

2007年度
オープンソースソフトウェア活用基盤整備事業
第I期 テーマ型（調査）

「Linux ディスク冗長化機能の適用評価と
最適な適用方法の調査」
ー ソフトウェア RAID 設定手順書 ー

2008年 1月

独立行政法人 情報処理推進機構

目次

1	目的と概要	1
2	mdadm	2
2.1	設定	2
2.1.1	RAID アレイ作成	2
2.1.2	RAID アレイ削除	5
2.1.3	RAID アレイへのディスクの追加	6
2.1.4	RAID アレイからのディスクの削除	8
2.1.5	ディスクへの不良マーク付与	9
2.1.6	RAID アレイのスーパーブロックの操作	10
2.1.7	RAID アレイの書き込み制限	11
2.1.8	RAID アレイ上でのイベント検出	12
2.1.9	RAID アレイのチェック	12
2.2	復旧	13
2.2.1	スペアディスク切り替え	13
2.2.2	進捗状況の確認	14
2.2.3	hotplug 機能	14
2.2.4	リブート時の動作	14
2.3	管理	15
2.3.1	故障ディスクの特定	15
2.3.2	RAID 構成情報のバックアップ・リストア	15
3	dmraid	17
3.1	設定	17
3.1.1	RAID アレイ作成	17
3.1.2	RAID アレイ削除	19
3.1.3	RAID アレイへのディスクの追加	19
3.1.4	RAID アレイからのディスクの削除	19
3.1.5	デバイスへの不良マーク付与	19
3.1.6	RAID アレイのスーパーブロックの操作	19
3.1.7	RAID アレイの書き込み制限	19
3.1.8	RAID アレイ上でのイベント検出	19
3.1.9	RAID アレイのチェック	19
3.2	復旧	20
3.2.1	スペアディスクの切り替え	20
3.2.2	進捗状況の確認	20

3.2.3	hotplug 機能	20
3.2.4	リブート時の動作	20
3.3	管理	20
3.3.1	故障ディスクの特定	20
3.3.2	RAID 構成情報のバックアップ・リストア	20
4	LVM2	22
4.1	設定	22
4.1.1	RAID アレイ作成	22
4.1.2	RAID アレイ削除	24
4.1.3	RAID アレイへのディスクの追加	24
4.1.4	RAID アレイからのディスクの削除	24
4.1.5	デバイスへの不良マーク付与	24
4.1.6	RAID アレイのスーパーブロックの操作	24
4.1.7	RAID アレイの書き込み制限	25
4.1.8	RAID アレイ上でのイベント検出	26
4.1.9	RAID アレイのチェック	26
4.2	復旧	27
4.2.1	スペアディスクの切り替え	27
4.2.2	進捗状況の確認	27
4.2.3	hotplug 機能	27
4.2.4	リブート時の動作	27
4.3	管理	27
4.3.1	故障ディスクの特定	27
4.3.2	RAID 構成情報のバックアップ・リストア	29
5	付録 構成情報	30

1 目的と概要

複数のハードディスクをまとめて一台のハードディスクとして管理し、データを分散して記憶することで、高速化や安全性の向上を図る技術を RAID という。RAID をソフトウェアで実現したものをソフトウェア RAID という。

本書は Linux 上でソフトウェア RAID を構築・管理するための手順書である。

Linux 上でソフトウェア RAID の設定を行うツールには mdadm、dmraid、LVM2 がある。

- mdadm では、RAID1、RAID5、RAID6、RAID10 を、dmraid および LVM2 では、RAID1 を構築することができる。
- mdadm ではスペアデバイスを利用した自動復旧が可能であるが、dmraid および LVM2 では自動復旧はサポートされていない。
- dmraid では、ハードウェアが dmraid のサポートしている ATARAID タイプに対応している必要がある。(詳細に関しては「3 dmraid」の項を参照)

本書ではソフトウェア RAID を設定・復旧・管理する手順をツールごとに説明する。

実行例の見方

コマンド入力は下線で表され、他は出力結果である。

```
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sde1[3] sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/4] [UUUU]

unused devices: <none>
```

2 mdadm

mdadm は md (Multiple Devices) デバイスドライバを通してソフトウェア RAID を実現するツールである。以下に設定・復旧・管理手順について記述する。mdadm では、RAID1、RAID5、RAID6、RAID10 を構築することができる。

2.1 設定

2.1.1 RAID アレイ作成

次のコマンドを実行し、RAID アレイを作成する。

```
/sbin/mdadm -C <作成するRAIDアレイ> -l <RAIDレベル> -n <アクティブデバイス数> -x <スペアデバイス数> <デバイス群>
```

<作成する RAID アレイ>には慣習的に/dev/md0 や/dev/md1 等を指定する。

<RAID レベル>には構築したい RAID レベルにしたがって次のように設定する。

- raid0, 0, stripe RAID0 (ストライピング)
- raid1, 1, mirror RAID1 (ミラーリング)
- raid4, 4 RAID4
- raid5, 5 RAID5
- raid6, 6 RAID6
- raid10, 10 RAID10
- multipath, mp マルチパス RAID
- linear リニア RAID

<アクティブデバイス数>には RAID アレイ内でアクティブなディスクの数を指定する。

<スペアデバイス数>にはスペアディスクの数を指定する。スペアディスクとは、ディスク故障の発生等により縮退した RAID アレイをディスク切り替えによって自動復旧するための余剰ディスクである。

<デバイス群>には使用するデバイス(名)群を指定する。記述方法として正規表現も可能である。なお、アクティブディスクとなるか、スペアディスクとなるかの切り分けは、mdadm が自動的に決定する。

次のコマンドで作成した RAID アレイを確認できる。

```
# cat /proc/mdstat
```

また、RAID アレイ作成時に/var/log/messages にログが出力される。

実行例 1

4つのデバイス/dev/sdb1, /dev/sdc1, /dev/sdd1, /dev/sde1 で構成された RAID1 アレイ /dev/md0 を作成する。

```
# /sbin/mdadm -C /dev/md0 -l 1 -n 4 /dev/sd[bcde]1
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sde1[3] sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/4] [UUUU]

unused devices: <none>
```

