

CLUSTERPRO

MC RootDiskMonitor 2.0 for Linux

CLUSTERPRO

MC StorageSaver for BootDisk 2.0 (for Linux)

ユーザーズガイド

© 2015(Mar) NEC Corporation

- ☐ 製品の概要
- ☐ OS ディスクの監視方式について
- ☐ 設定ファイル
- ☐ 操作・運用手順
- ☐ syslog メッセージ
- ☐ 注意・制限事項について
- ☐ リファレンス
- ☐ 付録

はしがき

本書は、CLUSTERPRO MC RootDiskMonitor 2.0 for Linux（以後 RootDiskMonitor と記載します）、および CLUSTERPRO MC StorageSaver for BootDisk 2.0 (for Linux) の OS ディスク監視に関する設定について記載したものです。

（注）StorageSaver for BootDisk は、以後 RootDiskMonitor と表記します。

なお、StorageSaver for BootDisk では、以下の機能は利用できません。

- パトロールシーク機能
- HW-RAID 監視機能

基本機能として以下の運用が可能です。

- I/O パス監視機能
- 障害レポート機能
- クラスタウェア連携機能

拡張機能として以下の運用が可能です。

- ミラー構成監視機能（本機能については次期バージョン以降でサポート予定）

（１） 商標および登録商標

- ✓ Red Hat は、米国およびその他の国における Red Hat,Inc.の登録商標または商標です。
- ✓ Oracle は、Oracle やその関連会社の米国およびその他の国における、商標または登録商標です
- ✓ Linux は、Linus Torvalds 氏の米国およびその他の国における、登録商標または商標です。
- ✓ EMC、Symmetrix DMX、Symmetrix VMAX、CLARiX、VNX は EMC Corporation の商標または登録商標です。
- ✓ その他、本書に登場する会社名および商品名は各社の商標または登録商標です。
- ✓ なお、本書では®、TM マークを明記していません。

(2) 本リリースの強化点について

RootDiskMonitor 2.0 (2015 年 4 月出荷版)では、下記の機能を強化しています。

- ・ Red Hat Enterprise Linux 7.0 および Oracle Linux 7.0 をサポートしました。
Red Hat Enterprise Linux 7.0 および Oracle Linux 7.0 環境における監視機能をサポートしました。
Red Hat Enterprise Linux 7.0 および Oracle Linux 7.0 環境ではプロセスの起動、停止方法を変更しています。
詳細は“4.1 運用管理コマンドの操作手順”を参照してください。
- ・ Hyper-V のゲスト OS における監視機能をサポートしました。
Hyper-V のゲスト OS における OS ディスクの監視機能をサポートしました。
設定手順については「CLUSTERPRO MC RootDiskMonitor 仮想環境(ゲスト OS)での設定手順」を参照してください。
- ・ セクタサイズが 4096 バイトのディスク装置をサポートしました。
セクタサイズが 4096 バイトのディスク装置に対する監視をサポートしました。
- ・ 設定ファイル(rdm.config)に定義するデバイスファイルの形式を変更しました。
設定ファイル(rdm.config)に定義するデバイスファイルの形式をブロックスペシャルファイル名から udev デバイスファイル名に変更しました。
これにより構成変更等でブロックスペシャルファイル名が変更された場合でも設定ファイルの再作成が不要になります。
なお、従来どおりブロックスペシャルファイル名も設定可能です。
- ・ デバイスファイルのチェック機能をサポートしました。
設定ファイル(rdm.config)にスペシャルファイル名と udev デバイスファイル名の両方を指定した場合、デーモンプロセス起動時にスペシャルファイル名の妥当性チェックを行います。
スペシャルファイル名が変更されていた場合には、本機能により当該デバイスを監視対象外とします。
これにより、誤った設定のまま監視を開始して障害を誤検知することを防止します。
なお、デバイスファイルのチェック機能は、設定ファイル自動生成(rdmconfig) のサポート対象外です。本機能を使用する場合は、手動で設定ファイルを編集してください。

- ・ パトロールシークの検査中断、再開機能をサポートしました。
運用管理コマンドによるパトロールシークの検査中断、再開がおこなえるよう強化しました。
詳細は「CLUSTERPRO MC RootDiskMonitor 2.0 for Linux パトロールシーク機能ユーザーズガイド」を参照してください。

(3) これまでの強化点について

RootDiskMonitor 1.1 (2013 年 10 月出荷版)では、下記の機能を強化しています。

- ・ TestI/O 発行方式の設定方法を変更しました。
TestI/O の発行方式を設定するパラメータとして、設定ファイルに新規パラメータ (TESTIO_MODE)を追加しました。
従来のバージョンにおけるパラメータ TESTIO_DIRECT は本パラメータに統合されました。
また、これに伴い TestUnitReady の発行抑止についても、本パラメータで設定するよう変更しました。
パラメータの詳細については “3.2 設定ファイルの記述” を参照してください。
なお、従来どおり TESTIO_DIRECT の指定、ならびに TIME_TUR_INTERVAL での TestUnitReady の発行抑止も可能ですので、1.0 以前のバージョンの設定ファイルもそのまま使用可能です。
- ・ 障害検出時のアクション内容を強化しました。
I/O ストール検出時に実行するアクションとして、クラスタウェア連携用デーモン (rdmstat)を使用した CLUSTERPRO 連携が指定可能となりました。
これに伴い、障害時のアクションを指定するパラメータ (VG_FAULT_ACTION および VG_STALL_ACTION) に指定する値も変更しました。
パラメータの詳細については “3.2 設定ファイルの記述” を参照してください。
なお、従来の設定値も本バージョンでは使用可能ですので、1.0 以前のバージョンの設定ファイルもそのまま使用可能です。

RootDiskMonitor 1.2 (2014 年 4 月出荷版)では、下記の機能を強化しています。

- ・ パトロールシーク機能にて 2TB 以上のディスクをサポートしました。
パトロールシーク機能において、2TB 以上のディスクの検査が可能となりました。

目次

1. 製品の概要	1
1.1. 製品概要について	1
1.2. 製品の構成について	2
2. OS ディスクの監視方式について	4
2.1. OS ディスクの監視	4
2.2. I/O パスの監視手順について	8
2.3. I/O パスの異常を検出すると	12
2.4. クラスタウェアとの連携について	14
3. 設定ファイル	16
3.1. 本製品の導入	16
3.2. 設定ファイルの記述	22
4. 操作・運用手順	27
4.1. 運用管理コマンドの操作手順	27
4.2. カスタムモニタリソースによる CLUSTERPRO との連携	34
4.3. CLUSTERPRO のサーバ管理プロセス (clpnm) の強制終了による CLUSTERPRO との連携	37
4.4. システムメモリダンプ採取と OS 強制停止による CLUSTERPRO との連携	39
5. syslog メッセージ	40
5.1. syslog に出力するメッセージについて	40
5.2. 警報対象として登録することを推奨するメッセージ一覧	41
5.3. その他のメッセージ一覧	44
5.4. 運用管理製品との連携	44
6. 注意・制限事項について	45
6.1. 注意・制限事項	45
7. リファレンス	48
8. 付録	54
8.1. 本製品のテスト手順について	54
8.2. カスタムモニタリソースによる CLUSTERPRO との連携手順	67
8.2.1. CLUSTERPRO 連携設定	67
8.2.2. 動作確認	74
8.3. HW-RAID 状態表示	75

1. 製品の概要

1.1. 製品概要について

(1) 製品の提供する主な機能

本製品は、Linux の OS ディスクを構成する I/O パスの動作状態を定期監視します。I/O パスに異常が見られるとエラーレポートを通知し、さらに OS ディスクが動作不能になるとクラスタウェアと連携しノードを切り替えることでクラスタシステムでの可用性を向上させます。

- ・ I/O パス監視機能
OS ディスクを構成する I/O パスに対して死活監視、I/O リクエストのストール監視を行います。シングル構成、ミラー構成の OS ディスクを監視できます。
- ・ 障害レポート機能
I/O パスを定期監視し異常を検出すると syslog、コンソールに異常レポートを通知します。
- ・ クラスタウェア連携機能
OS ディスクを構成する I/O パスがすべて障害となり、LUN(論理ディスク装置)へのアクセスが不可能になると、クラスタウェアと連携することによりノード切り替えを実現します。

クラスタウェアと連携しノード切り替えを実現するには以下の3つの手法があります。

- (1) CLUSTERPRO のサーバ管理プロセス (clpnm) を強制終了させる方式
- (2) OS ディスク監視専用のクラスタリソースを作成する方式
- (3) Linux のシステムメモリダンプを採取し、OS を強制的に停止させる方式

クラスタウェアを利用しない非クラスタシステムでは、ノード切り替え機能はご利用いただけません。

1.2. 製品の構成について

(1) プロダクト構成

本製品は Linux の OS ディスク監視を行います。

(2) ソフトウェア構成

プロセス構成は以下の通りです。

- ・ rdmdiagd(1M) OS ディスク監視エンジン
- ・ rdmping(1M) OS ディスク監視モニタ
- ・ rdmadmin(1M) 運用管理コマンド
- ・ rdmconfig(1M) 設定ファイル自動生成コマンド
- ・ rdmstat(1M) クラスタウェア連携用デーモン

ファイル、ディレクトリ構成は以下の通りです。

- ・ 実行形式ディレクトリ /opt/HA/RDM/bin
- ・ 設定ファイル /opt/HA/RDM/conf/rdm.config
- ・ rc ファイル格納ディレクトリ
 - 【Red Hat Enterprise Linux 5.x】
 - 【Red Hat Enterprise Linux 6.x】
 - 【Oracle Linux 6.x】
 - /etc/init.d
 - /etc/rc.d/rc0.d
 - /etc/rc.d/rc1.d
 - /etc/rc.d/rc2.d
 - /etc/rc.d/rc3.d
 - /etc/rc.d/rc4.d
 - /etc/rc.d/rc5.d
 - /etc/rc.d/rc6.d
- ・ Unit 定義ファイル格納ディレクトリ
 - 【Red Hat Enterprise Linux 7.x】
 - 【Oracle Linux 7.x】
 - /usr/lib/systemd/system

(3) サポート範囲

Linux OS ディスクが対象となります。

SCSI インタフェース接続の OS ディスク装置、増設ディスク装置

FC 接続のディスクアレイ装置

- ・ NEC 社製 iStorage 全シリーズ(ただし、E1 シリーズは除きます)
- ・ EMC 社製 CLARiX シリーズ
- ・ EMC 社製 VNX シリーズ
- ・ EMC 社製 Symmetrix DMX シリーズ、Symmetrix VMAX シリーズ

下記のボリューム管理製品を対象とします。

- LVM 上に構築された OS ディスク(/dev/VolGroup00,etc)
- 物理ディスク上に直接構築された OS ディスク(/dev/sdx)

注意: StorageSaver for BootDisk の場合は、サーバ搭載の内蔵ディスクの監視には使用できません。

2. OS ディスクの監視方式について

2.1. OS ディスクの監視

(1) 監視のフレームワークについて

本製品では、OS ディスクの障害を検出するために、I/O パスに対して定期的に Test I/O を行います。Test I/O で監視対象となる項目は下記の通りです。

- I/O パスの死活監視
- I/O リクエストのストール監視

Test I/O は SCSI パススルードライバ経由で行われますが、下記の SCSI コマンドを使用します。

- Inquiry command
- TestUnitReady command

(2) 監視対象となる I/O パスについて

OS ディスクを構成する I/O パスが監視対象となります。
なお、コンフィグレーションファイルに OS ディスクを構成する I/O パスのスペシャルファイルと監視ルールの設定が必要です。

(3) I/O パスの異常を検出すると

Test I/O で異常を検出した I/O パスは、syslog、コンソールに障害レポートを通知します。
さらに、LUN(論理ディスク装置)へのアクセスが不可能になると、コンフィグレーションファイルで指定されたアクションを実行します。

(4) I/O パスの状態について

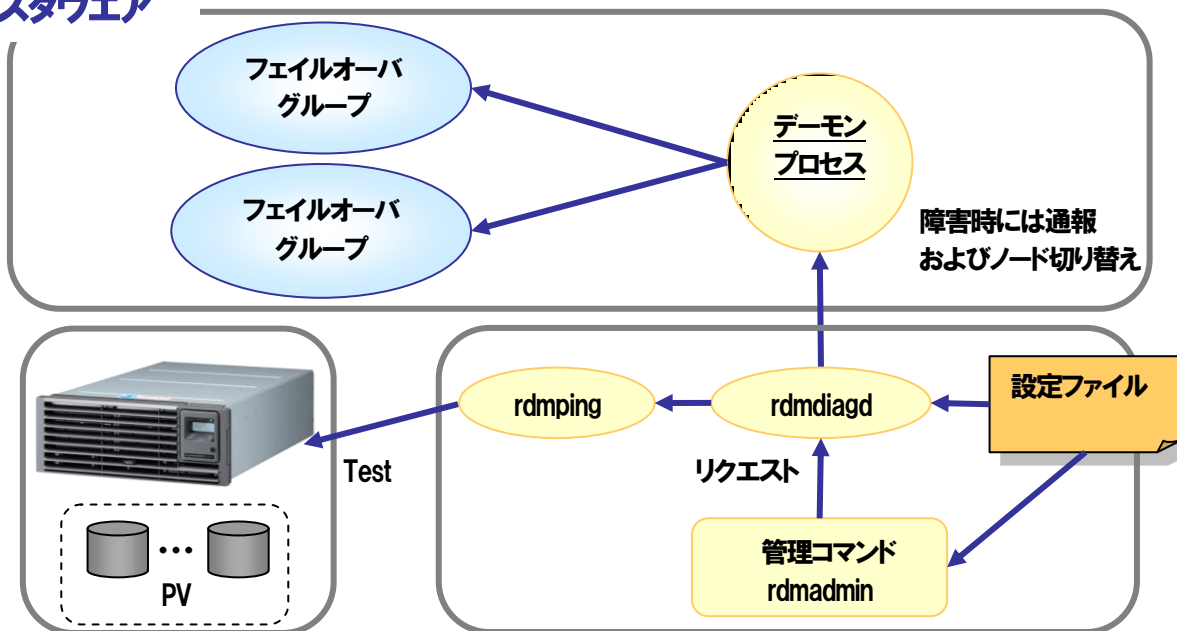
LUN レベルのステータス (I/O パスの監視状態) として、以下の状態をレポートします。

