdg - readsector0 checker
reports path is down
Jun 11 18:13:16 ol-r5 multipathd: checker failed path 8:96 in map app_space
Jun 11 18:13:16 ol-r5 multipathd: app_space: remaining active paths: 2
Jun 11 18:13:17 ol-r5 iscsid: connect to 192.168.3.13:3260 failed (No route
to host)
Jun 11 18:13:17 ol-r5 multipathd: dm-2: add map (uevent)
Jun 11 18:13:17 ol-r5 multipathd: dm-2: devmap already registered
Jun 11 18:13:17 ol-r5 multipathd: dm-3: add map (uevent)
Jun 11 18:13:17 ol-r5 multipathd: dm-3: devmap already registered
Jun 11 18:13:21 ol-r5 multipathd: ora_redo: sdd - readsector0 checker reports
path is down
Jun 11 18:13:21 ol-r5 multipathd: app_space: sdg - readsector0 checker
reports path is down
Jun 11 18:13:23 ol-r5 iscsid: connect to 192.168.3.13:3260 failed (No route
to host)
Jun 11 18:13:26 ol-r5 multipathd: ora_redo: sdd - readsector0 checker reports
path is down
Jun 11 18:13:26 ol-r5 multipathd: app_space: sdg - readsector0 checker
reports path is down
Jun 11 18:13:29 ol-r5 iscsid: connect to 192.168.3.13:3260 failed (No route
to host)
Jun 11 18:13:31 ol-r5 multipathd: ora_redo: sdd - readsector0 checker reports
path is down
Jun 11 18:13:31 ol-r5 multipathd: app_space: sdg - readsector0 checker
reports path is down
Jun 11 18:13:35 ol-r5 iscsid: connect to 192.168.3.13:3260 failed (No route
to host)
Jun 11 18:13:36 ol-r5 multipathd: ora_redo: sdd - readsector0 checker reports
path is down
```

```
Jun 11 18:13:36 ol-r5 multipathd: app_space: sdg - readsector0 checker
reports path is down
Jun 11 18:13:41 ol-r5 iscsid: connect to 192.168.3.13:3260 failed (No route
to host)
Jun 11 18:13:41 ol-r5 multipathd: ora_redo: sdd - readsector0 checker reports
path is down
Jun 11 18:13:41 ol-r5 multipathd: app_space: sdg - readsector0 checker
reports path is down
Jun 11 18:13:45 ol-r5 iscsid: connection1:0 is operational after recovery (23
attempts)
Jun 11 18:13:46 ol-r5 multipathd: ora_redo: sdd - readsector0 checker reports
path is up
Jun 11 18:13:46 ol-r5 multipathd: 8:48: reinstated
Jun 11 18:13:46 ol-r5 multipathd: ora_redo: remaining active paths: 3
Jun 11 18:13:46 ol-r5 multipathd: app_space: sdg - readsector0 checker
reports path is up
Jun 11 18:13:46 ol-r5 multipathd: 8:96: reinstated
Jun 11 18:13:46 ol-r5 multipathd: app_space: remaining active paths: 3
Jun 11 18:13:47 ol-r5 multipathd: dm-2: add map (uevent)
Jun 11 18:13:47 ol-r5 multipathd: dm-2: devmap already registered
Jun 11 18:13:47 ol-r5 multipathd: dm-3: add map (uevent)
Jun 11 18:13:47 ol-r5 multipathd: dm-3: devmap already registered
```

デバイスの所有権

特定の状況では、たとえば、データベースなどのアプリケーションやユーザー領域のファイル・システムに、ファイル・システム構造が不要なストレージのRAWブロックへのアクセスが必要な場合、ボリュームの所有権をデフォルト以外のユーザーに付与することが重要になります。

Oracleシステムの場合、これらの2つの例はOracle Automatic Storage Managementの形式でマージされ、ブロック・ストレージおよびスケーラブルでクラスタ化可能なファイル・システム同様の構造を使用して、Oracle環境内で効率的かつ管理可能な形式でOracle環境用に統合されます。

Oracle ASMには、root以外のユーザーに割り当てられたブロック・ストレージを表しているデバイス・ノードの所有権が必要です。従来どおり、これが"oracle"ユーザーとなります。デバイス・ノードのグループに関しても、役立つ推奨事項がいくつかあります。

Oracle Linux Release 5

Oracle Linux Release 5では、DM-Multipathで所有権、およびDM-Multipathで作成したデバイスの保護モードを変更できます。ファイルの'uid'、'gid、または'mode'を指定すると、必要に応じてroot以外のユーザーにストレージへのアクセスを許可するように、デバイスを容易に作成できます。ユーザーとグループのシンボリック名または整数をuidフィールドとgidフィールドに使用し、標準の8進数のアクセス権表示をmodeフィールドに使用します。

/etc/multipath.conf構成ファイルの変更は、次のように行います。

```
 multipaths {
         multipath {
                wwid           3600144f0c532bda00000512a754e0001
                alias          ora_redo
               uid          oracle
               gid          orainst
               mode         660
         }