/var/log/messages

```
Sep 19 18:51:32 XXX kernel: md: bind<sdb1>
Sep 19 18:51:32 XXX kernel: md: bind<sdc1>
Sep 19 18:51:32 XXX kernel: md: bind<sdd1>
Sep 19 18:51:32 XXX kernel: md: bind<sde1>
Sep 19 18:51:32 XXX kernel: md: md0: raid array is not clean -- starting backg
round reconstruction
Sep 19 18:51:32 XXX kernel: md: raid1 personality registered for level 1
Sep 19 18:51:32 XXX kernel: raid1: raid set md0 active with 4 out of 4 mirrors
Sep 19 18:51:32 XXX kernel: md: syncing RAID array md0
Sep 19 18:51:32 XXX kernel: md: minimum _guaranteed_ reconstruction speed: 100
0 KB/sec/disc.
Sep 19 18:51:32 XXX kernel: md: using maximum available idle IO bandwidth (but
not more than 200000 KB/sec) for reconstruction.
Sep 19 18:51:32 XXX kernel: md: using 128k window, over a total of 4016128 blo
cks.
Sep 19 19:18:20 XXX kernel: md: md0: sync done.
Sep 19 19:18:20 XXX kernel: RAID1 conf printout:
Sep 19 19:18:20 XXX kernel: --- wd:4 rd:4
Sep 19 19:18:20 XXX kernel: disk 0, wo:0, o:1, dev:sdb1
Sep 19 19:18:20 XXX kernel: disk 1, wo:0, o:1, dev:sdc1
Sep 19 19:18:20 XXX kernel: disk 2, wo:0, o:1, dev:sdd1
Sep 19 19:18:20 XXX kernel: disk 3, wo:0, o:1, dev:sde1
```

実行例 2

実行例 1 と同じ構成にスペアディスク/dev/sdf1 を追加し、RAID アレイ/dev/md0 を作成する。

```
# /sbin/mdadm -C /dev/md0 -l 1 -n 4 -x 1 /dev/sd[bcdef]1
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sdf1[4] (S) sde1[3] sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/4] [UUUU]

unused devices: <none>
```

/var/log/messages

```
Sep 19 18:51:32 XXX kernel: md: bind<sdb1>
Sep 19 18:51:32 XXX kernel: md: bind<sdc1>
Sep 19 18:51:32 XXX kernel: md: bind<sdd1>
Sep 19 18:51:32 XXX kernel: md: bind<sde1>
Sep 19 18:51:32 XXX kernel: md: bind<sdf1>
Sep 19 18:51:32 XXX kernel: md: md0: raid array is not clean -- starting background reconstruction
Sep 19 18:51:32 XXX kernel: md: raid1 personality registered for level 1
Sep 19 18:51:32 XXX kernel: raid1: raid set md0 active with 4 out of 4 mirrors
Sep 19 18:51:32 XXX kernel: md: syncing RAID array md0
Sep 19 18:51:32 XXX kernel: md: minimum _guaranteed_ reconstruction speed: 1000 KB/sec/disc.
Sep 19 18:51:32 XXX kernel: md: using maximum available idle IO bandwidth (but not more than 200000 KB/sec) for reconstruction.
Sep 19 18:51:32 XXX kernel: md: using 128k window, over a total of 4016128 blocks.
Sep 19 19:18:20 XXX kernel: md: md0: sync done.
Sep 19 19:18:20 XXX kernel: RAID1 conf printout:
Sep 19 19:18:20 XXX kernel: --- wd:5 rd:4
Sep 19 19:18:20 XXX kernel: disk 0, wo:0, o:1, dev:sdb1
Sep 19 19:18:20 XXX kernel: disk 1, wo:0, o:1, dev:sdc1
Sep 19 19:18:20 XXX kernel: disk 2, wo:0, o:1, dev:sdd1
Sep 19 19:18:20 XXX kernel: disk 3, wo:0, o:1, dev:sde1
Sep 19 19:18:20 XXX kernel: disk 4, wo:0, o:1, dev:sdf1
```

2.1.2 RAID アレイ削除

次のコマンドで RAID アレイを削除する。

```
/sbin/mdadm -S <削除するRAIDアレイ>
```

RAID アレイを削除しただけでは”mdadm -A~”で RAID アレイが再構築されてしまう。そのため、物理ディスクを完全に RAID から切り離すためには、以下のようにスーパーブロックを 0 で初期化する必要がある。

```
/sbin/mdadm --zero-superblock <対象となるRAIDアレイ>
```

実行例

RAID アレイ/dev/md0 を削除し、スーパーブロックを 0 で初期化する。

```
# /sbin/mdadm -S /dev/md0
# /sbin/mdadm --zero-superblock /dev/md0
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
unused devices: <none>
```

/var/log/messages

```
Sep 19 19:06:21 XXX kernel: md: md0 stopped.
Sep 19 19:06:21 XXX kernel: md: unbind<sde1>
Sep 19 19:06:21 XXX kernel: md: export_rdev(sde1)
Sep 19 19:06:21 XXX kernel: md: unbind<sdd1>
Sep 19 19:06:21 XXX kernel: md: export_rdev(sdd1)
Sep 19 19:06:21 XXX kernel: md: unbind<sd1>
Sep 19 19:06:21 XXX kernel: md: export_rdev(sd1)
Sep 19 19:06:21 XXX kernel: md: unbind<sdb1>
Sep 19 19:06:21 XXX kernel: md: export_rdev(sdb1)
```

2.1.3 RAID アレイへのディスクの追加

次のコマンドで RAID アレイに対してディスクを追加する。

```
/sbin/mdadm <追加対象のRAIDアレイ> -a <追加するディスク名>
```