- UP
I/O パスが正常に動作している状態です。
- DOWN
I/O パ스에異常があり、利用不可な状態です。

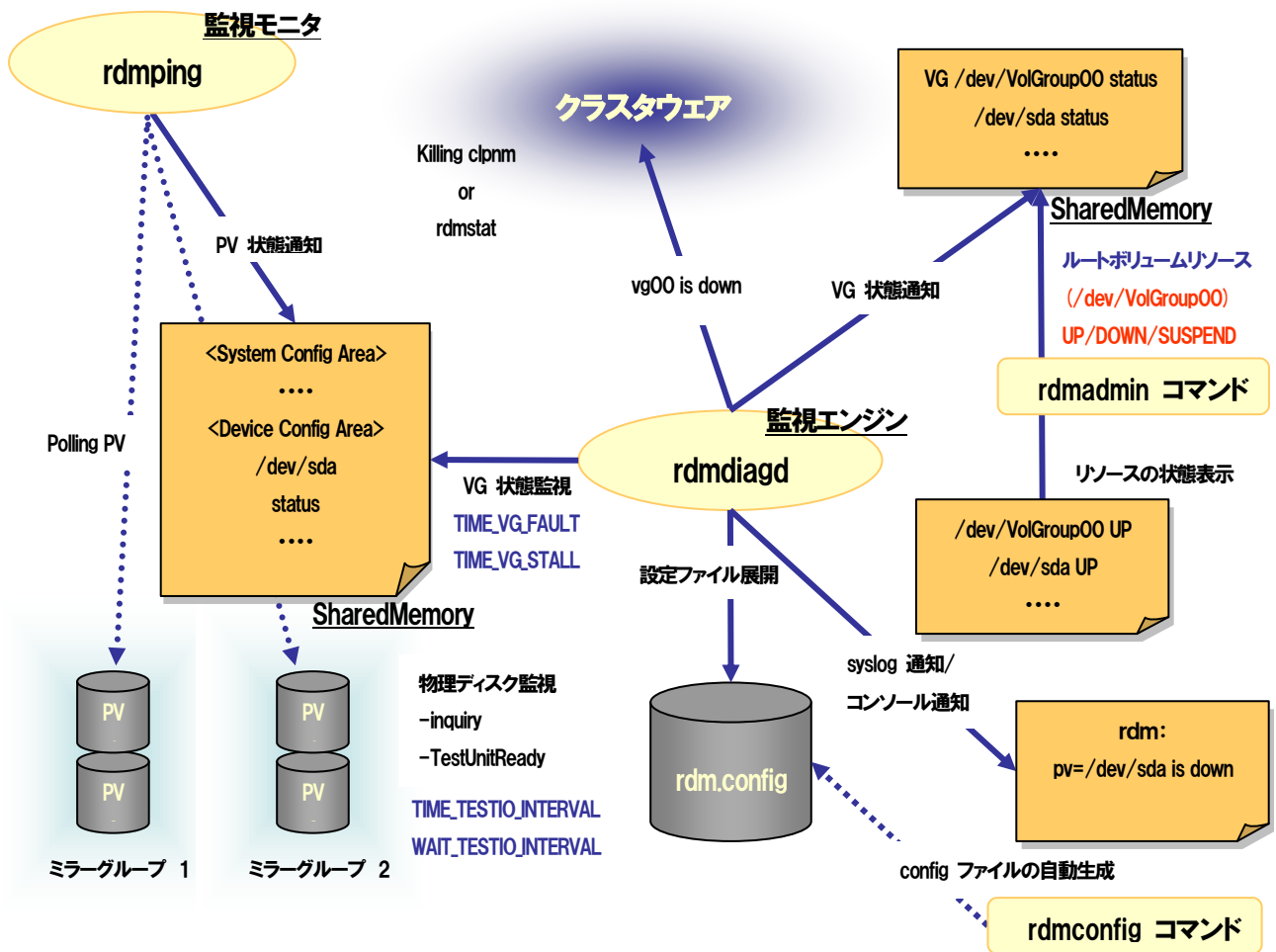
VG レベルのステータス (I/O パスをミラーグループとして束ねた監視状態) として、以下の状態をレポートします。

- UP
VG が正常に動作している状態です。
- SUSPEND
VG を構成するミラー片系の I/O パ스에異常がある状態です。
- DOWN
VG に異常があり、利用不可な状態です。

クラスタウェア



【RootDiskMonitor の構成】



【RootDiskMonitor のプロセスモデル】

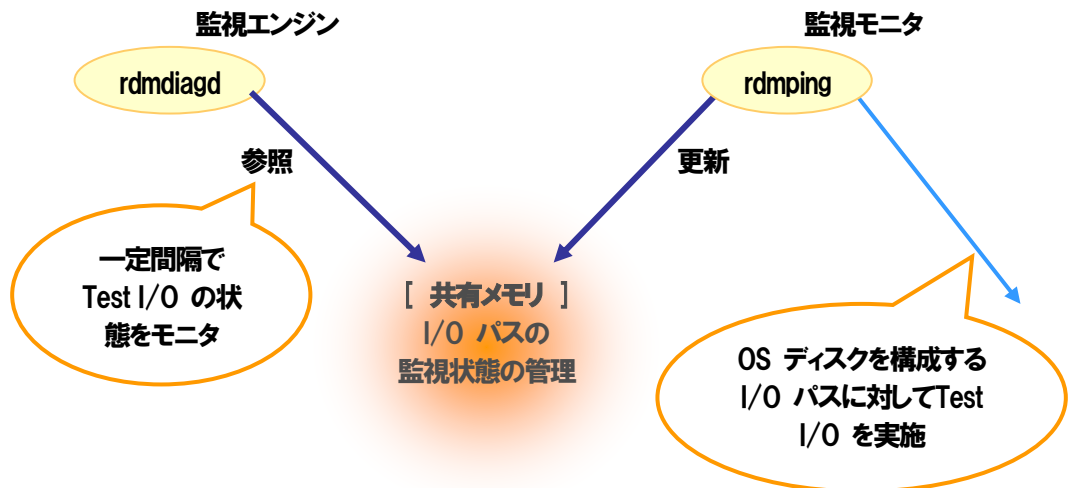
2.2. I/O パスの監視手順について

(1) I/O パスの死活監視

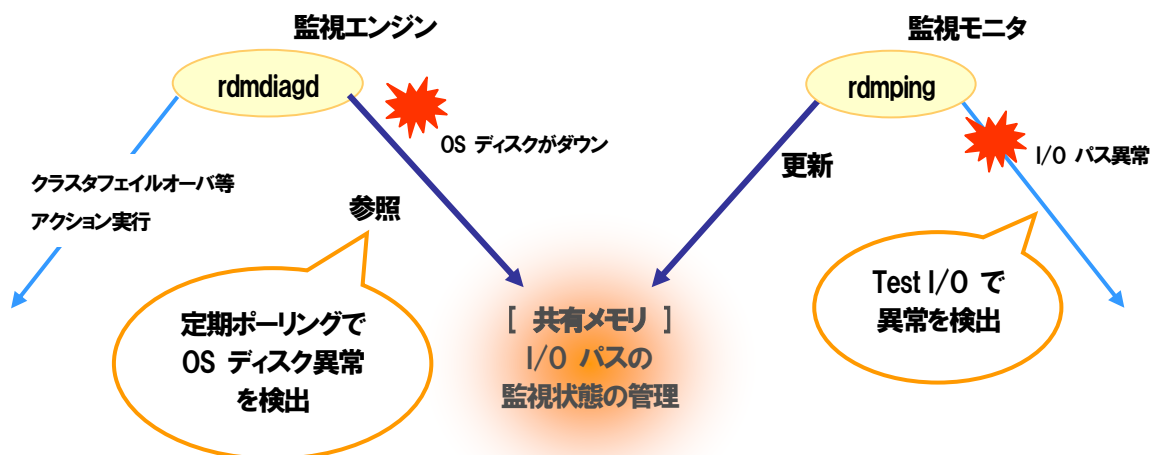
OS ディスクを構成する I/O パスに対し定期的に SCSI パススルー機能を利用して Test I/O を発行することで、I/O パスの動作状態を監視します。
Test I/O が正常終了しない、またはタイムアウトした場合は I/O パスを異常と判定します。

注意 RootDiskMonitor を動作させる場合、Linux パススルードライバ(sg ドライバ)がインストールされ、事前にカーネルにロードされている必要があります。

【Test I/O のフレームワーク】

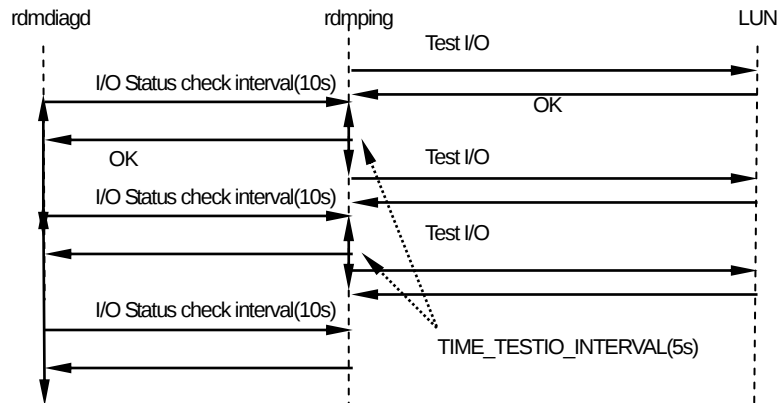


【Test I/O でOS ディスクの異常を検出】



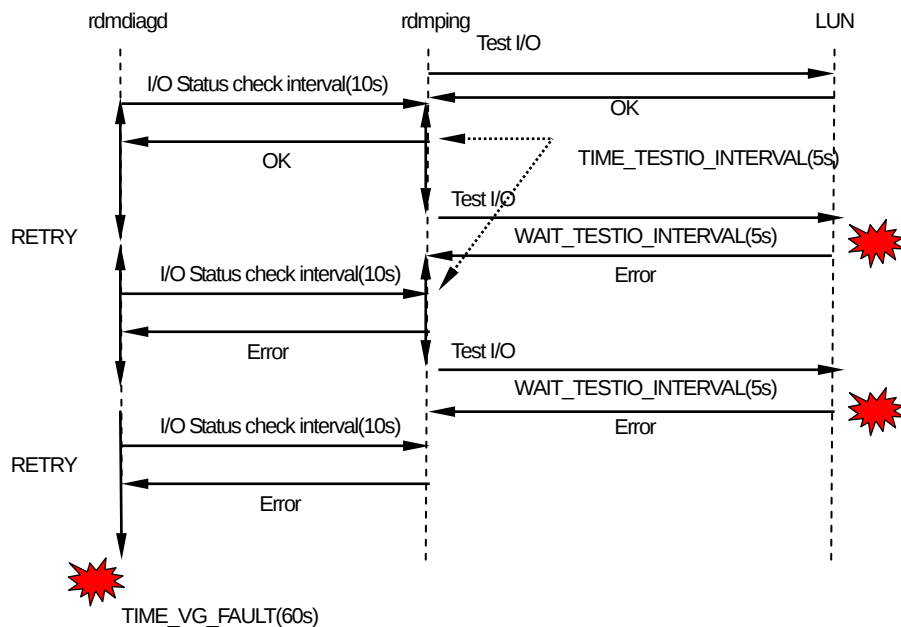
(2) Test I/O の正常なシーケンスは、以下のような動作になります。

基本タイマ(I/O Status Check Interval)	:10(秒)
TIME_VG_FAULT	:60(秒)
TIME_TESTIO_INTERVAL	:5(秒)
WAIT_TESTIO_INTERVAL	:5(秒)



(3) Test I/O で異常を検出すると、以下のような動作になります。

基本タイマ(I/O Status Check Interval)	:10(秒)
TIME_VG_FAULT	:60(秒)
TIME_TESTIO_INTERVAL	:5(秒)
WAIT_TESTIO_INTERVAL	:5(秒)

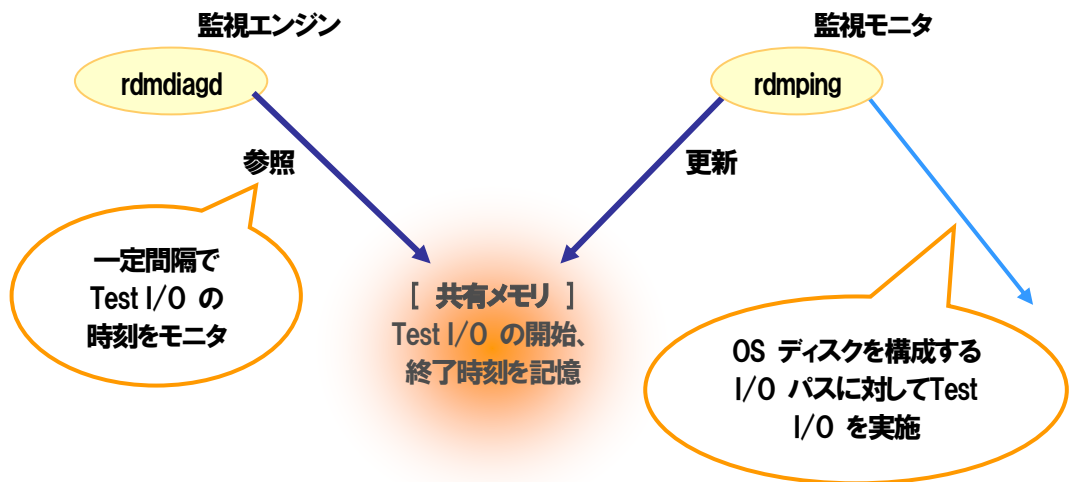


タイムオーバ
syslog 通知 or クラスタウェアデーモン強制停止 or rdmstat 停止 or os system dump + panic

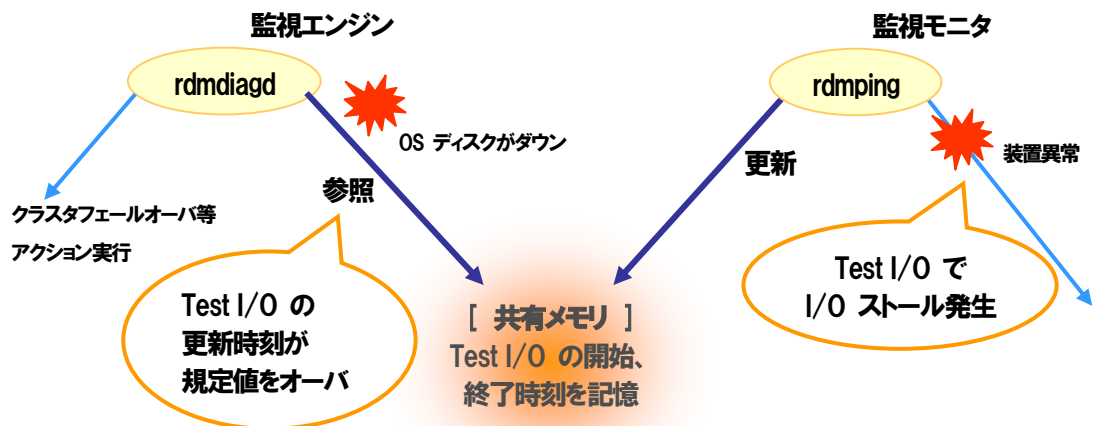
(4) I/O パスのストール監視

OS ディスクを構成する I/O パスに対し定期的に SCSI パススルー機能を利用して Test I/O を発行することで、OS 全体のストール状態を監視します。
Test I/O が一定時間以内に正常完了しない場合は I/O パスを異常と判定します。