         multipath {
                 wwid            3600144f0c532bda00000512a75600002
                 alias           app_space
         }
 }
```

Oracle Linux Release 6

Oracle Linux Release 5では簡単な方法でDM-Multipath構成ファイルの所有権を指定できますが、Oracle Linux Release 6ではこの方法は使用できません。

Oracle Linux Release 6では、Oracle Linuxの事実上のデバイス・マネージャであるUDEV機能を使用する方法があります。この機能は、Linuxのhotplugとdevfsの後継ですが、ユーザー領域で機能します（名前に'U'が含まれているのは、このためです）。

UDEVは、サービスを必要としているハードウェア・イベントが通知される、デーモン・ユーザー・プロセスで構成されています。デーモンは、ハードウェア・イベントの必要なシンボリック表示を作成するために使用される、数多くのルールを処理します。この例として、USBフラッシュ・ドライブの挿入が挙げられます。ブロック・ストレージのデバイス・ノードが自動的に作成されて、マウントされたファイル・システム内のフラッシュ・ドライブのコンテンツが表示されます。

ルールは強力であり、厳密な構文が使用されます。UDEVルールの記述とデバッグに焦点をあてた、数多くの役立つWebサイトがあります。

所有権を決定する簡単な方法として、プリフィックスや他のネーミング規則に基づいてLUNのアクセス権を付与するユーザーを決定する方法があります。

これまでの例では、ボリューム`ora_redo`および`app_space`を使用しています。前述のとおり、Oracle ASMでは、Oracle ASM構造をホストするために使用するLUNをroot以外のユーザーが書き込める必要があります。次の例では、ユーザーは`oracle`となり、デバイス・ノードをこのユーザーおよびグループ`orainst`のすべてのメンバーが書き込める必要があります。

UDEVルール・ファイルが作成されて、起動時などにデバイスが作成または変更されると、該当するデバイスの所有権が変更されます。

ファイル99-diskownership.rulesは/etc/udev/rules.dに作成され、次のUDEVルールが含まれます。

```
SUBSYSTEM!="block",  GOTO="quickexit"
KERNEL!="dm-[0-9]*", GOTO="quickexit"
PROGRAM=="/sbin/dmsetup info -c --noheadings -o name -m %m -j %M"
RESULT=="ora_*", OWNER="oracle", GROUP="dba", MODE="0660"
RESULT=="app_*", OWNER="geosurvey", GROUP="geology", MODE="0660"
LABEL="quickexit
```

UDEVでは、さまざまなデバイス・タイプの大規模な配列が処理されるため、必要となる場合を除き、このスクリプトが実行から除外されていることを確認することが重要です。イベントを除外するために、このスクリプトでは、ブロック・ストレージ1以外のカーネル・サブシステムの処理が試みられないことを確認し、処理が試みられた場合は、即座にルール・ファイルの最後に移動して終了します。

次に、スクリプトでは、処理対象のデバイス名の先頭がdm-x（ここで、xは数字）以外であることを確認し、dm-xである場合は、即座にルール・ファイルの最後に移動します。

実行が3行目まで続行された場合、先頭がdm-xのブロック・デバイスが処理されるため、DM-Multipathの構成ファイルで割り当てられている別名を検出する必要があります。プログラムの出力は、空の文字列（別名が割り当てられていない場合）になるか、このマルチパス化されたデバイスの適切に割り当てられた別名になります。

使用するネーミング規則によって、すべてのOracle ASMボリュームの先頭に文字列ora_が追加される場合、この条件と一致するすべてのデバイスをユーザー'oracle'とグループ'dba'が所有し、セキュリティ・モード'rw-rw----'を指定する必要があることを推測できます。

この最後の3つの属性は、適切な値を割り当てることで、UDEVの宣言'OWNER'、'GROUP'、および'MODE'によって処理されます。

このスクリプトのチェック・プロセスは、ボリューム名の先頭に適切な文字列を追加するだけで、さまざまなアプリケーション/所有権の要件に対応するように拡張できます。

前述の例では、先頭に文字列'app_'が付いたすべてのボリュームのエイリアスは、ユーザー'geosurvey'およびグループ'geology'の所有権に自動的に変更されます。ブロック・ストレージへのアクセスが必要な地質解析パッケージに、これを適用できます。

結論

Oracle LinuxとOracle ZFS Storage Applianceは、高可用性環境で適切に連携して動作するように設計されており、ストレージへの冗長接続を効率的かつ効果的に使用できます。

データ・アクセスの継続性と一貫したボリュームのネーミングにより、ビジネス・アプリケーションを構築する、安定性に優れたプラットフォームを実現し、変化する環境でOracle ZFS Storage Applianceによって引き続きビジネス・ニーズを満たすことができます。

参考資料

Oracle ZFS Storage Applianceドキュメント	http://www.oracle.com/technetwork/jp/server-storage/sun-unified-storage/documentation/index.html
Oracle Linux製品ページ	http://www.oracle.com/jp/technologies/linux/overview/index.html
Oracle ZFS Storage Appliance製品ページ	http://www.oracle.com/jp/products/servers-storage/storage/nas/overview/index.html

Oracle LinuxとOracle ZFS Storage Applianceの
マルチパスの構成
2013年10月、バージョン1.0
著者：Andrew Ness

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

お問い合わせ窓口

Oracle Direct

TEL 0120-155-096
URL oracle.com/jp/direct

Oracle is committed to developing practices and products that help protect the environment

Copyright © 2013, Oracle and/or its affiliates.All rights reserved.
本文書は情報提供のみを目的として提供されており、ここに記載される内容は予告なく変更されることがあります。本文書は一切間違いがないことを保証するものではなく、さらに、口述による明示または法律による黙示を問わず、特定の目的に対する商品性もしくは適合性についての黙示的な保証を含み、いかなる他の保証や条件も提供するものではありません。オラクル社は本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクル社の書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。

OracleおよびJavaはOracleおよびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。

IntelおよびIntel Xeonは Intel Corporationの商標または登録商標です。すべてのSPARC商標はライセンスに基づいて使用されるSPARC International, Inc.の商標または登録商標です。AMD、Opteron、AMDロゴおよびAMD Opteronロゴは、Advanced Micro Devicesの商標または登録商標です。UNIXはX/Open Company, Ltd.によってライセンス提供された登録商標です。0611

Hardware and Software, Engineered to Work Together