実行例 1

縮退していない RAID アレイ/dev/md0 に対してディスク/dev/sdf1 を追加する。

```
# /sbin/mdadm /dev/md0 -a /dev/sdf1
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sdf1[4] (S) sde1[3] sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/4] [UUUU]

unused devices: <none>
```

/var/log/messages

```
Sep 19 18:51:32 XXX kernel: md: bind<sdf1>
```

実行例 2

縮退している RAID アレイ/dev/md0 に対してディスク/dev/sdf1 を追加する。

```
# /sbin/mdadm /dev/md0 -a /dev/sdf1
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sdf1[3] sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/3] [UUU_]
      [>.....] recovery = 0.3% (12992/4016128) finish=5.1min
      speed=12992K/sec

unused devices: <none>
```

/var/log/messages

```
Sep 19 18:51:32 XXX kernel: md: bind<sdf1>
Sep 19 19:18:20 XXX kernel: RAID1 conf printout:
Sep 19 19:18:20 XXX kernel: --- wd:3 rd:4
Sep 19 19:18:20 XXX kernel: disk 0, wo:0, o:1, dev:sdb1
Sep 19 19:18:20 XXX kernel: disk 1, wo:0, o:1, dev:sdcl
Sep 19 19:18:20 XXX kernel: disk 2, wo:0, o:1, dev:sdd1
Sep 19 19:18:20 XXX kernel: disk 3, wo:1, o:1, dev:sdf1
Sep 19 18:51:32 XXX kernel: md: minimum _guaranteed_ reconstruction speed: 100
0 KB/sec/disc.
Sep 19 18:51:32 XXX kernel: md: using maximum available idle IO bandwidth (but
not more than 200000 KB/sec) for reconstruction.
Sep 19 18:51:32 XXX kernel: md: using 128k window, over a total of 4016128 blo
cks.
Sep 19 19:18:20 XXX kernel: md: md0: sync done.
Sep 19 19:18:20 XXX kernel: RAID1 conf printout:
Sep 19 19:18:20 XXX kernel: --- wd:4 rd:4
Sep 19 19:18:20 XXX kernel: disk 0, wo:0, o:1, dev:sdb1
Sep 19 19:18:20 XXX kernel: disk 1, wo:0, o:1, dev:sdcl
Sep 19 19:18:20 XXX kernel: disk 2, wo:0, o:1, dev:sdd1
Sep 19 19:18:20 XXX kernel: disk 3, wo:0, o:1, dev:sdf1
```

2.1.4 RAID アレイからのディスクの削除

次のコマンドで RAID アレイからディスクを削除する。削除できるディスクは不良マークがついているかスペアディスクである必要がある。

```
/sbin/mdadm <削除対象のRAIDアレイ> -r <削除するディスク名>
```

RAID アレイからディスクが削除されていることは次のコマンドで確認できる。

```
/sbin/mdadm -D <削除対象のRAIDアレイ>
```

実行例

RAID アレイ/dev/md0 から/dev/sde1 ディスクを削除し、削除されていることを確認している。

```
# /sbin/mdadm /dev/md0 -r /dev/sde1
# /sbin/mdadm -D /dev/md0
/dev/md0:
  Version : 00.90.03
  Creation Time : Wed Sep 19 19:28:02 2007
  Raid Level : raid1
  Array Size : 4016128 (3.83 GiB 4.11 GB)
  Device Size : 4016128 (3.83 GiB 4.11 GB)
  Raid Devices : 4
  Total Devices : 3
  Preferred Minor : 0
  Persistence : Superblock is persistent

  Update Time : Wed Sep 19 20:19:31 2007
  State : clean, degraded
  Active Devices : 3
  Working Devices : 3
  Failed Devices : 0
  Spare Devices : 0

  UUID : 32b9c2f4:c2c11b11:37143974:5e75140f
  Events : 0.16

  Number   Major   Minor   RaidDevice State
  0         8       17     0         active sync  /dev/sdb1
  1         8       33     1         active sync  /dev/sdc1
  2         8       49     2         active sync  /dev/sdd1
  3         0       0      1         removed
```

/var/log/messages

```
Sep 19 19:06:21 XXX kernel: md: unbind<sde1>
Sep 19 19:06:21 XXX kernel: md: export_rdev(sde1)
```

2.1.5 ディスクへの不良マーク付与

次のコマンドでディスクに不良マークを付与する。これにより、擬似的なディスク故障からの復旧のテストができる。

```
/sbin/mdadm <対象となるRAIDアレイ> -f <不良マークを付与するディスク>
```

実行例

スペアディスクを含まない RAID アレイ/dev/md0 において、アクティブなディスク/dev/sde1 に不良マークを付与する。RAID アレイ/dev/md0 は縮退する。

```
# /sbin/mdadm /dev/md0 -f /dev/sde1
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sde1[3] (F) sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/3] [UUU_]

unused devices: <none>
```

スペアディスクを含む RAID アレイ/dev/md1 において、アクティブなディスク/dev/sde1 に不良マークを付与すると、スペアディスクがアクティブに切り替わり自動で復旧処理が行われる。詳細は「2.2.1 スペアディスク切り替え」を参照。

2.1.6 RAID アレイのスーパーブロックの操作

(1) スーパーブロックの内容を表示

次のコマンドで指定したデバイスの md に関するスーパーブロックの内容を表示する。

```
/sbin/mdadm -E <表示するデバイス名>
```

実行例

ディスク/dev/sde1 の md に関するスーパーブロックの内容を表示する。

```
# /sbin/mdadm -E /dev/sde1
/dev/sde1:
    Magic : a92b4efc
    Version : 00.90.03
    UUID : 5a194984:f53f1370:87f7da55:86dc54d2
    Creation Time : Sat Sep 15 18:33:03 2007
    Raid Level : raid1
    Raid Devices : 4
    Total Devices : 5
    Preferred Minor : 0

    Update Time : Sat Sep 15 19:16:24 2007
    State : clean
    Active Devices : 4
    Working Devices : 5
    Failed Devices : 0
    Spare Devices : 1
    Checksum : 974e43b7 - correct
    Events : 0.2440

    Layout : left-symmetric
    Chunk Size : 64K

    Number Major Minor RaidDevice State
this      3      8      65      3      active sync  /dev/sdc1
0         0      8      17      0      active sync  /dev/sdb1
1         1      8      33      1      active sync  /dev/sdc1
2         2      8      49      2      active sync  /dev/sdd1
3         3      8      65      3      active sync  /dev/sde1
4         4      8      81      4      spare  /dev/sdf1
```