【I/O ストール監視のフレームワーク】

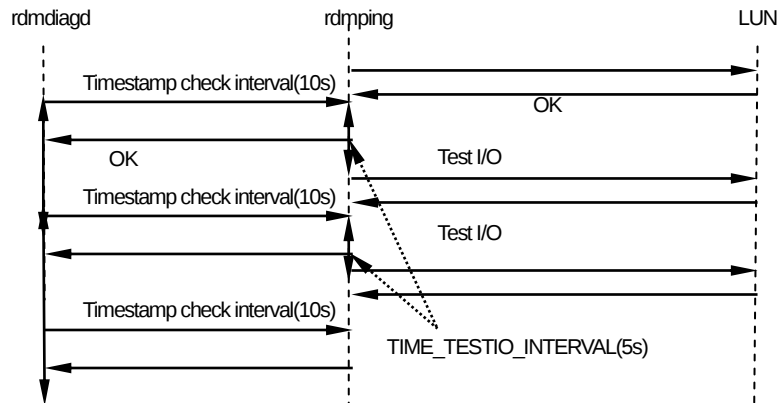


【I/O ストールを検出すると】



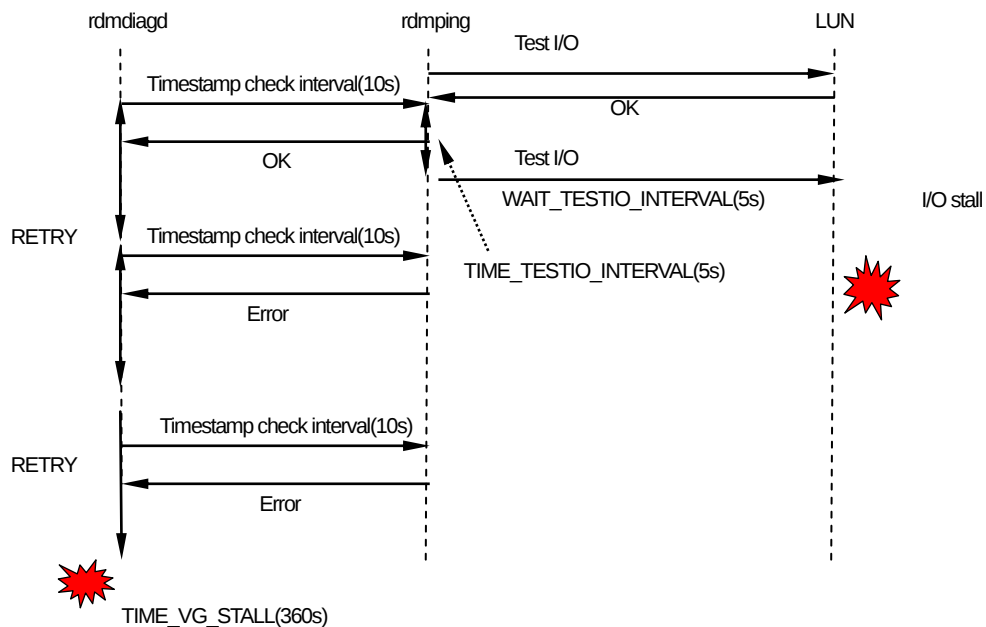
(5) Test I/O の正常なシーケンスは、以下のような動作になります。

基本タイマ(timestamp Check Interval)	:10(秒)
TIME_VG_STALL	:360(秒)
TIME_TESTIO_INTERVAL	:5(秒)
WAIT_TESTIO_INTERVAL	:5(秒)



(6) Test I/O で I/O ストールを検出すると、以下のような動作になります。

基本タイマ(timestamp Check Interval)	:10(秒)
TIME_VG_STALL	:360(秒)
TIME_TESTIO_INTERVAL	:5(秒)
WAIT_TESTIO_INTERVAL	:5(秒)



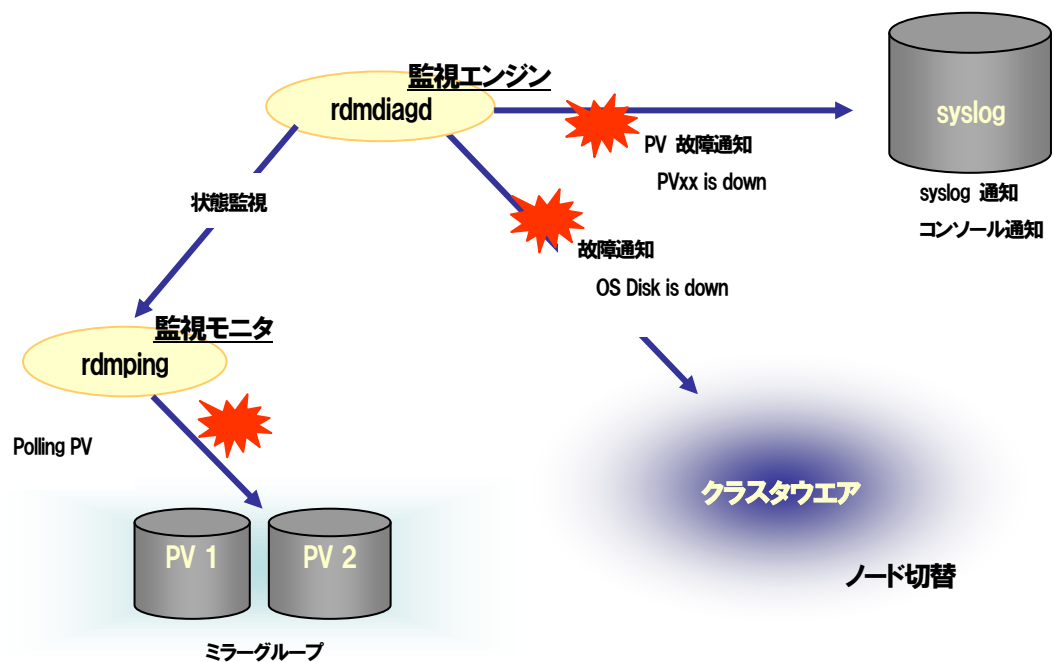
タイムオーバ
syslog 通知 or クラスタウェアデーモン強制停止 or rdmstat 停止 or os system dump + panic

2.3. I/O パスの異常を検出すると

(1) I/O パスの異常を検出すると

I/O パスの異常を検出すると、syslog にエラーメッセージを出力します。
当該 I/O パスの監視は継続しますので I/O パスが復旧次第、
正常状態として監視を続けます。

【Test I/O (Polling) 方式によるディスク監視】



(2) メディアエラーで異常を検出すると ※本機能は次期バージョン以降でサポート予定です。

syslog の監視を行い、ログ内に監視キーワードを検出すると、
メディアエラーと判断し、障害ディスクの切り離しを行います。

(3) 両系障害レベルの異常を検出すると

OS ディスクを構成する I/O パスで異常が発生し、LUN へのアクセスが不可能になると、syslog、コンソールにエラーメッセージを出力します。

- シングルパス構成での I/O パス異常
- ソフトミラー構成での正副 LUN 異常

両系障害レベルで異常となると、設定ファイルで規定されたアクションを実行します。

- 警報通知のみ
- CLUSTERPRO のサーバ管理プロセス (clpnm) を強制終了しノード切り替え
- クラスタ監視リソース連動によるノード切り替え
- システムメモリダンプを採取し、強制的に OS を停止しノード切り替え
- ユーザー定義コマンドの実行(本リリースでは未サポート)

2.4. クラスタウェアとの連携について

OS ディスクの障害で動作不能な状態に陥った場合にクラスタウェアと連携することで待機ノードへ切り替え業務を継続することができます。

本書ではクラスタウェア製品として、CLUSTERPRO を例にクラスタウェア連携について説明します。
(以降の章でも特に断りが無い限り、クラスタウェア連携については CLUSTERPRO を対象とします)

RootDiskMonitor が CLUSTERPRO と連携するには、以下の 3 つの方式があります。

1. CLUSTERPRO のカスタムモニタリソースにクラスタウェア連携デーモン(rdmstat)を登録する方式
2. CLUSTERPRO のサーバ管理プロセス (clpnm) を強制終了し、ノードを切り替える方式
3. システムメモリダンプを採取し、強制的に OS を停止し、ノードを切り替える方式

RootDiskMonitor としては 1 の方式を推奨します。

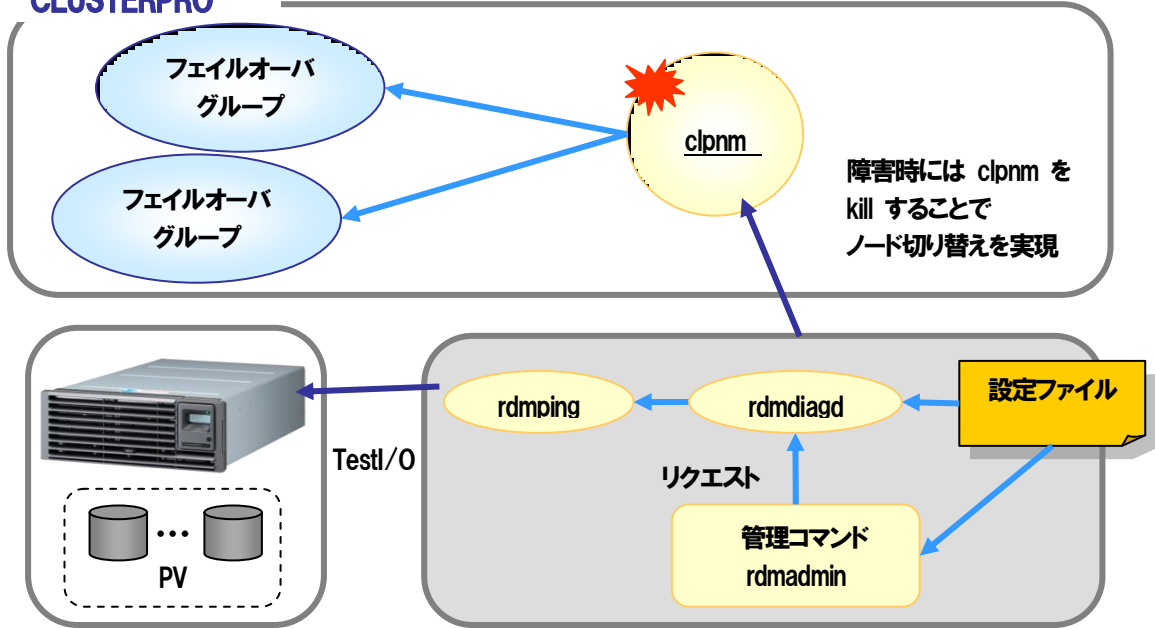
具体的な連携の設定手順については、後述の CLUSTERPRO との連携設定を参照してください。

また、CLUSTERPRO と連携しノード切り替え、ノードダウンを行うには、以下の注意事項があります。

- ノード切り替えは、CLUSTERPRO を利用したクラスタシステムで有効です。
CLUSTERPRO を導入していないシステムではご利用になれません。
- OS ディスクが壊れている場合は、ファイル I/O が停止するため syslog やコンソールにエラーメッセージを出力できない場合があります。
また、システムメモリダンプを採取し OS を強制停止する方式では、システムメモリダンプが採取できない場合があります、OS の強制停止が正しく完了しない場合があります。

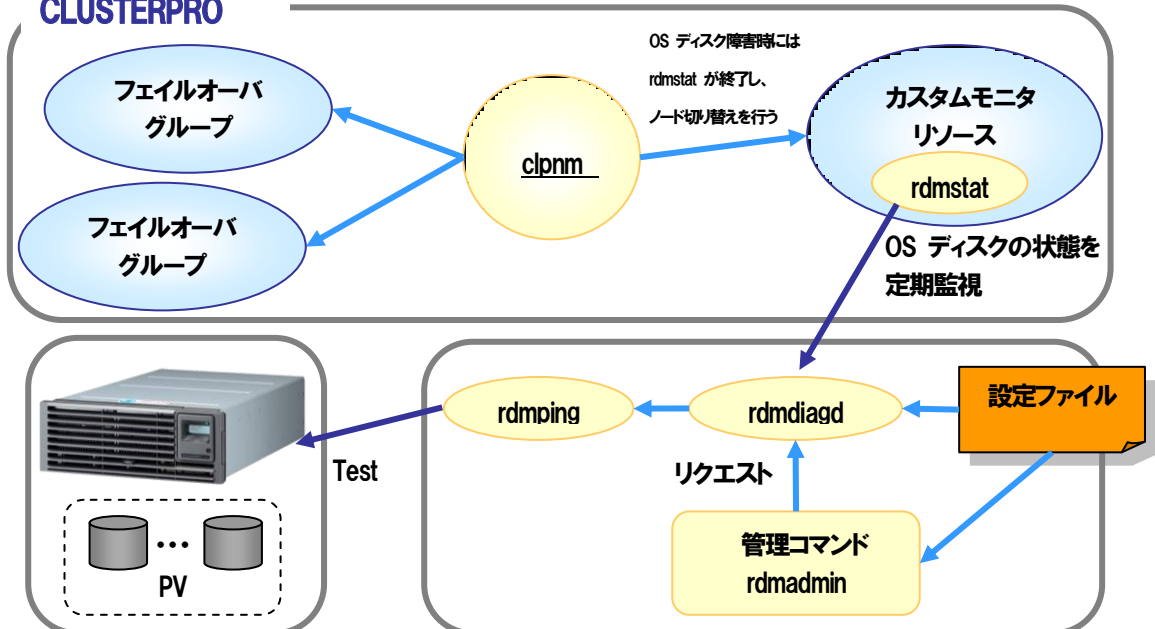
ただし、この場合でも CLUSTERPRO で OS の異常を検出することはできるため、ノードの切り替え自体は問題なく実行されます。