(2) スーパーブロックの内容を更新

次のコマンドで RAID アレイのスーパーブロックの内容を更新する。

```
/sbin/mdadm -A <対象となるRAIDアレイ> -U=<引数>
```

<引数>には `sparc2.2`、`summaries`、`super-minor` のいずれかを指定する。`sparc2.2` の場合、RAID パッチが施された `Liux2.2` における `Sparc` マシン用の形式にアレイのスーパーブロックを変換する。`summaries` の場合、アレイのスーパーブロック内のサマリ情報を訂正する。`super-minor` の場合、各デバイスのスーパーブロックのマイナー番号を編成されるアレイのマイナー番号に変更する。

2.1.7 RAID アレイの書き込み制限

(1) RAID アレイの書き込み制限

次のコマンドで RAID アレイに対して書き込み制限を行う。

```
/sbin/mdadm -o <対象となるRAIDアレイ>
```

実行例

RAID アレイ/`dev/md0` に対して書き込み制限を行う。RAID アレイ/`dev/md0` は読み取り専用になる。

```
# /sbin/mdadm -o /dev/md0
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active (read-only) raid1 sde1[3] sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/4] [UUUU]

unused devices: <none>
```

(2) RAID アレイの書き込みの書き込み制限の解除

次のコマンドで RAID アレイの書き込み制限の解除を行う。

```
/sbin/mdadm -w <対象となるRAIDアレイ>
```

実行例

RAID アレイの書き込み制限の解除を行う。RAID アレイ/`dev/md0` は読み書き可能となる。

```
# /sbin/mdadm -w /dev/md0
```

2.1.8 RAID アレイ上でのイベント検出

次のコマンドで RAID アレイ上でのイベント検出時にメールで通知する機能を使用する。なお、sendmail サービスが起動していなければならない。

```
/sbin/mdadm -F -m=<メールアドレス>
```

次のコマンドで RAID アレイ上でのイベント検出時に特定のプログラムを実施する機能を使用する。

```
/sbin/mdadm --monitor -p <プログラム>
```

イベントの種類は次のとおり

DeviceDisappeared	RAID アレイの削除
RebuildStarted	再構築開始
RebuildNN	再構築 NN (20,40,60,80) %完了
Fail	アクティブなディスクへの不良マーク付与
FailSpare	スペアディスクへの不良マーク付与
SpareActive	再構築によりアクティブへの切り替え完了
NewArray	RAID アレイの追加
DegradedArray	RAID アレイの縮退
MoveSpare	スペアグループ内でのスペアディスク切り替え
SpareMissing	RAID アレイ内のスペアディスクの消失
TestMessage	テストメッセージ

メール送信の対象となるイベントは、Fail、FailSpare、DegradedArray、TestMessage の 4 つである。

2.1.9 RAID アレイのチェック

ディスクの内容同士で整合性があるかチェックする機能は、md では未サポートである。

2.2 復旧

2.2.1 スペアディスク切り替え

RAID アレイ中にスペアディスクが存在している場合、ディスク故障の発生等によりアクティブなディスクに対して不良マークが付与されると、スペアディスクへの切り替えが自動で行われる。

RAID アレイ中にスペアディスクが存在しない場合は、縮退状態の RAID アレイにディスクを追加することによって手動で復旧することができる。

実行例 1

スペアディスク/dev/sdf1 を含む RAID アレイに対して/dev/sde1 に不良マークを付与し、自動で復旧させる。

```
# /sbin/mdadm /dev/md0 -f /dev/sde1
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sdf1[4] sde1[3] (F) sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/3] [UUU_]
      [>.....] recovery = 0.3% (12992/4016128) finish=5.1min
      speed=12992K/sec

unused devices: <none>
```

実行例 2

スペアディスクを含まない RAID アレイに対して/dev/sde1 に不良マークを付与し、手動でディスク/dev/sdf1 を追加して復旧させる。

```
# /sbin/mdadm /dev/md0 -a /dev/sdf1
```

2.2.2 進捗状況の確認

復旧処理の進捗状況は次のコマンドで確認できる。

```
cat /proc/mdstat
```

実行例

復旧処理中の場合、md デバイスの復旧処理の進捗状況を確認できる。

```
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sdf1[4] sde1[3] (F) sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/3] [UUU_]
      [>.....] recovery = 0.3% (12992/4016128) finish=5.1min
      speed=12992K/sec

unused devices: <none>
```

2.2.3 hotplug 機能

hotplug 機能による自動復旧は未サポートである。

縮退状態の RAID アレイにディスクを追加することによって手動で復旧することができる。

実行例

物理的にディスク/dev/sdf1 を挿入し、縮退状態の RAID アレイ/dev/md0 に対してディスク/dev/sdf1 を手動で追加する。

```
# /sbin/mdadm /dev/md0 -a /dev/sdf1
```

2.2.4 リブート時の動作

縮退状態で再起動を行った場合、/etc/mdadm.conf に構成情報がバックアップされていれば、縮退状態のまま起動する。

復旧処理中にリブートを実施しようとした場合、/etc/mdadm.conf に構成情報がバックアップされていれば、復旧処理が中断され、リブートが実施される。リブート後、復旧処理の最初から開始される。

/etc/mdadm.conf に構成情報がバックアップされていなければ、RAID アレイは停止状態である。構成情報のバックアップについては「2.3.2 RAID 構成情報のバックアップ・リストア」を参照。

2.3 管理

2.3.1 故障ディスクの特定

ディスクが故障した場合、次のコマンドで故障ディスクの特定ができる。

```
cat /proc/mdstat
```

実行例

/dev/sde1 が故障している場合、(F)が付与されていることから故障ディスクを特定できる。

```
# cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sde1[3] (F) sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/3] [UUU_]

unused devices: <none>
```

2.3.2 RAID 構成情報のバックアップ・リストア

(1) RAID 構成情報のバックアップ

次のコマンドで RAID 構成情報のバックアップができる。

```
echo 'DEVICE <バックアップするRAIDアレイ中のデバイス群>' > /etc/mdadm.conf
/sbin/mdadm --detail --scan >> /etc/mdadm.conf
```

実行例

RAID アレイ/dev/md0 を構成するデバイスが/dev/sd[bcde]1 であるとした場合、次のコマンドを実行することで RAID 構成情報のバックアップを行う。

```
# echo 'DEVICE /dev/sd[bcde]1' > /etc/mdadm.conf
# /sbin/mdadm --detail --scan >> /etc/mdadm.conf
DEVICE /dev/sd[bcde]1
ARRAY /dev/md0 level=raid1 num-devices=4 UUID=218084e6:394d12a6:e60478ee:42f47f12
```

(2) RAID 構成情報のリストア

ディスクに RAID アレイ構成時の情報が残っていれば、次のコマンドでリストアすることができる。

```
/sbin/mdadm -A <対象となるRAIDアレイ>
```

実行例

md0 をリストアする。

```
# /sbin/mdadm -A /dev/md0
```

リストア後の構成例

```
Personalities : [raid6] [raid5] [raid4] [raid1]
md0 : active raid1 sde1[3] sdd1[2] sdc1[1] sdb1[0]
      4016128 blocks [4/4] [UUUU]

unused devices: <none>
```

3 dmraid

dmraid は Device-mapper を備えた Linux 上でベンダ固有のドライバをインストールせずに RAID デバイスをサポートするツールである。以下に設定・復旧・管理手順について記述する。dmraid では、RAID1 を構築することができる。

3.1 設定

3.1.1 RAID アレイ作成

RAID アレイを作成する手順はハードウェアでサポートされている ATARAID タイプによって異なる。dmraid では、以下の ATARAID タイプがサポートされており、各々独自のツールで RAID アレイを作成する必要がある。

- Highpoint HPT37X
- Highpoint HPT45X
- Intel Software RAID
- LSI Logic MegaRAID
- NVidai NForce
- Promise FastTrack
- Silicon Image Medley
- VIA Software RAID

例えば、Intel Software RAID では、BIOS から RAID のサポートを有効にすることでマシン起動時に”Press <Ctrl - I> to enter the RAID Configuration Utility.”と表示されるようになり、指示に従うと Intel Matrix Storage Manager という RAID の設定が行えるツールが起動する。

RAID 作成後、次のコマンドでデバイスが認識できているかどうか確認する。

```
dmraid -r [<対象となるデバイス>]
```

また、次のコマンドで、RAID アレイが正しく作成できたことを確認する。

```
dmraid -s [<対象となるRAIDアレイ>]  
または  
/sbin/dmsetup table [<対象となるデバイス>]
```

実行例 1

RAID 作成後、デバイスが認識できているかどうかを確認する。

```
# dmraid -r
/dev/sdc: isw, "isw_dcjeehgfed", GROUP, ok, 156250078 sectors, data@ 0
/dev/sdd: isw, "isw_dcjeehgfed", GROUP, ok, 156301486 sectors, data@ 0
/dev/sde: isw, "isw_cdafhccbeb", GROUP, ok, 488397166 sectors, data@ 0
/dev/sdf: isw, "isw_cdafhccbeb", GROUP, ok, 488397166 sectors, data@ 0
```

実行例 2

RAID アレイが正しく作成できたことを確認する。

```
# dmraid -s
*** Group superset isw_dcjeehgfed
--> Active Subset
name   : isw_dcjeehgfed_Volume0
size   : 156243968
stride : 128
type   : mirror
status : ok
subsets: 0
devs   : 2
spares : 0
*** Group superset isw_cdafhccbeb
--> Active Subset
name   : isw_cdafhccbeb_Volume1
size   : 488390656
stride : 128
type   : mirror
status : ok
subsets: 0
devs   : 2
spares : 0
```

実行例 3

RAID アレイが正しく作成できたことを確認する。

```
# /sbin/dmsetup table
isw_cdafhccbeb_Volume1: 0 488390656 mirror core 2 131072 nosync 2 8:64 0 8:80 0
isw_dcjeehgfed_Volume0: 0 156243968 mirror core 2 131072 nosync 2 8:32 0 8:48 0
VolGroup00-LogVol101: 0 4063232 linear 8:2 33554816
VolGroup00-LogVol100: 0 33554432 linear 8:2 384
```

3.1.2 RAID アレイ削除

RAID アレイの削除は、ハードウェア (ATA RAID タイプ) に合わせて用意されているツールを使用して行う。

3.1.3 RAID アレイへのディスクの追加

dmraid では未サポートの機能である。

代替手段として、以下の手順で RAID アレイへのディスクの追加を行うことができる。

- ② RAID アレイを削除する。削除手順は「3.1.2 RAID アレイ削除」を参照。
- ③ RAID アレイ構成デバイス群に追加したいデバイスを含めて RAID アレイを再構築する。RAID アレイの再構築手順は「3.1.1 RAID アレイ作成」を参照。

3.1.4 RAID アレイからのディスクの削除

dmraid では未サポートの機能である。

代替手段として、以下の手順で RAID アレイの削除を行うことができる。

- ① RAID アレイを削除する。削除手順は「3.1.2 RAID アレイ削除」を参照。
- ② RAID アレイ構成デバイス群に追加したいデバイスを含めないで RAID アレイを再構築する。RAID アレイの再構築手順は「3.1.1 RAID アレイ作成」を参照。