CLUSTERPRO



【 clpnm (1M) を強制終了する手法】

CLUSTERPRO



【カスタムモニタリソースによるフェイルオーバーグループ連動】

3. 設定ファイル

3.1. 本製品の導入

本製品の導入手順について説明します。

導入手順の詳細については、あわせて『CLUSTERPRO MC RootDiskMonitor 2.0 for Linux リリースメモ』もご覧ください。

(1) インストールの前に

コードワードの取得

本製品を導入する前に、あらかじめロック解除のためのコードワードを取得する必要があります。
製品添付の「コードワードについて」の手順にしたがって、コードワードを取得してください。

コードワードの登録

製品添付の「コードワードについて」及び「コードワード登録手順」の「ライセンスツールのインストール」及び「コードワード登録」の手順にしたがって本製品をインストールするマシンに、取得したコードワードを登録してください。

(2) インストール

本製品を導入するため、RootDiskMonitor をインストールします。

・パッケージ名

clusterpro-mc-rdm-w.x.y-z.i386.rpm

※w, x, y, z は任意のバージョン番号が入ります

・インストール

```
# rpm -ivh /mnt/cdrom/Linux/rpm/clusterpro-mc-rdm-w.x.y-z.i386.rpm
```

※w, x, y, z は任意のバージョン番号が入ります

注意: Red Hat Enterprise Linux 7.x と 6.x (64bit) または Oracle Linux 7.x と 6.x (64bit) のシステムに本製品をインストールする場合、事前に互換ライブラリ(glibc-x-y.z.i686.rpm)がインストールされている必要があります。

・アンインストール

```
# rpm -e clusterpro-mc-rdm-w.x.y-z
```

※w, x, y, z は任意のバージョン番号が入ります

アンインストール時に /opt/HA ディレクトリは削除されません。不要な場合、手動で削除してください。

(3) 依存パッケージについて

- sg3_utils

RootDiskMonitor は内部で以下のパッケージを利用します。

sg3_utils Utils for Linux's SCSI generic driver devices + raw devices

本パッケージがインストールされていない場合、事前にインストールしてください。
以下のコマンドでインストールの有無を確認できます。

```
# rpm -qa sg3_utils  
sg3_utils-w.x.y.z
```

- ・ インストールされていない場合、何も出力されません

注意:w, x, y, z には sg3_utils パッケージのバージョン番号が入ります。

本パッケージは標準で OS インストール媒体中に含まれます。

- glibc (32bit 互換ライブラリ)

Red Hat Enterprise Linux 7.x と 6.x (64bit) または Oracle Linux 7.x と 6.x (64bit) のシステムに本製品をインストールする場合、事前に互換ライブラリ(glibc-x.y.z.i686.rpm)がインストールされている必要があります。

互換ライブラリがインストールされていない場合、事前にインストールしてください。
以下のコマンドでインストールの有無を確認できます。

```
# rpm -qa glibc  
:  
glibc-x.y.z.i686
```

- ・ インストールされていない場合、"glibc-x.y.z.i686" の行が出力されません。

注意:x, y, z には互換ライブラリのバージョン番号が入ります。

本パッケージは標準で OS インストール媒体中に含まれます。

(4) セットアップ

OS ディスクを監視するには、設定ファイルの作成が必要です。

設定ファイル名は、`/opt/HA/RDM/conf/rdm.config` です。

サンプルファイルが `/opt/HA/RDM/conf/rdm.config.default` として提供されていますので、このファイルをコピーした後に、OS ディスクを構成するデバイス情報を登録してください。

設定ファイル自動生成コマンド `/opt/HA/RDM/bin/rdmconfig(1M)` を利用するとデバイス情報を検索し設定ファイルのテンプレートを自動生成できます。

既に、設定ファイルが存在する場合は、上書き要否を問い合わせますので「Y/N」から Y を選択してください。N を選択すると、設定ファイル自動生成を中止します。

自動生成したテンプレートファイルについては、監視ルール、OS ディスクを構成するデバイス情報の妥当性を確認してください。

なお、OS ディスク以外については自動生成対象とはなりませんので、手動で設定ファイルを編集してください。

注意:

- マルチパスデバイスを使用した SANBoot 構成は設定ファイル自動生成 (rdmconfig) のサポート対象外です。
以下の手動で設定ファイルを作成してください。

① OS ディスクのデバイスファイル名を確認

```
# df
Filesystem                1K-blocks    Used Available Use% Mounted on
/dev/mapper/mpatha2         8181760    874112   7307648   11% /
/dev/mapper/mpatha1         201388     84216   117172   42% /boot
```

② マルチパスデバイスを使用している場合は、対応する sd デバイスファイル名、udev デバイスファイル名を確認

(例) Device Mapper Multipath のデバイスを使用している場合

```
# multipath -ll /dev/mapper/mpatha
size=20G features='0' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=1 status=active
|  `-- 7:0:0:0 sdb 8:16   active ready running
`-+- policy='service-time 0' prio=1 status=enabled
    `-- 8:0:0:0 sdl 8:176 active ready running
```

```
# ls -l /dev/disk/by-path
lrwxrwxrwx 1 root root 9 Mar 9 18:03 pci-0000:15:00.0-fc-0x220000255c3a0266-lun-0 -> ../../sdb
lrwxrwxrwx 1 root root 9 Mar 9 18:03 pci-0000:15:00.1-fc-0x2a0000255c3a0266-lun-0 -> ../../sdl
```

③ 設定ファイルのサンプルファイルをコピー

```
# cp -p /opt/HA/RDM/conf/rdm.config.default /opt/HA/RDM/conf/rdm.config
```

④ 設定ファイルの編集

【CLUSTERPRO MC StorageSaver for BootDisk 2.0 以降の場合】

```
# vi /opt/HA/RDM/conf/rdm.config
<中略>
VG          VG_NONE
MIRROR      group01
## PV Name: /dev/sdb
PV          pci-0000:15:00.0-fc-0x220000255c3a0266-lun-0
## PV Name: /dev/sdl
PV          pci-0000:15:00.1-fc-0x2a0000255c3a0266-lun-0
```

【CLUSTERPRO MC StorageSaver for BootDisk 1.2 以前の場合】

```
# vi /opt/HA/RDM/conf/rdm.config
<中略>
VG          VG_NONE
MIRROR      group01
PV          /dev/sdb
PV          /dev/sdl
```

(5) 設定ファイルの変更

- ・ CLUSTERPRO との連携方式によって、設定ファイルの変更が必要です。
カスタムモニタリソースによる CLUSTERPRO との連動を使ったノード切り替えを行う場合は、下記のパラメータを変更してください。

パラメータ名 : OVER_ACTION
設定値 : SERVICE_CMD_DISABLE を SERVICE_CMD_ENABLE へ変更

CLUSTERPRO のサーバ管理プロセス (clpnm) を強制終了する手法によるノード切り替えを行う場合は、下記のパラメータを変更してください。

パラメータ名 : OVER_ACTION
設定値 : SERVICE_CMD_DISABLE を CLPNM_KILL へ変更

OS ディスク故障時に、システムメモリダンプを採取し OS を強制終了する手法によるノード切り替えを行う場合は、下記のパラメータを変更してください。

パラメータ名 : OVER_ACTION
設定値 : SERVICE_CMD_DISABLE を TOC_EXEC へ変更

CLUSTERPRO と連携したノード切り替えを行わない場合は OVER_ACTION の変更は不要です。

- ・ 仮想環境 (ゲスト OS) で本製品を使用する場合は、設定ファイルの変更が必要です。
仮想環境 (ゲスト OS) で本製品を使用する場合は、下記のパラメータを変更してください。

パラメータ名 : TESTIO_MODE
設定値 : INQ_TUR を READ へ変更

- ・ Hyper-V 上のゲスト OS で本製品を使用する場合は、設定ファイルに udev デバイスファイル名を定義することができません。
設定ファイルを自動生成する際、以下のコマンドを実行してください。

```
# /opt/HA/RDM/bin/rdmconfig -sd
```

(6) プロセスの再起動

① モニタプロセスの再起動

設定ファイルを作成した後にモニタプロセスを再起動してください。

【Red Hat Enterprise Linux 7.x】

【Oracle Linux 7.x】

```
# systemctl stop rdmd  
# systemctl start rdmd
```

【Red Hat Enterprise Linux 6.x と 5.x】

【Oracle Linux 6.x】

```
# /etc/init.d/rdmd stop  
# /etc/init.d/rdmd start
```

② モニタプロセスの確認

モニタプロセスが起動されていることを確認してください。

```
# ps -ef|grep rdm  
root 2169      1  0 10:43:40 ?  0:00 /opt/HA/RDM/bin/rdmdiagd  
root 2179  2169  0 10:43:40 ?  0:00 rdmping
```

③ ステータスの確認

監視リソースの状態を確認してください。

```
# /opt/HA/RDM/bin/rdmdadmin
```

3.2. 設定ファイルの記述

(1) 設定ファイルの設定について

設定ファイル名は以下の通りです。

`/opt/HA/RDM/conf/rdm.config`

以下に使用するキーワードを記述します。

監視ルール	
項目	説明
TIME_VG_FAULT	監視リソースを異常と判定する時間を指定します。 このパラメータはデフォルト値を使用することを推奨します。 指定値は 30 秒～, デフォルト 60 秒
TIME_VG_STALL	監視リソースの I/O ストールを判定する時間を指定します。 このパラメータはデフォルト値を使用することを推奨します。 指定値は 60 秒～, デフォルト 360 秒 なお、0 秒を指定すると I/O ストール監視を行いません。
TIME_TESTIO_INTERVAL	Test I/O の発行間隔を指定します。 このパラメータはデフォルト値を使用することを推奨します。 指定値は 1 秒～, デフォルト 5 秒
WAIT_TESTIO_INTERVAL	Test I/O でパススルードライバに指定する I/O 待ち合わせ時間を指定します。 このパラメータはデフォルト値を使用することを推奨します。 指定値は 1 秒～, デフォルト 5 秒
OVER_ACTION	OS ディスク異常検出時のアクションを指定します。 VG 単位に異常検出時のアクションを制御する場合は後述の VOLTYPE パラメータを指定してください。
	SERVICE_CMD_DISABLE アクション指定なし、デフォルトです。 syslog にメッセージを出力します。
	SERVICE_CMD_ENABLE クラスタウェア連携用デーモン (rdmstat)を使用してクラスタウェア 連携を行うことで、ノードを切り替えま す。
	CLPNM_KILL CLUSTERPRO のサーバ管理プロ セス (clpnm) を強制終了することで ノードを切り替えます。 クラスタウェア連携時にはこの設定 を推奨します。
	TOC_EXEC システムメモリダンプを採取し、OS を強制停止することでノードを切り替 えます。
	POWER_OFF ソフトウェア watchdog を利用し、OS を停止します。

VG_STALL_ACTION	I/O ストール検出時のアクションを指定します。 ※本パラメータは変更しないことを推奨します。	
	SERVIE_CMD_DISABLE	アクション指定なし、デフォルトです。 I/O ストールを検出すると、syslog にメッセージを出力します。
	SERVICE_CMD_ENABLE	クラスタウェア連携用デーモン(rdmstat)を使用してクラスタウェア連携を行うことで、ノードを切り替えます。
	CLPNM_KILL	CLUSTERPRO のサーバ管理プロセス(clpnm)を強制終了することでノードを切り替えます。
	TOC_EXEC	システムメモリダンプを採取し、OS を強制停止することでノードを切り替えます。
	POWER_OFF	ソフトウェア watchdog を利用し、OS を停止します。
TESTIO_MODE	TestI/O の発行方法を指定します。 TUR_TESTIO_USE は本パラメータに統合されました。 一部ディスク装置 (EMC 社製 CLARiX シリーズ、VNX シリーズ、NEC 社製 iStorageM シリーズ等) は TestUnitReady の発行を許可していません。 そのようなディスク装置を監視する場合は、本パラメータを INQ に設定することで TestUnitReady の発行を抑制します。 ※本パラメータは特別な要件がない限り変更しないことを推奨します。	
	INQ	パススルードライバ経由で Inquiry を発行します。
	INQ_TUR	パススルードライバ経由で Inquiry と TestUnitReady を発行します。デフォルトです。
	INQ_TUR_READ	パススルードライバ経由で Inquiry と TestUnitReady と Read を発行します。
	DIRECT	設定ファイルに定義されたデバイスファイルに対して直接 Inquiry と TestUnitReady を発行します。
	READ	設定ファイルに定義されたデバイスファイルに対して直接 read を発行します。 仮想環境 (ゲスト OS) で使用する場合は、この値を指定します。

デバイス定義	
項目	説明
VG	OS ディスクのボリューム名を指定します。 LVM 構成の場合は VG 名 (/dev/VolGroup00 など)を指定します。 物理ディスク構成の場合はダミーの VG 名として VG_NONE を指定します。 複数の VG を監視する場合は VG-MIRROR-PV の組み合わせを指定してください。
VOLTYPE ※通常は指定しない。省略可	ディスクの種別を指定します。 ※本パラメータは通常指定する必要はありません。OVER_ACTION の動作を変更する必要がある場合は指定しないでください。
	ROOT_VOLUME 通常の OS ディスクの場合に指定します。また、指定されていない場合のデフォルトです。 VG ダウンを検出した場合に通常通り OVER_ACTION の動作を実行します。
	OTHER OS ディスク以外のデータディスクの場合に指定します。OTHER が指定された VG は、VG ダウンを検出した場合でも OVER_ACTION の動作を実行せず、syslog への通報のみとなります。
MIRROR	OS ディスクのミラーグループを構成する組み合わせを指定します。 指定値は 16 文字以内のノード一意である任意の英数字です。 設定ファイル自動生成コマンドを使用すると、groupxy(xy は 01 からの連番)を登録します。 OS ディスクの中で同一データを構成する I/O パスをグルーピングします。
PV	I/O パスへの udev デバイスファイルまたはスペシャルファイルを指定します。 スペシャルファイルはブロック型のファイル名を指定します。(例えば /dev/sda など)物理ディスクをパーティション分割して利用している場合は、パーティション名の指定ではなく物理ディスク名を指定してください。 ■ シングルパス構成の場合は、LUN の udev デバイスファイルまたはスペシャルファイルをひとつ指定してください。 ■ ソフトミラー構成の場合は、正副 LUN 両方の udev デバイスファイルまたはスペシャルファイルを指定してください。 ■ FC 接続の代替パス構成の場合は、LUN への正副 I/O パス両方の udev デバイスファイルまたはスペシャルファイルを指定してください。 設定ファイルにスペシャルファイル名と udev デバイスファイル名の両方を指定した場合、デーモンプロセス起動時にスペシャルファイル名の妥当性チェックを行います。 スペシャルファイル名が変更されていた場合には、本機能により当該デバイスを監視対象外とします。 これにより、誤った設定のまま監視を開始して障害を誤検知することを防止します。 スペシャルファイル名と udev デバイスファイル名の間にはスペースが必要です。