3.1.5 デバイスへの不良マーク付与

dmraid では未サポートの機能である。

3.1.6 RAID アレイのスーパーブロックの操作

dmraid では未サポートの機能である。

3.1.7 RAID アレイの書き込み制限

dmraid では未サポートの機能である。

3.1.8 RAID アレイ上でのイベント検出

dmraid では未サポートの機能である。

3.1.9 RAID アレイのチェック

dmraid では未サポートの機能である。

3.2 復旧

3.2.1 スペアディスクの切り替え

dmraid では未サポートの機能であり、自動復旧は行われません。故障ディスクの入れ替え後に、ハードウェア(ATARAID タイプ)に合わせて用意されているツールでアレイを構築しなおす必要があります。

3.2.2 進捗状況の確認

dmraid では未サポートの機能である。

3.2.3 hotplug 機能

dmraid では未サポートの機能であり、自動復旧は行われません。故障ディスクの入れ替え後に、ハードウェア(ATARAID タイプ)に合わせて用意されているツールでアレイを構築しなおす必要があります。

3.2.4 リブート時の動作

縮退状態で再起動を行った場合、縮退状態のまま起動する。

3.3 管理

3.3.1 故障ディスクの特定

故障ディスクの特定は次のコマンドでできる。

```
dmraid -r [<対象となるRAIDアレイ>]
```

実行例

以下は出力例である。出力結果からは故障ディスク/dev/sdd が取り除かれ、RAID アレイ "isw_dcjeehgfd"が一つのディスクで運用(縮退運用)していることがわかる

```
# dmraid -r
/dev/sdc: isw, "isw_dcjeehgfd", GROUP, ok, 156250078 sectors, data@ 0
/dev/sde: isw, "isw_cdafhccebb", GROUP, ok, 488397166 sectors, data@ 0
/dev/sdf: isw, "isw_cdafhccebb", GROUP, ok, 488397166 sectors, data@ 0
```

3.3.2 RAID 構成情報のバックアップ・リストア

次のコマンドで、構成情報のバックアップを保存する。

```
dmraid -rD [<対象となるRAIDアレイ>]
```

これを実行することにより、以下の形式でバックアップのファイルができる。

```
<デバイス名>_<タイプ>.{dat,offset,size}
```

また、構成情報のバックアップから RAID アレイをリストアするには以下のように手動で書き戻す必要がある。

```
# dd if=sdb_isw.dat of=/dev/sdb_isw bs=1 seek=$(cat sdb_isw.offset) count=$(  
cat sdb_isw.size)
```

実行例

構成情報のバックアップを保存し、ls でファイルが存在しているか確認する。

```
# dmraid -rD  
/dev/sdb: isw, "isw_cahjiaggba", GROUP, ok, 156301486 sectors, data@ 0  
/dev/sdc: isw, "isw_cahjiaggba", GROUP, ok, 156250078 sectors, data@ 0
```

```
# ls  
sdb_isw.dat sdb_isw.offset sdb_isw.size sdc_isw.dat sdc_isw.offset sdc_isw.size
```

4 LVM2

LVM2 は複数台のディスクを一つにまとめたグループから論理ボリューム(パーティション)を作成する機能である。LVM2 の機能の一つとして、論理ボリュームで RAID アレイ(LV:論理ボリューム)を構成することができる。以下に設定運用管理手順について記述する。LVM2 では、RAID1 を構築することができる。

4.1 設定

4.1.1 RAID アレイ作成

LVM2 を使用して RAID アレイ(LV)を設定する手順は以下の通りである。

- ① RAID アレイに組み込む各デバイスに PV (物理ボリューム)があるかを確認する。

```
/usr/sbin/pvdisplay
```

- ② PV を持たないデバイスに対して、PV を作成する。

```
/usr/sbin/pvcreate <対象となるデバイスファイル名>
```

- ③ VG (ボリュームグループ)を作成する。

```
/usr/sbin/vgcreate <VG名> <対象となるデバイスファイル名>
```

- ④ RAID1 で構成されたアレイ(LV)を作成する。"-m"は(オリジナルのディスクを除く)ミラーディスクの個数を指定する。実際には VG 作成時に組み込んだ PV の数-2(オリジナルのディスクおよびログ採取用ディスク)を指定しなければならない。

```
/usr/sbin/lvcreate -m <ミラーディスクの個数> -L <RAIDアレイのサイズ> <VG名>
```

また、明示的に RAID アレイ(LV)に組み込むディスクとログ採取用ディスクを指定する場合には、引数として RAID アレイ(LV)に組み込むデバイスファイル名をログ採取用ディスクが最後になるように指定することで実現できる。

```
/usr/sbin/lvcreate -m <ミラーディスクの個数> -L <RAIDアレイのサイズ> <VG名> <対象となるデバイスファイル名>
```

作成した RAID1(mirror)は、次のコマンドで確認できる。

```
/usr/sbin/lvdisplay -vvv <RAIDアレイ(LV)名>
```

実行例 1

- ② /dev/sdb1、/dev/sdc1、/dev/sdd1、/dev/sde1、/dev/sdf1 の 5 つの PV を作成する。

```
# /usr/sbin/pvcreate /dev/sdb1 /dev/sdc1 /dev/sdd1 /dev/sde1 /dev/sdf1
```

- ③ /dev/sdb1、/dev/sdc1、/dev/sdd1、/dev/sde1、/dev/sdf1 の 5 つの PV で構成された VolGroup10 を作成する。

```
# /usr/sbin/vgcreate VolGroup10 /dev/sdb1 /dev/sdc1 /dev/sdd1 /dev/sde1 /dev/sdf1
```