注意:

- ・ POWER_OFF 機能を利用する場合は softdog モジュールが必要となります。
事前に softdog モジュールがインストールされていることを確認してください。

- ・ TOC_EXEC 機能、POWER_OFF 機能を利用する場合は事前に kdump の設定を行い、SysRq キーを発行することによりカーネルパニックが発生することを確認してください。
- ・ POWER_OFF 機能を利用する場合、ソフトウェア watchdog を利用した OS 停止に失敗した際には自動的に TOC_EXEC 機能にて OS を停止させます。
- ・ CLUSTERPRO を導入し、クラスタを起動している場合は CLUSTERPRO によるソフトウェア watchdog を利用した OS 停止機能を優先させるため、RootDiskMonitor の POWER_OFF 機能を利用することができません。どちらも同等の機能のため、CLUSTERPRO の softdog 機能を利用してください。また、CLUSTERPRO と他の方式にて連携を行う場合は、前述の "2.4 クラスタウェアとの連携について" を参照してください。
- ・ EMC 社製 CLARiX シリーズ、VNX シリーズ、NEC 社製 iStorage M シリーズを使用した SANBoot 構成の場合には TESTIO_MODE の設定を INQ にしてください。
- ・ 仮想環境 (ゲスト OS) で本製品を使用する場合は、TESTIO_MODE 値に READ を設定してください。
- ・ 上記タイム値の上限値は MAXINT まで指定可能ですが、常識的な運用での適用を推奨します。
- ・ デバイスファイルのチェック機能を使用する場合は、"/opt/HA/RDM/local" ディレクトリが存在するか確認してください。存在しない場合は、該当のディレクトリを作成してください。

(2) 設定ファイルの設定例について

rdm.config (RootDiskMonitor (Linux) Configuration)

#####

System Config Area

#####

Test/I/O interval timer for Root Volume is failed (seconds)

Root Volume status changes fail between this timer

minimum = 30, default = 60

TIME_VG_FAULT 60

I/O stall interval timer for Root Volume is failed (seconds)

Root Volume status changes fail between this timer

minimum = 60, default = 360. 0 mean I/O stall no check.

TIME_VG_STALL 360

Test/I/O interval timer value (seconds)

exec normal Test/I/O for PV between this timer

minimum = 1, default = 5

TIME_TESTIO_INTERVAL 5

Wait I/O for sg driver timer value (seconds)

wait Test/I/O between this timer

minimum = 1, default = 5

WAIT_TESTIO_INTERVAL 5

CLUSTERPRO のサーバ管理プロセス
(clpnm) 強制停止によるノード切り替えを
行う事例

Root Volume fault action

select SERVICE_CMD_DISABLE(default),SERVICE_CMD_ENABLE,TOC_EXEC,CLPNM_KILL,POWER_OFF
OVER_ACTION CLPNM_KILL

変更しないことを推奨

Root Volume stall find action

select SERVICE_CMD_DISABLE(default),SERVICE_CMD_ENABLE,TOC_EXEC,CLPNM_KILL,POWER_OFF
VG_STALL_ACTION SERVICE_CMD_DISABLE

Test/I/O mode

select INQ,INQ_TUR_READ,READ,DIRECT,INQ_TUR(default)

TESTIO_MODE INQ_TUR

#####

Device Config Area

#####

VG volume group for LVM(VolGroup00 , etc)

MIRROR PV set for Mirror(set any number)

PV PV Block Device File(/dev/sda , etc)

自動生成にて定義されるフォーマット

VG /dev/VolGroup00

MIRROR group01

PV pci-0000:0c:00.0-scsi-0:2:0:0

ブロックデバイスファイル名を指定する場
合の定義例

MIRROR group02

PV /dev/sda

MIRROR group03

PV /dev/sdb pci-0000:0c:00.0-scsi-0:2:0:0

ブロックデバイスファイルのチェック機能
を利用する場合の定義例
ブロックデバイスファイル名と udev デ
バイスファイル名を定義します

4. 操作・運用手順

4.1. 運用管理コマンドの操作手順

- (1) リソース監視の状態を表示します。

# /opt/HA/RDM/bin/rdmadmin (monitor status = TRUE)				リソース監視の on/off を表示します
=====				
type	: H/W Path	: Logical : status	: I/O : status	VG レベルの監視状態を 表示します
=====				
VG	: /dev/VolGroup00	: up		
MIRROR	:	:		
PV	: /dev/sda	: up	: up	I/O パスの論理・物理状態を 表示します
MIRROR	:	:		
PV	: /dev/sdb	: up	: up	

- (2) ディスクに対するすべての I/O パスが異常になると VG レベルのステータスもダウン状態になります。

以下の例ではシングル構成のため、/dev/sda が故障した時点で VG ダウンとなります。

# /opt/HA/RDM/bin/rdmadmin (monitor status = TRUE)				
=====				
type	: H/W Path	: Logical : status	: I/O : Status	すべての経路が障害となるため VG レベルで down となります
=====				
VG	: /dev/VolGroup00	: down		
MIRROR	:	:		
PV	: /dev/sda	: down	: down	障害を検出
MIRROR	:	:		
PV	: /dev/sdb	: up	: up	

- (3) 内蔵ディスクがソフトミラー構成で冗長化されている場合は、片系の I/O パスのみが異常になると PV レベルのステータスがダウン状態になります。

このとき VG レベルでは縮退状態で運用されているため、VG のステータスは suspend になります。

```
# /opt/HA/RDM/bin/rmdadmin
(monitor status = TRUE)
```

```
=====
type      :                               : Logical   : I/O
          : H/W Path                     : status    : status
=====
VG         : /dev/VolGroup00              : suspend   :
MIRROR     :                               :
PV         : /dev/sda                     : down      : down
PV         : /dev/sdb                     : up        : up
MIRROR     :                               :
PV         : /dev/sdc                     : up        : up
PV         : /dev/sdd                     : up        : up
```

ソフトミラーの片系が
障害となった状態です

- (4) リソース監視の停止と再開についてリソース監視を一時的に停止および再開する場合は以下のコマンドで行います。

```
# /opt/HA/RDM/bin/rdmadmin -c stop
Change TESTIO.
START -> STOP
```

```
# /opt/HA/RDM/bin/rdmadmin -c start
Change TESTIO.
STOP -> START
```

なお、リソース監視停止中は、モニタステータスが FALSE になります。

```
# /opt/HA/RDM/bin/rdmadmin
(monitor status = FALSE)
```

type	:	H/W Path	:	Logical status	:	I/O status
VG	:	/dev/VolGroup00	:	up	:	
MIRROR	:		:		:	
PV	:	/dev/sda	:	up	:	up
MIRROR	:		:		:	
PV	:	/dev/sdb	:	up	:	up

FALSE になります

- (5) 3 秒間隔でリソースの状態を定期表示します。

```
# /opt/HA/RDM/bin/rdmadmin -c status -t 3
(monитор status = TRUE)
```

```
=====
type      :      : Logical   : I/O
           : HW Path : status   : status
=====
VG         : /dev/VolGroup00 : up
MIRROR     :               :
PV         : /dev/sda       : up      : up
MIRROR     :               :
PV         : /dev/sdb       : up      : up
```

<... 3秒経過 ...>

```
(monитор status = TRUE)
```

```
=====
type      :      : Logical   : I/O
           : HW Path : status   : status
=====
VG         : /dev/VolGroup00 : up
MIRROR     :               :
PV         : /dev/sda       : up      : up
MIRROR     :               :
PV         : /dev/sdb       : up      : up
```

(注)コマンドを終了させたい場合、ctrl+c で終了できます。

(6) コンフィグレーション情報を表示します。

```
# /opt/HA/RDM/bin/rdmadmin -c param
SG parameters.
-----
TIME_VG_FAULT          60
TIME_VG_STALL          360
TIME_TESTIO_INTERVAL   5
WAIT_TESTIO_INTERVAL   5
OVER_ACTION            SERVICE_CMD_DISABLE
VG_STALL_ACTION         SERVICE_CMD_DISABLE
TUR_TESTIO_USE          TRUE
TESTIO_DIRECT           INQ_TUR
testioYN                TRUE
TESTIO_MODE             INQ_TUR

SG device list.
-----
SYSTEM
VG          /dev/VolGroup00
VOLTYPE     ROOT_VOLUME
MIRROR
PV          /dev/sda
MIRROR
PV          /dev/sdb
```

(7) モニタプロセスの起動、終了

- rc ファイルからの起動、終了

OS 起動(boot)を契機に自動起動、OS 終了を契機に自動終了されます。

デフォルトで自動起動が設定されるのはランレベル3および5です。

- 手動起動、終了

以下のコマンドを root 権限で投入することで起動できます。

【Red Hat Enterprise Linux 7.x】

【Oracle Linux 7.x】

```
# systemctl start rdmd
```

【Red Hat Enterprise Linux 6.x と 5.x】

【Oracle Linux 6.x】

```
# /etc/init.d/rdmd start
```

以下のコマンドを root 権限で投入することで終了できます。

【Red Hat Enterprise Linux 7.x】

【Oracle Linux 7.x】

```
# systemctl stop rdmd
```

【Red Hat Enterprise Linux 6.x と 5.x】

【Oracle Linux 6.x】

```
# /etc/init.d/rdmd stop
```

上記コマンドで終了しない場合は、ps -ef|grep rdm で rdm から始まるプロセスの pid を検索して、kill -9 <pid>で終了させてください。

また、以下のコマンドを root 権限で投入することで、デーモンプロセスを再起動できます。

【Red Hat Enterprise Linux 7.x】

【Oracle Linux 7.x】

```
# systemctl restart rdmd
```

【Red Hat Enterprise Linux 6.x と 5.x】

【Oracle Linux 6.x】

```
# /etc/init.d/rdmd restart
```

- (8) デバッグ機能を利用するとディスク障害を擬似できます。

設定ファイルの設定値の正当性を検証するためにコマンドオペレーションでディスク障害を擬似できます。物理ディスクの抜き差し等の操作をする必要がないためシステムへの影響を与えず評価が実現できます。

なお、本機能は開発用の機能ですので、サポート対象にはなりませんので御承知おきください。

/opt/HA/RDM/bin/rdmadmin -c debug -v on/off [-f SpecialFile]

off -> I/O status modify up

on -> I/O status modify down

```
# /opt/HA/RDM/bin/rdmadmin -c debug -v on -f /dev/sda
(monitor status = TRUE)
Change debug value.
special file = /dev/sda
0 -> 1
```

```
# /opt/HA/RDM/bin/rdmadmin
(monitor status = TRUE)
```

```
=====
type      :                               : Logical   : I/O
           : HW Path                     : status    : status
=====
VG         : /dev/VolGroup00              : up
MIRROR    :                               :
PV        : /dev/sda                      : down      : down
MIRROR    :                               :
PV        : /dev/sdb                      : up        : up
```

強制的にメモリ上の
ステータスを塗り替える

一定の時間が経過すると VG ステータスも異常値に変わります。

ミラー構成の場合、両方のスペシャルファイルを down 状態に変更すると
ノード切り替えの評価が可能です。

4.2. カスタムモニタリソースによる CLUSTERPRO との連携

(1) CLUSTERPRO との連携について

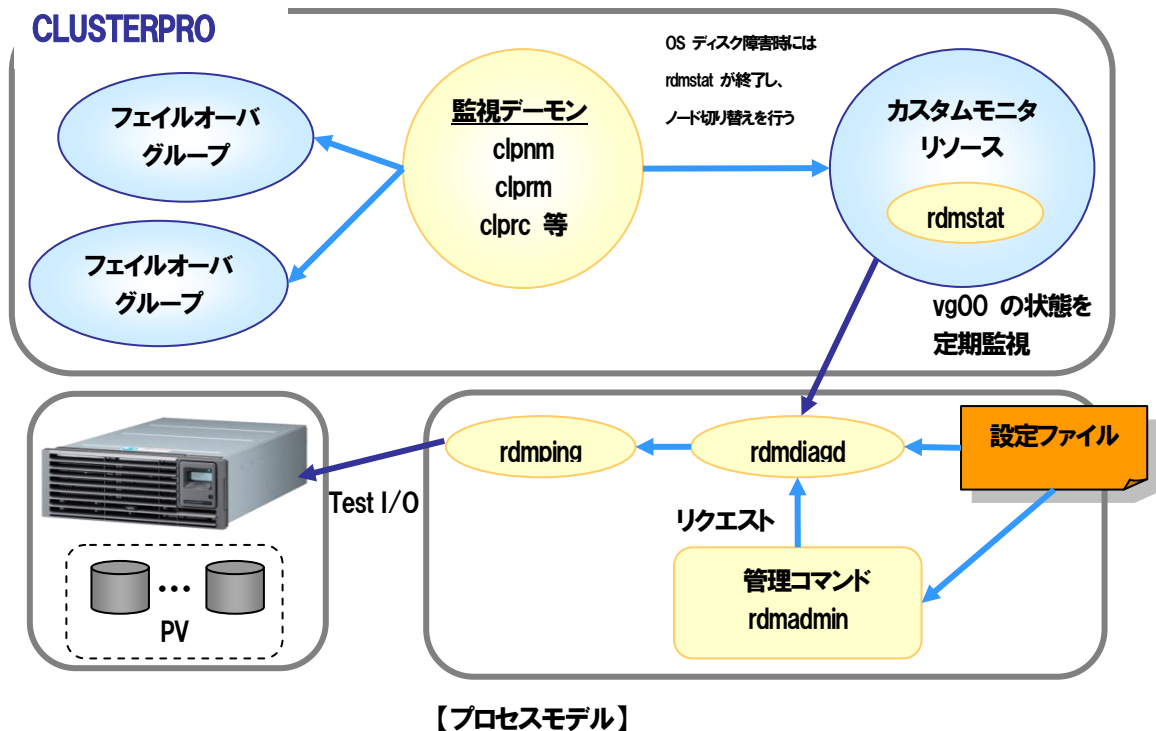
OS ディスクの動作状態をモニタするコマンド `rdmstat(1M)` を CLUSTERPRO のカスタムモニタリソースとして登録することで、OS ディスクの障害時のノードダウン、ノード切り替えを実現します。

本機能を利用する場合は、不必要に CLUSTERPRO のサーバ管理プロセス (`clpnm`) を kill しないために RootDiskMonitor のコンフィグレーションの **OVER_ACTION** には、**SERVICE_CMD_ENABLE** を指定してください。

この方式であれば、複数ノードクラスタシステムでのノード切り替えだけでなく縮退した状態でのノードダウンや1ノードのクラスタシステムでのノードダウンを実現できますので、非常に有用な手法です。

RootDiskMonitor としては本方式を推奨します。

本方式で連携する場合、障害時に確実にフェイルオーバーできるよう、フェイルオーバー時の CLUSTERPRO の動作設定は「クラスタサービス停止と OS シャットダウン」を選択してください。「クラスタサービス停止と OS シャットダウン」を選択していない場合、I/O パス障害の影響でフェイルオーバー処理が正しく完了せず、フェイルオーバーに失敗したり、フェイルオーバー完了が遅延したりする場合があります。



(2) rdmstat の運用について

OS ディスクに障害が発生すると、rdmdiagd(1M)が I/O パスおよび VG レベルの管理ステータスを down 状態に変更し、syslog、コンソールにエラーメッセージを出力します。

rdmstat(1M)は共有メモリを経由して VG レベルの管理ステータスをモニタします。
VG が down 状態に遷移した時点で、rdmstat(1M)は異常終了し、CLUSTERPRO が
カスタムモニタリソースのダウンを検出しノード切り替え、ノードダウンが発生します。

rdmstat(1M)は、rdmdiagd(1M)および rdmping(1M)のプロセスが起動され、OS ディスクの
監視を行っている場合に有効に機能します。

以下のようなリソース監視を停止している場合は、VG 障害を検出できません。

- rdmdiagd(1M)および rdmping(1M)のプロセスが起動していない。
- rdmdadmin(1M)のオペレーション操作でリソース監視停止を指示している。

<syslog メッセージの出力例>

LVM かつシングルディスク構成での事例です。

下記の順序で syslog にメッセージが出力されます。

OS ディスクへのファイル I/O が停止すると、syslog に記録されない場合もあります。

- OS ディスクの障害を検出し VG ステータスをダウンに変更します。
Jan 10 18:05:29 node1 rdm[24031]: PV down find .(sf=/dev/sda)
Jan 10 18:05:29 node1 rdm[24031]: VG status change down .(vg=/dev/VolGroup00)

また、LVM かつソフトウェアミラー構成の例は以下となります。

- 最初に、ソフトミラーの片系障害(正系 LUN 障害)を検出
Jan 10 18:03:28 node1 rdm[24031]: PV down find .(sf=/dev/sda)
Jan 10 18:03:28 node1 rdm[24031]: VG status change suspend .(vg=/dev/VolGroup00)
- 次に、ソフトミラーの両系障害(副系 LUN 障害)を検出
Jan 10 18:05:29 node1 rdm[24031]: PV down find .(sf=/dev/sdb)
Jan 10 18:05:29 node1 rdm[24031]: VG status change down .(vg=/dev/VolGroup00)

(3) カスタムモニタリソースの設定手順について

rdmstat を CLUSTERPRO のカスタムモニタリソースに登録する手順については、後述の

「8.2 カスタムモニタリソースによる CLUSTERPRO との連携手順」

を参照してください。

4.3. CLUSTERPRO のサーバ管理プロセス (clpnm) の強制終了による

CLUSTERPRO との連携

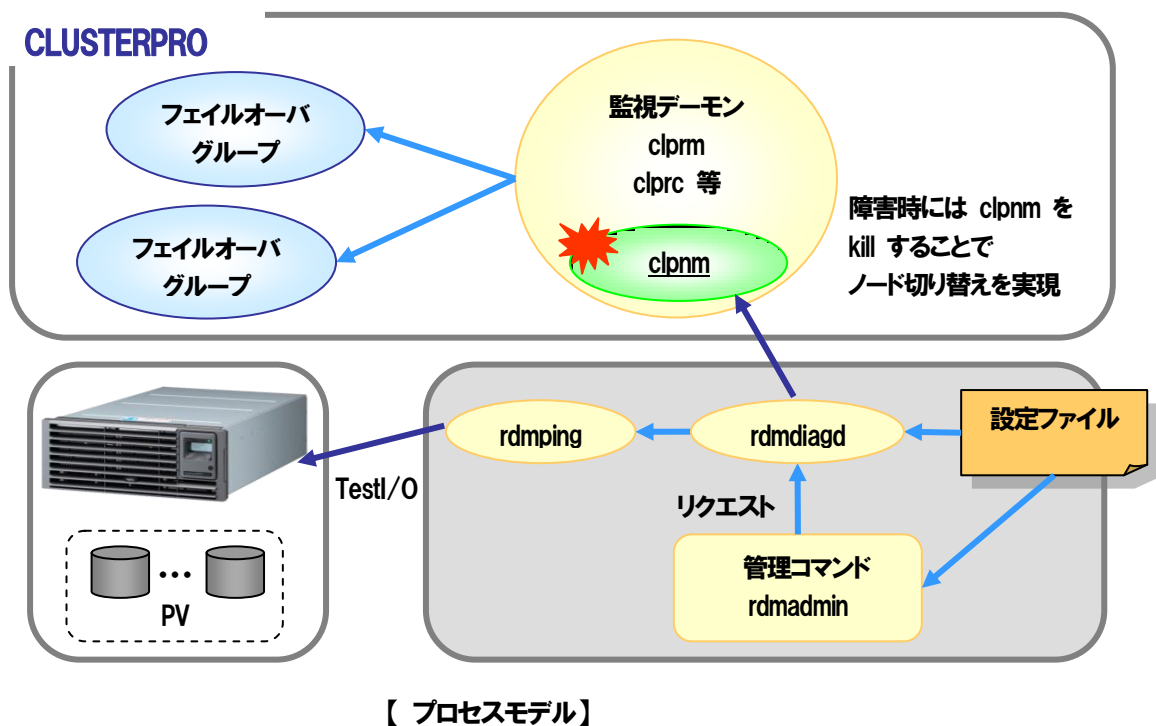
(1) CLUSTERPRO との連携について

OS ディスクの障害時に、CLUSTERPRO のサーバ管理プロセス (clpnm) を強制終了させることで、ノード切り替えを実現する方式です。

RootDiskMonitor のコンフィグレーションである **OVER_ACTION** に **CLPNM_KILL** を指定すると、OS ディスクの障害時に CLUSTERPRO のサーバ管理プロセス (clpnm) を kill することができます。

この方式であれば、内蔵ディスク障害が発生した場合でも、確実にノード切り替えが実現可能です。

このノード切り替え機能は、2ノード以上のクラスタシステムで有効です。



(2) OS ディスクの障害を検出すると

OS ディスクに障害が発生すると、rdmdiaqd(1M)が I/O パスおよび VG レベルの管理ステータスを down 状態に変更し、syslog、コンソールにエラーメッセージを出力します。

rdmdiaqd(1M)はコンフィグレーションの OVER_ACTION に CLPNM_KILL を指定していると CLUSTERPRO のサーバ管理プロセス (clpnm) を強制終了させ、CLUSTERPRO によるノード切り替えを行います。
また、rdmdiaqd(1M)自身も abort します。

/opt/HA/RDM/log 配下には、core ファイルやトレースファイルを出力しますので、OS ディスク故障時にこれらのファイルを確認してください。
なお、OS ディスク故障時には、ファイル I/O が失敗する可能性がありますので core ファイル等が残っていないケースもあります。

<syslog メッセージの出力例>

LVM かつシングルディスクの構成の例です。

下記の順序で syslog にメッセージが出力されます。

OS ディスクへのファイル I/O が停止すると、syslog に記録されない場合もあります。

● OS ディスクの障害を検出し VG ステータスをダウンに変更

Jan 10 18:05:29 node1 rdm[24031]: PV down find .(sf=/dev/sda)

Jan 10 18:05:29 node1 rdm[24031]: VG status change down .(vg=/dev/VolGroup00)

● VG へのアクセス不可を検出し、予備ノードへ切り替え

Jan 10 18:05:29 node1 rdm[24031]: start KILL clpnm.

Jan 10 18:05:29 node1 rdm[24031]: send signal clpnm.

Jan 10 18:05:29 node1 rdm[24031]: abort rdmdiaqd.

また、LVM かつソフトウェアミラー構成の例は以下となります。

● 最初に、ソフトミラーの片系障害(正系 LUN 障害)を検出

Jan 10 18:03:28 node1 rdm[24031]: PV down find .(sf=/dev/sda)

Jan 10 18:03:28 node1 rdm[24031]: VG status change suspend .(vg=/dev/VolGroup00)

● 次に、ソフトミラーの両系障害(副系 LUN 障害)を検出

Jan 10 18:05:29 node1 rdm[24031]: PV down find .(sf=/dev/sdb)

Jan 10 18:05:29 node1 rdm[24031]: VG status change down .(vg=/dev/VolGroup00)

● VG へのアクセス不可を検出し、予備ノードへ切り替え

Jan 10 18:05:29 node1 rdm[24031]: start KILL clpnm.

Jan 10 18:05:29 node1 rdm[24031]: send signal clpnm.

Jan 10 18:05:29 node1 rdm[24031]: abort rdmdiaqd.

4.4. システムメモリダンプ採取と OS 強制停止による CLUSTERPRO との連携

(1) CLUSTERPRO との連携について

OS ディスク故障時にシステムメモリダンプの採取と OS 強制停止 (panic) により CLUSTERPRO と連携してノード切り替えを実現します。

OS ディスク故障時には OS やその他監視製品なども正常に動作できない場合がありますので、この方式による OS 強制停止でノード切り替えを行うことは有効です。

また、システムメモリダンプが採取されますので、障害状態の解析なども可能です。

(注) 内蔵ディスクの故障パターンによっては正しくシステムメモリダンプが採取できない場合があります。システムメモリダンプが採取できない場合も、CLUSTERPRO が予備ノードから現用ノードの異常を検出しますので系切り替えは可能です。

本機能を使ってシステムメモリダンプを採取する場合には、あらかじめ kdump の設定が完了している必要があります。また、システムメモリダンプは /var/crash 配下に作成されます。

万が一システムメモリダンプ採取に失敗した場合、続いて CLUSTERPRO のサーバ管理プロセス (clpnm) を強制停止することで待機系への切り替えを試みます。

5. syslog メッセージ

5.1. syslog に出力するメッセージについて

本製品では、リソース監視で致命的な異常を検出すると syslog にメッセージを出力します。

syslog の facility と level は以下の通りです。

facility: LOG_DAEMON

level : LOG_ALERT、LOG_ERR、LOG_WARNING

LOG_ALERT、LOG_ERR は、ハードエラーなどの致命的な障害の場合に使用されます。

5.2. 警報対象として登録することを推奨するメッセージ一覧

特に重要度の高い syslog メッセージを記述します。(下線部はメッセージの固定部分を示します)
これらのメッセージが出力された場合は、HW 保守担当者に HW 検査を依頼してください。

- (1) Test I/O のリソース監視で異常を検出した場合

LOG_ERROR の出力契機は以下の通りです。

PV down find .(sf=/dev/sdx)

説明: Test I/O で PV レベルの異常を検出

処置: I/O パス異常を検出しましたので、早急に該当ディスクの点検を行ってください。

シングル構成の場合は、その後 VG レベルも down になります。

ミラー構成の場合は、VG レベルが suspend になります。

VG status change suspend .(vg=/dev/VolGroupXX)

説明: Test I/O で VG レベルの異常(suspend)を検出

処置: ミラー構成の片系のディスクが障害となっています。

早急に該当ディスクの点検を行ってください。

VG status change down .(vg=/dev/VolGroupXX)

説明: Test I/O で VG レベルの異常(down)を検出

処置: ミラー構成の両系のディスクが障害となっています。

早急に該当ディスクの点検を行ってください。

VG status change up .(vg=/dev/VolGroupXX)

説明: Test I/O で VG レベルの復旧(up)を検出

処置: ディスクの障害から復旧しました。

ディスクの監視を開始していますので、特に必要ありません。

I/O stall find , timeover occurred (sf=/dev/sdx)

説明: Test I/O で I/O ストールタイムオーバを検出

処置: ディスクが故障している可能性があります。

早急に該当ディスクの点検を行ってください。

(2) コンフィグレーション関係

LOG_ERROR の出力契機は以下の通りです。

ConfigFile error(xxx).

説明: 設定ファイルの設定誤りを検出

処置: 設定ファイル(/opt/HA/RDM/conf/rdm.config)を確認してください。

記述例は『3.2 設定ファイルの記述』を参照してください。

(3) システムメモリダンプ採取と OS 強制停止による CLUSTERPRO 連携

LOG_ALERT の出力契機は以下の通りです。

start kernel system dump.

説明: システムメモリダンプの採取と OS 強制停止を開始

処置: 内蔵ディスクの障害により、クラスタ系切り替えが発生しています。

早急に該当ディスクの点検を行ってください。

(4) CLUSTERPRO のサーバ管理プロセス (clpnm) 強制終了による CLUSTERPRO 連携

LOG_ALERT の出力契機は以下の通りです。

start KILL clpnm.

説明:CLUSTERPRO のサーバ管理プロセス (clpnm) 強制終了を開始

処置:内蔵ディスクの障害により、クラスタ系切り替えが発生しています。

早急に該当ディスクの点検を行ってください。

send signal clpnm.

説明:CLUSTERPRO のサーバ管理プロセス (clpnm) 強制終了を完了

処置:内蔵ディスクの障害により、クラスタ系切り替えが発生しています。

早急に該当ディスクの点検を行ってください。

abort rdmdiagd.