- ④ VolGroup10 でミラーディスク 3 つ、RAID アレイのサイズ 2G で RAID アレイ(LV)を作成する。

```
# /usr/sbin/lvcreate -m3 -L 2G VolGroup10
# /usr/sbin/lvdisplay -vvv /dev/VolGroup10/lvol0
...
/dev/sdb1 0:      0    500: lvol0_mimage_0(0:0)
/dev/sdb1 1:    500    855: NULL(0:0)
/dev/sdc1 0:      0    500: lvol0_mimage_1(0:0)
/dev/sdc1 1:    500    855: NULL(0:0)
/dev/sdd1 0:      0    500: lvol0_mimage_2(0:0)
/dev/sdd1 1:    500    855: NULL(0:0)
/dev/sde1 0:      0    500: lvol0_mimage_3(0:0)
/dev/sde1 1:    500    855: NULL(0:0)
/dev/sdf1 0:      0      1: lvol0_mlog(0:0)
/dev/sdf1 1:      1    979: NULL(0:0)
Getting device info for VolGroup10-lvol0
...
```

実行例 2

/dev/sdb1 をログ採取用ディスクに指定して RAID アレイ(LV)を作成する。

```
# /usr/sbin/lvcreate -m3 -L2G VolGroup10 /dev/sdc1 /dev/sdd1 /dev/sde1 /dev/sdf1 /dev/sdb1
# /usr/sbin/lvdisplay -vvv /dev/VolGroup10/lvol0
...
/dev/sdb1 0:      0      1: lvol0_mlog(0:0)
/dev/sdb1 1:      1    979: NULL(0:0)
/dev/sdc1 0:      0    500: lvol0_mimage_0(0:0)
/dev/sdc1 1:    500    855: NULL(0:0)
/dev/sdd1 0:      0    500: lvol0_mimage_1(0:0)
/dev/sdd1 1:    500    855: NULL(0:0)
/dev/sde1 0:      0    500: lvol0_mimage_2(0:0)
/dev/sde1 1:    500    855: NULL(0:0)
/dev/sdf1 0:      0    500: lvol0_mimage_3(0:0)
/dev/sdf1 1:    500    855: NULL(0:0)
Getting device info for VolGroup10-lvol0
...
```

4.1.2 RAID アレイ削除

- ① 次のコマンドで、RAID アレイ(LV)を削除する。

```
/usr/sbin/lvremove <RAIDアレイ (LV) 名>
```

- ② 次のコマンドで、VG を削除する。

```
/usr/sbin/vgremove <VG名>
```

- ③ 次のコマンドで、PV を削除する。

```
/usr/sbin/pvremove <対象となるデバイスファイル名>
```

実行例

RAID アレイ(VG 名 VolGroup10 、LV 名 lvol0)を削除する。

```
# /usr/sbin/lvremove /dev/VolGroup10/lvol0
# /usr/sbin/vgremove /dev/VolGroup10
# /usr/sbin/pvremove /dev/sdb1 /dev/sdc1 /dev/sdd1 /dev/sde1 /dev/sdf1
```

4.1.3 RAID アレイへのディスクの追加

LVM2 では未サポートの機能である。

代替手段として、以下の手順で RAID アレイ(LV)へのディスクの追加を行うことができる。

- ① RAID アレイ(LV)、VG を削除する。LV、VG の削除手順は「4.1.2 RAID アレイ削除」を参照。
- ② 追加したいデバイスを含めた PV、VG、RAID アレイ(LV)を作成することで追加したいデバイスを含んだ RAID アレイを再構築する。RAID アレイの再構築手順は「4.1.1 RAID アレイ作成」を参照。

4.1.4 RAID アレイからのディスクの削除

次のコマンドで RAID アレイ(LV)からディスクを削除する。

```
/dev/sbin/lvconvert -m<削除するミラーディスクの個数> <RAIDアレイ (LV) 名>
```

4.1.5 デバイスへの不良マーク付与

LVM2 では未サポートの機能である。

4.1.6 RAID アレイのスーパーブロックの操作

- (1) スーパーブロックの内容を表示

次のコマンドで RAID アレイのスーパーブロックの内容を表示する。

```
/usr/sbin/vgcfgbackup <表示するVG名>
```

(2) スーパーブロックの内容を更新

次のコマンドで RAID アレイのスーパーブロックの内容を更新する。

```
/usr/sbin/vgcfgrestore -n <対象となるVG名>
```

(3) スーパーブロックを 0 で初期化

次のコマンドで RAID アレイのスーパーブロックを 0 で初期化する。

```
/usr/sbin/vgreduce <対象となるVG名>
```

4.1.7 RAID アレイの書き込み制限

(1) RAID アレイの書き込み制限

次のコマンドで、RAID アレイ(LV)に書き込み制限を行う。

```
/usr/sbin/lvchange -p r <LV名>
```

次のコマンドで RAID アレイ(LV)への書き込みが制限されていることを確認する。

```
/usr/sbin/lvdisplay <LV名>
```

実行例

RAID アレイ(VG 名 VolGroup10、LV 名 lvol0)への書き込み制限を行う。

```
# /usr/sbin/lvchange -p r /dev/VolGroup10/lvol0
# /usr/sbin/lvdisplay /dev/VolGroup10/lvol0
--- Logical volume ---
LV Name                /dev/VolGroup10/lvol0
VG Name                VolGroup10
LV UUID                Bdi1NG-82jg-dBJ9-60hk-Fywk-20wJ-MoGceT
LV Write Access        read only
LV Status               available
# open                 0
LV Size                2.00 GB
Current LE             512
Segments               1
Allocation              inherit
Read ahead sectors     0
Block device           253:7
```

(2) RAID アレイの書き込み制限の解除

次のコマンドで、RAID アレイ(LV)の書き込み制限を解除する。

```
/usr/sbin/lvchange -p rw <LV名>
```

RAID アレイ(LV)への書き込み制限が解除されていることは、次のコマンドで確認できる。

```
/usr/sbin/lvdisplay <LV名>
```

実行例

RAID アレイ(VG 名 VolGroup10、LV 名 lvol0)への書き込み制限を解除する。

```
# /usr/sbin/lvchange -p rw /dev/VolGroup10/lvol0
# /usr/sbin/lvdisplay /dev/VolGroup10/lvol0
--- Logical volume ---
LV Name           /dev/VolGroup10/lvol0
VG Name           VolGroup10
LV UUID           Bdi iNG-82jg-dBJ9-60hk-Fywk-20wJ-MoGceT
LV Write Access   read/write
LV Status         available
# open            0
LV Size           2.00 GB
Current LE        512
Segments          1
Allocation        inherit
Read ahead sectors 0
Block device      253:7
```