説明:rdmdiagd が abort しました

処置:内蔵ディスクの障害により、クラスタ系切り替えが発生しています。

早急に該当ディスクの点検を行ってください。

5.3. その他のメッセージ一覧

(1) ライセンス管理

LOG_ERROR の出力契機は以下の通りです。

Activation failed. Product key name is not been entry.

説明: 認証に失敗しました。有償ロックキーが入力されていません。

処置: ライセンスファイルに有償ロックキーを入力してください。

Activation failed. Code word is generated by different product key name.

説明: 認証に失敗しました。有償ロックキーが一致していません。

処置: 発行されたコードワードが正しく登録できていることを確認してください。

Activation failed. Code word is generated by different host ID.

説明: 認証に失敗しました。ホスト情報が一致していません。

処置: 発行されたコードワードが正しく登録できていることを確認してください。

Activation failed. Trial term is expired.

説明: 認証に失敗しました。試用期限を過ぎています。

処置: 正式版のライセンスを登録してください。

5.4. 運用管理製品との連携

本製品は、運用管理製品と連携し syslog メッセージを監視することができます。

これにより、syslog に出力される重要なログをアラートとしてリアルタイムで通知でき、障害発生時も早急な発見、迅速な対応が可能になります。

本製品で連携可能な運用管理製品は、以下となります。

◆ WebSAM SystemManager

RootDiskMonitor が異常を検知し syslog にその内容が出力されると、

WebSAM SystemManager のログ監視機能にて通知が行われます。

※ 連携手順については、「CLUSTERPRO MC (HA シリーズ) WebSAM SystemManager メッセージ監視連携手順書」を参照してください。

◆ ESMPRO/ServerAgent

RootDiskMonitor が異常を検知し syslog にその内容が出力されると、

ESMPRO/ServerAgent のアラート通報機能にて通知が行われます。

※ 連携手順については、「ESMPRO/ServerAgent アラート通報機能との連携手順」を参照してください。

6. 注意・制限事項について

6.1. 注意・制限事項

(1) 下記の注意事項があります。

- 本製品を利用する場合、Linux SCSI パススルードライバ(sg ドライバ)がインストールされ、事前にカーネルモジュールとしてロードされている必要があります。

モジュールがロードされているかどうかは lsmod(8)コマンドの結果から確認できます。
以下の sg ドライバのエントリ行が出力されることを確認してください。

```
# lsmod | grep sg
sg                38369  2  sg ドライバのエントリ行
```

- RootDiskMonitor は内部で以下のパッケージを利用します。

sg3_utils Utils for Linux's SCSI generic driver devices + raw devices

本パッケージがインストールされていない場合、事前にインストールしてください。
以下のコマンドでインストールの有無を確認できます。

```
# rpm -qa sg3_utils
sg3_utils-w.x.y.z
```

- ・ インストールされていない場合、何も出力されません

注意:w, x, y, z には sg3_utils パッケージのバージョン番号が入ります。

本パッケージは標準で OS インストール媒体中に含まれます。

- Red Hat Enterprise Linux 7.x と 6.x (64bit) または Oracle Linux 7.x と 6.x (64bit) のシステムに本製品をインストールする場合、事前に互換ライブラリ(glibc-x.y.z.i686.rpm)がインストールされている必要があります。

互換性ライブラリがインストールされていない場合、事前にインストールしてください。
以下のコマンドでインストールの有無を確認できます。

```
# rpm -qa glibc
:
glibc-x.y.z.i686
```

- ・ インストールされていない場合、" glibc-x.y.z.i686" の行が出力されません

注意:x, y, z には互換性ライブラリのバージョン番号が入ります。

本パッケージは標準で OS インストール媒体中に含まれます。

- 本製品を運用中には共有メモリを約 1MB 程度使用します。

- ログ用のディレクトリ(`/opt/HA/RDM/log`)配下に、トレースファイルや core ファイルを保存するために、約 2MB 程度使用します。
トレースファイルは、サイクリックとなっていますので、2MB を超えることはありません。
- ディスクの間欠障害、部分的なメディアエラー等で異常を検出できない場合があります。
- アンインストール時に `/opt/HA` ディレクトリは削除されません。不要な場合、手動で削除してください。
- CLUSTERPRO を導入し、クラスタを起動している場合は CLUSTERPRO によるソフトウェア watchdog を利用した OS 停止機能を優先するため、RootDiskMonitor の POWER_OFF 機能を利用することができません。
どちらも同等の機能のため、CLUSTERPRO の softdog 機能を利用してください。
- EMC 社製 CLARiX シリーズ、VNX シリーズ、NEC 社製 iStorage M シリーズは監視ソフトによる TestUnitReady command の発行を許可していないため LUN への Test/I/O は実行できません。
本ディスクアレイ装置を使用した SANBoot 構成の場合は設定ファイルの TESTIO_MODE を INQ に設定する必要があります。
- マルチパスデバイスを使用した SANBoot 構成は設定ファイル自動生成 (rdmconfig) のサポート対象外です。
"3.1 本製品の導入 (4) セットアップ "を参照し、手動で設定ファイルを作成してください。
- 仮想環境 (ゲスト OS) において本製品を使用する場合は、設定ファイルの TESTIO_MODE に READ を設定する必要があります。
- StorageSaver for BootDisk の場合は、サーバ搭載の内蔵ディスクの監視には使用できません。
- Hyper-V のゲスト OS において本製品を使用する場合は、設定ファイルに udev デバイスファイル名を定義することができません。
設定ファイルを自動作成する際、以下のコマンドを実行してください。

```
# /opt/HA/RDM/bin/rdmconfig -sd
```

(2) 下記の制限事項があります。

- LVM で構築した OS ディスクについては、OS ディスクが設定ファイル自動生成の対象になります。それ以外の VG については、手動で設定ファイル情報を設定してください。

また、設定ファイル自動生成が可能な構成は、LVM および物理ディスク構成で、ディスク1つによるシングル構成、およびディスク2つによるソフトミラー構成の場合のみとなります。

ディスク4つでのソフトミラー構成や OS ディスク以外に同一 VG 内にデータディスクを有する構成の場合などでは、正しく設定ファイル自動生成できない場合があります。

このような構成の場合には手動で設定ファイルを作成してください。

- FC 接続のディスクアレイ装置を OS ディスクとして監視する場合、VG が活性化されていない、またはファイルシステムとしてマウントされていないと、正常に動作しない場合があります。必ず、ファイルシステムとしてマウントした状態で監視してください。

(3) その他

- メディアエラー監視機能、ソフトミラー構成監視機能については次期バージョン以降でサポート予定です。

7. リファレンス

rdmadmin(1M)

名称

rdmadmin – OSディスク監視モニタの制御

構文

```
/opt/HA/RDM/bin/rdmadmin [-h] [-c param] [-c status [-f file] [-t time]]  
                        [-c start] [-c stop] [-c trace [-l diag|ping]]  
                        [-c pdstatus]
```

説明

rdmadmin コマンドは、OSディスク監視モニタ (RootDiskMonitor)を制御するコマンドです。
パラメータの表示やOSディスク監視の開始／停止、OSディスク監視の設定ファイルの生成等を行います。

オプション

-h

コマンドの説明を表示します。

-c param

OSディスク監視モニタのパラメータ、監視リソースの一覧を表示します。

-c status [-f file] [-t time]

OSディスクの状態を表示します。

-f オプションにスペシャルファイルを指定すると、指定したI/Oパスの状態のみ表示します。

-f オプションを省略すると、全てのOSディスクの状態を表示します。

-t オプションに時間 (単位は秒) を指定すると、指定した時間毎に status を実行します。

-t オプションを省略すると、rdmadmin は status を一度だけ表示して終了します。

-c start

OSディスクの監視を開始します。

-c stop

OSディスクの監視を停止します。

-c trace [-l diag|ping]

OSディスク監視コマンドの内部トレースを、標準出力に表示します。

-l diag は、rdmdiagd コマンドのトレースを出力します。

-l ping は、rdmping コマンドのトレースを出力します。

-l オプションを省略すると、rdmdiagd と rdmping の両方のコマンドのトレースを出力します。

-c pdstatus

監視対象ディスクがHW-RAID構成の場合に、論理ディスクおよび物理ディスクの状態を表示します。

RootDiskMonitor HW-RAID監視機能 未インストール時は、-c status と同様の情報を表示します。

使用例

- 全てのOSディスクの状態を表示します。
/opt/HA/RDM/bin/rdmadmin
または
/opt/HA/RDM/bin/rdmadmin -c status
- OSディスク監視モニタのパラメータを表示します。
/opt/HA/RDM/bin/rdmadmin -c param
- 全てのOSディスクの状態を30秒間隔で表示します。
/opt/HA/RDM/bin/rdmadmin -c status -t 30
- HW-RAID構成の各ディスクの状態を表示します。
/opt/HA/RDM/bin/rdmadmin -c pdstatus

関連項目

rdmconfig(1M)

関連ファイル

/opt/HA/RDM/bin/rdmadmin

rdmadmin コマンドの標準のパス。

rdmstat(1M)

名称

rdmstat - OSディスク カスタムモニタリソース

構文

```
/opt/HA/RDM/bin/rdmstat [-h] [-v] [-V vgname] [-w wait-time]
```

機能説明

rdmstat は、OSディスク監視モニタ(Root Disk Monitor)がレポートするリソースステータスを監視するコマンドです。rdmstat のプロセス状態、あるいは終了ステータスを参照することで、RootDiskMonitor の監視対象リソースの状態を知ることができます。

-w オプションを指定すると、rdmstat は常駐して RootDiskMonitor がレポートするステータスを監視し続け、指定されたVGが異常状態(down)になると、異常終了します。CLUSTERPRO と連携する場合のカスタムモニタリソースとして有効なパラメータです。

RootDiskMonitor のプロセスが動作していない場合、また、rdmadmin(1M)でリソース監視の停止を指示した場合は、VGの異常を検出できないため、正常状態として報告します。

-w オプションを指定しない場合は、一回だけ RootDiskMonitor がレポートするステータスを調べ、終了します。

rdmstat は、-w オプションと RootDiskMonitor がレポートするステータスによって、以下のように動作します。

-w オプション有りの場合:

RootDiskMonitor が監視するディスクの状態	rdmstatコマンド	終了ステータス
ディスク正常時	終了せずに常駐	---
ディスク異常時	終了	1 を返す
RootDiskMonitor が動作していない	終了せずに常駐	---

-w オプション無しの場合:

RootDiskMonitor が監視するディスクの状態	rdmstatコマンド	終了ステータス
ディスク正常時	終了	0 を返す
ディスク異常時	終了	1 を返す
RootDiskMonitor が動作していない	終了	2 を返す

オプション

-h

コマンドの説明を表示します。

-v

動作トレースを標準出力に表示します。

-V vname

監視したい VG 名を指定します。

-V を指定した場合は、対象 VG 名を一つだけ指定できます。

-V を指定しない場合は、RootDiskMonitor が監視する全ての VG が監視対象になります。

複数の VG が監視対象になっている場合、少なくとも一つのVG で DOWN が検出された時点で、rdmstat は「RootDiskMonitor の監視対象ディスクが異常」と判断します。

-w wait-time

ディスク監視を常駐させたい場合に指定します。

wait-time には、監視間隔の時間 (単位は秒) を指定します。

1 以上の値を指定してください。もし 0 を指定した場合は、強制的に 1 に補整されます。

補足事項

本コマンドはメモリ上に常駐します。swap 領域に待避されません。

RootDiskMonitor のプロセスを起動していない状態、rdmadm(1M)でリソース監視の停止を指示された状態であればVGリソースの異常を検出できません。

使用例

・ヘルプを表示します。

```
##/opt/HA/RDM/bin/ rdmstat -h
```

・動作の詳細なトレースを表示します。監視対象ディスクは、RootDiskMonitor が監視する全ての VG です。

```
# /opt/HA/RDM/bin/rdmstat -v
```

・/dev/VolGroup00 のみを監視対象とします。

```
# /opt/HA/RDM/bin/rdmstat -V /dev/VolGroup00
```

・/dev/VolGroup01 で異常が発生するまで、rdmstat を常駐させます。

この時、RootDiskMonitor のステータスを 5 秒間隔で調べます。

```
# /opt/HA/RDM/bin/rdmstat -V /dev/VolGroup01 -w 5
```

rdmconfig(1M)

名称

rdmconfig - OSディスク監視モニタの設定ファイルテンプレートの自動生成

構文

/opt/HA/RDM/bin/rdmconfig [-sd]

説明

rdmconfig は、OS ディスク監視モニタ (RootDiskMonitor) の設定ファイルのテンプレートを自動生成します。
本コマンドで設定ファイルを作成した後に、監視ポリシーの設定、監視リソースの妥当性を確認してください。

rdmconfig を実行すると、/opt/HA/RDM/conf/rdm.config.default を元に/opt/HA/RDM/conf/rdm.config を生成します。コマンド実行前に既に /opt/HA/RDM/conf/rdm.config ファイルが存在する場合は上書き要否の問い合わせを行います。上書きした場合、元の設定ファイルは、/opt/HA/RDM/conf/rdm.config.old として保存されます。

オプション

-sd

設定ファイルに定義する PV 情報をブロックデバイスファイル名 (/dev/sda等) で定義します。

注意事項

- ・ 本コマンドで作成した設定ファイルは OS ディスクの構築状況によっては監視対象リソースの修正が必要です。特に、LVM で規定したミラーを構成する LUN(PV) の組み合わせについては、妥当性を確認してください。
- ・ OS ディスク以外については自動生成対象とはなりませんので、手動で設定ファイルを編集してください。
- ・ マルチパスデバイスを使用したSANBoot構成は設定ファイル自動生成 (rdmconfig) のサポート対象外です。
"3.1 本製品の導入 (4) セットアップ"を参照し、手動で設定ファイルを作成してください。
- ・ デバイスファイルのチェック機能は、設定ファイル自動生成 (rdmconfig) のサポート対象外です。本機能を使用する場合は、"3.2 設定ファイルの記述 (2) 設定ファイルの記述例について"を参照し、手動で設定ファイルを編集してください。