4.1.8 RAID アレイ上でのイベント検出

LVM2 では未サポートの機能である。

4.1.9 RAID アレイのチェック

LVM2 では未サポートの機能である。

4.2 復旧

4.2.1 スペアディスクの切り替え

LVM2 では未サポートの機能である。

4.2.2 進捗状況の確認

次のコマンドで、復旧の進捗状況が確認できる。”Copy%”の欄の値が進捗状況である。

```
/usr/sbin/lvs
```

実行例

復旧の進捗状況を確認する。

#	/usr/sbin/lvs								
LV	VG	Attr	LSize	Origin	Snap%	Move	Log	Copy%	
lv10	VolGroup10	mwi-a-	2.00G				lv10_mlog	100.00	

4.2.3 hotplug 機能

LVM2 では未サポートの機能であり、故障ディスクの入れ替え後に、LVM2でRAIDアレイ(LV)を構築しなおす必要がある。

4.2.4 リブート時の動作

縮退状態で再起動を行った場合、縮退状態のまま起動する。

4.3 管理

4.3.1 故障ディスクの特定

以下の手順で故障ディスクを検出することができる。

(1) LVM2 から構成変更を認識できる場合

ディスク撤去後等、LVM2 が構成変更情報を認識できる場合は、lvs コマンドにより、故障部位を特定できる。

① 故障ディスクの UUID を検出する。

次のコマンドで故障ディスクの UUID を表示する。

```
/usr/sbin/lvs -a -o+devices <VG名>
```

② UUID から故障ディスクを特定する。

①で表示された UUID を VG のメタデータ(/etc/lvm/backup/<VG名>)と照らし合わせることで故障ディスクを特定できる。以下の形式のコマンドで出力される「device =」の値が故障ディスクとなる。

```
grep -C1 <故障ディスクのUUID> /etc/lvm/backup/<VG名>
```

実行例

- ① VG "VolGroup10"に対して lvs コマンドを実行し、故障ディスクの UUID を検出する。

```
# /usr/sbin/lvs -a -o+devices VolGroup10
...
Couldn't find device with uuid 'e1R2u6-YH63-EH3C-zMcE-lfut-Ulzp-k1Dvjh'.
Couldn't find all physical volumes for volume group VolGroup10.
Volume group "VolGroup10" not found
```

- ② UUID を VG のメタデータ(/etc/lvm/backup/VolGroup10)と照らし合わせて故障ディスクを特定する。出力結果から故障ディスクが/dev/sde1 であることがわかる。

```
# grep -C1 e1R2u6-YH63-EH3C-zMcE-lfut-Ulzp-k1Dvjh /etc/lvmbbackup/VolGroup10
pv3 {
    id = "1uQKIQ-NY2r-Sgqf-RaXU-7cAX-Ep6l-3A2mcZ"
    device = "/dev/sde1"    # Hint only
```

(2) LVM2 から構成変更を認識できない場合

セクタ障害等、現状の LVM2 が検出できない障害については、dmsetup コマンドにより LVM2 の下位に位置する device-mapper ドライバを利用するデバイスとしての状態を確認できる。

- ① ディスクの構成情報からデバイス名を取得する。
次のコマンドでディスクの構成情報を表示し、デバイス名を取得する。

```
/usr/sbin/dmsetup ls --tree
```

- ② ディスクの状態を表示する。
①で表示された構成情報と照らし合わせることにより、故障ディスクを特定できる。以下の形式のコマンドで出力される「device =」の値が故障ディスクとなる。

```
/usr/sbin/dmsetup status <デバイス名>
```

実行例

以下の例では 253:15(ディスク d7 上の lvol10_mimage_0)と 253:17(ディスク d1 上の lvol0_mimage_1 がミラーされている構成であるとわかる。このときの状態は 253:15 が”D”(故障)、253:17 が”A”(正常)であることがわかる。

- ① `dmsetup ls` コマンドを実行し、デバイス名を取得する。

```
# /usr/sbin/dmsetup ls --tree
testvg-lvol0 (253:18)
|-testvg-lvol0_mimage_1 (253:17)
|  `--d1 (253:1)
|-testvg-lvol0_mimage_0 (253:15)
|  `--d7 (253:7)
`--testvg-lvol0_mlog0 (253:10)
   `--d5 (253:5)
```

- ② `dmsetup status` コマンドでデバイスのステータスを表示する。

これより、“253:15”が故障“D”していると判断できる。先の結果より、“253:15”は“lvol0_mimage_0”である。論理ボリューム“lvol0_mimage_0”と物理ディスクとの対応は `lvdisplay` コマンドで判断できる。

```
# dmsetup status testvg-lvol0
0 4096 mirror 2 253:15 253:17 3/4 1 DA 3 disk 253:10 A
```

4.3.2 RAID 構成情報のバックアップ・リストア

次のコマンドで RAID アレイ(を含む VG)のバックアップができる。

```
/usr/sbin/vgcfgbackup <VG名>
```

バックアップファイルとして `/etc/lvm/backup/<VG名>` が作成される。

次のコマンドで RAID アレイ(を含む VG)のリストアができる。

```
/usr/sbin/vgcfgbackup <VG名>
```

5 付録 構成情報

本手順の記載内容は以下の構成で確認した。

[HW 構成]

Express5800/120Lh(Xeon 3.4GHzx2, メモリ 6GB, SCSI 36GB×6)

Super Server 6025B-T(Xeon 3.4GHzx, メモリ 6GB, SATA 80G×6)

[OS]

Red Hat Enterprise Linux 4.0 Update 5

[ツール]

mdadm Version 1.12.0

dmraid Version 1.0.0.rc14

lvm2 Version 2.02.21