関連ファイル

/opt/HA/RDM/bin/rdmadmin

rdmadmin コマンドの標準のパス。

/opt/HA/RDM/conf/rdm.config

rdmconfigで生成される、RootDiskMonitor の設定ファイル。

/opt/HA/RDM/conf/rdm.config.default

rdm.config を生成する際に、元となるファイル。

8. 付録

8.1. 本製品のテスト手順について

■ はじめに

RootDiskMonitor を導入するシステムにおいて、設定ファイルの検証および性能チューニングの検証を擬似的に行う手順を説明します。

コマンドオペレーションでディスク障害を擬似することにより、物理ディスクの抜き差し等の操作を行う必要がなくなり、システムへ影響を与えず評価が実現できます。

○ RootDiskMonitor の評価

- ・ 物理ディスク故障(OS ディスク障害、クラスタウェア連携)
- ・ I/O ストール障害

■ 物理ディスク疑似障害

物理ディスクの障害には以下のパターンがあります。

- ・ 物理ディスクの障害
ソフトミラー構成の場合には片系障害を経て両系障害となります
- ・ CLUSTERPRO 連携

物理ディスクの擬似的な障害発生の手順について説明します。

本手順により、Test I/O の実行結果を擬似的に異常にすることで、監視ステータスを up から down に切り替えることや、VG ステータスを down 状態にすることができますので、評価を容易に行うことができます。

【コマンド書式】

/opt/HA/RDM/bin/rdmadmin -c debug -v [on | off] [-f SpecialFile]

off -> I/O status modify up I/O ステータスを up にします。

on -> I/O status modify down I/O ステータスを down にします。

■ シングルディスク構成の擬似障害試験手順

シングルディスク構成で物理ディスクの障害を擬似する手順を説明します。

- (1) 障害前に現在の状態をモニタリングします。

# /opt/HA/RDM/bin/rdmadmin (monitor status = TRUE)			
=====			
type	: H/W Path	: Logical status	: I/O status
=====			
VG	: /dev/VolGroup00	: up	
MIRROR	:	:	
PV	: /dev/sda	: up	: up

VG ステータスが up となっていることを確認

PV ステータスが up となっていることを確認

- (2) -f SpecialFile オプションで障害を擬似するディスクを指定します。
指定されたディスクは強制的に I/O ステータスが up から down に変更され擬似的に障害を起こすことができます。

```
# /opt/HA/RDM/bin/rdmadmin -c debug -v on -f /dev/sda
(monitor status = TRUE)
Change debug value.
special file = /dev/sda
0 -> 1
```

約 60 秒後

# /opt/HA/RDM/bin/rdmadmin (monitor status = TRUE)			
=====			
type	: H/W Path	: Logical status	: I/O status
=====			
VG	: /dev/VolGroup00	: down	
MIRROR	:	:	
PV	: /dev/sda	: down	: down

シングルディスク構成では down となります

メモリ上のステータスを強制的に書き換えます

- (3) syslog にメッセージが出力されます。

以下のメッセージが出力されることを確認してください。

Apr 19 10:24:24 host1 rdm[16517]: PV down find .(sf=/dev/sda)

Apr 19 10:24:24 host1 rdm[16517]: VG status change down .(vg=/dev/VolGroup00) → ①

Apr 19 10:24:24 host1 rdm[16517]: start KILL clpnm.

Apr 19 10:24:24 host1 rdm[16517]: send signal clpnm. → ②

Apr 19 10:24:24 host1 rdm[16517]: abort rdmdiagd. → ③

- ① OS ディスク(全パス)のステータスが down になったため、
VG の状態も down になります。

- ② CLUSTERPRO のサーバ管理プロセス (clpnm) の強制終了。

- ③ 監視エンジン(rdmdiagd) の強制終了。

- (4) CLUSTERPRO のサーバ管理プロセス (clpnm) が強制終了されていることを確認します。

```
# ps -ef |grep clpnm
```

何も表示されません。(強制終了した事を示します。)

- (5) 監視エンジン(rdmdiagd)、監視モニタ(rdmping)が強制終了されていることを確認します。

```
# ps -ef |grep rdm
```

何も表示されません。(強制終了した事を示します。)

CLUSTERPRO のサーバ管理プロセス (clpnm) の強制終了後、クラスタを再度起動する際は
以下どちらかの方法で復旧を行ってください。

方法1: マシンの再起動を行う。

方法2: CLUSTERPRO のサーバ管理プロセス (clpnm) のコマンド起動

```
# clpcl -s
```

■ ソフトミラー構成の擬似障害試験手順

ソフトミラー構成のディスクで片系の物理ディスクの障害を擬似する手順を説明します。

- (1) 障害前に現在の状態をモニタリングします。

```
# /opt/HA/RDM/bin/rdmadmin
(monitor status = TRUE)
```

		: Logical	: I/O
type	: H/W Path	: status	: status
VG	: /dev/VolGroup00	: up	
MIRROR	:	:	
PV	: /dev/sda	: up	: up
PV	: /dev/sdb	: up	: up

VG ステータスが
up であることを確認

PV ステータスが
up であることを確認

- (2) `-f SpecialFile` にて、ミラーを構成する片系のディスクを指定します。
指定されたディスクは強制的に I/O ステータスが up から down に変更され擬似的に片系障害を起こすことができます。

```
# /opt/HA/RDM/bin/rdmadmin -c debug -v on -f /dev/sdb
(monitor status = TRUE)
Change debug value.
special file = /dev/sdb
0 -> 1
```

約 60 秒後

```
# /opt/HA/RDM/bin/rdmadmin
(monitor status = TRUE)
```

		: Logical	: I/O
type	: H/W Path	: status	: status
VG	: /dev/VolGroup00	: suspend	
MIRROR	:	:	
PV	: /dev/sda	: up	: up
PV	: /dev/sdb	: down	: down

ソフトミラー構成では
suspend になります

メモリ上のステータスを
強制的に書き換えます

- (3) 片系障害を発生させた場合、syslog にメッセージが出力されます。
以下のメッセージが出力されることを確認してください。

Apr 19 10:19:44 host1 rdm[16517]: PV down find.(sf=/dev/sdb)

Apr 19 10:19:44 host1 rdm[16517]: VG status change suspend.(vg=/dev/VolGroup00)

次に、ソフトミラー構成のディスクで両系の物理ディスクの障害を擬似する手順を説明します。

本手順により、CLUSTERPRO のサーバ管理プロセス (clpnm) を強制終了し、CLUSTERPRO との連携を実現することが可能になります。

CLUSTERPRO との連携方式は2通りあり、ここでは『 CLUSTERPRO のサーバ管理プロセス (clpnm) の強制終了による CLUSTERPRO との連携 』の動作確認方法を記載します。

- (1) 設定ファイル内の OVER_ACTION に CLPNM_KILL を指定します。

```
# vi /opt/HA/RDM/conf/rdm.config
.....
# Root Volume fault action
# select SERVICE_CMD_DISABLE(default),SERVICE_CMD_ENABLE,TOC_EXEC,
# CLPNM_KILL,POWER_OFF
#OVER_ACTION          SERVICE_CMD_DISABLE
OVER_ACTION          CLPNM_KILL
.....
.....
```

設定ファイルを変更した後、デーモンプロセスを再起動してください

【Red Hat Enterprise Linux 7.x】

【Oracle Linux 7.x】

```
# systemctl restart rdmd
```

【Red Hat Enterprise Linux 6.x と 5.x】

【Oracle Linux 6.x】

```
# /etc/init.d/rdmd restart
```

- (2) CLUSTERPRO のサーバ管理プロセス (clpnm) が起動していることを確認します。

```
# ps -ef |grep clpnm
root 15691 15690  0 15:36:45 ?        0:00 clpnm
```

- (3) RootDiskMonitor のプロセス rdmdiagd と rdmping が起動していることを確認します。

```
# ps -ef |grep rdm
root  9351      1  0 14:53:12 ?        0:03 /opt/HA/RDM/bin/rdmdiagd
root  9354  9351  0 14:53:12 ?        0:00 rdmping
```

- (4) 障害前に現在の状態をモニタリングします。(既に片系のディスクが異常であること。)

```
# /opt/HA/RDM/bin/rmdadmin
(monitor status = TRUE)
```

		Logical	I/O
type	H/W Path	status	status
VG	/dev/VolGroup00	suspend	
MIRROR			
PV	/dev/sda	up	up
PV	/dev/sdb	down	down

片系障害で suspend 状態になっていることを確認

up になっていることを確認

down になっていることを確認

- (5) ソフトミラー構成のディスクの両系の物理ディスクを異常状態にします。

すでに片系が異常な状態で、残りの正常なディスクを異常にすることで両系障害を擬似します。

```
# /opt/HA/RDM/bin/rmdadmin -c debug -v on -f /dev/sda
(monitor status = TRUE)
Change debug value.
special file = /dev/sda
0 -> 1
```

約 60 秒後

```
# /opt/HA/RDM/bin/rmdadmin
(monitor status = TRUE)
```

		Logical	I/O
type	H/W Path	status	status
VG	/dev/VolGroup00	down	
MIRROR			
PV	/dev/sda	down	down
PV	/dev/sdb	down	down

両系障害で down 状態になっていることを確認

メモリ上のステータスを強制的に書き換えます

- (6) I/O パスの両系障害を発生させた場合、syslog にメッセージが出力されます。

以下のメッセージが出力されることを確認してください。

Apr 19 10:24:24 host1 rdm[16517]: PV down find .(sf=/dev/sda)

Apr 19 10:24:24 host1 rdm[16517]: VG status change down .(vg=/dev/VolGroup00) → ①

Apr 19 10:24:24 host1 rdm[16517]: start KILL clpnm.

Apr 19 10:24:24 host1 rdm[16517]: send signal clpnm. → ②

Apr 19 10:24:24 host1 rdm[16517]: abort rdmdiagd. → ③

①ソフトミラーを構成する両系のディスクのステータスが down になったため、VG の状態も down になります。

②CLUSTERPRO のサーバ管理プロセス (clpnm) の強制終了。

③監視エンジン(rdmdiagd) の強制終了。

- (7) CLUSTERPRO のサーバ管理プロセス (clpnm) が強制終了されていることを確認します。

```
# ps -ef |grep clpnm
```

何も表示されません。(強制終了した事を示します。)

- (8) 監視エンジン(rdmdiagd)、監視モニタ(rdmping)が強制終了されていることを確認します。

```
# ps -ef |grep rdm
```

何も表示されません。(強制終了した事を示します。)

CLUSTERPRO のサーバ管理プロセス (clpnm) の強制終了後、クラスタを再度起動する際は以下どちらかの方法で復旧を行ってください。

方法1: マシンの再起動を行う。

方法2: CLUSTERPRO のサーバ管理プロセス (clpnm) のコマンド起動

```
# clpcl -s
```

■ CLUSTERPRO 連携

CLUSTERPRO 連携による デバッグ 手順を説明します。

本手順により、rdmstat(1M) を CLUSTERPRO のカスタムモニタリソースとして登録し、OS ディスクの障害時のノードダウン、ノード切り替えを実現することが可能になります。
ここでは『カスタムモニタリソースによる CLUSTERPRO との連携』の動作確認方法を記載します。
クラスタ環境の構築がされていることが前提となります。

- (1) クラスタ環境構築
クラスタ環境構築については、
『 8 章(8.2) カスタムモニタリソースによる CLUSTERPRO との連携手順 』
をご覧ください。
- (2) クラスタの起動
クラスタ環境構築後、クラスタの起動を行います。

クラスタの起動によりサービスとして登録した rdmstat が状態監視を開始します。

※ rdmstat は/opt/HA/RDM/bin 配下から任意のディレクトリ配下へコピーしてご利用ください。

以下は rdmstat を/usr/local/bin 配下にコピーした場合の例です。

```
root 26949 26929  0 10:40:04 ?          0:00 /usr/local/bin/rdmstat -w 10
```

(3) 物理ディスク障害による VG ステータスのダウン

- ① 前述の手順により、物理ディスクの障害を発生させ VG ステータスを down 状態にします。
- ② rdmstat が VG の異常を検出します。
- ③ rdmstat 自身が異常終了します。
- ④ 障害が発生したノードをダウンさせ、ノード切り替えを行います。

syslog を確認してください。

```
Apr 22 10:42:53 host1 rdm[19483]: PV down find .(sf=/dev/sdb)
```

```
Apr 22 10:42:53 host1 rdm[19483]: VG status change down .(vg=/dev/VolGroup00) → ①
```

```
Apr 22 10:42:53 host1 rdm[19483]: Action is not define.
```

動作確認後は、マシンの再起動を行ってください。

■ 物理ディスク疑似障害の復旧

- (1) 物理ディスク(/dev/sda)の疑似障害の復旧を行います。

```
# /opt/HA/RDM/bin/rdmadmin -c debug -v off -f /dev/sda
Change debug value.
special file = /dev/sda
1 -> 0
```

```
# /opt/HA/RDM/bin/rdmadmin
(monitor status = TRUE)
```

	:		:	Logical	:	I/O
type	:	H/W Path	:	status	:	status
VG	:	/dev/VolGroup00	:	up	:	
MIRROR	:		:		:	
PV	:	/dev/sda	:	up	:	up

up になることを
確認します

- (2) syslog メッセージを確認してください。

Apr 19 11:57:40 host1 rdm[16517]: VG status change up.(vg=/dev/VolGroup00)

■ I/O ストール擬似障害

本手順により、I/O ストール状態を擬似的に発生させることができます。
設定ファイル内の VG_STALL_ACTION に SERVICE_CMD_DISABLE を指定し、
syslog に I/O ストールのエラーメッセージが出力されることを確認してください。

【コマンド書式】

/opt/HA/RDM/bin/rdmadmin -c debug2 -v [on | off]

off -> I/O stall stop I/O ストール擬似障害を復旧します。

on -> I/O stall start I/O ストール擬似障害を開始します。

- (1) 設定ファイル内の VG_STALL_ACTION に SERVICE_CMD_DISABLE を指定します。

```
# vi /opt/HA/RDM/conf/rdm.config
.....
# RootVolume stall find action
# select SERVICE_CMD_DISABLE(default),SERVICE_CMD_ENABLE,TOC_EXEC,
# CLPNM_KILL,POWER_OFF
VG_STALL_ACTION          SERVICE_CMD_DISABLE
.....
.....
```

設定ファイルを変更した後、デーモンプロセスを再起動してください

【Red Hat Enterprise Linux 7.x】

【Oracle Linux 7.x】

```
# systemctl restart rdmd
```

【Red Hat Enterprise Linux 6.x と 5.x】

【Oracle Linux 6.x】

```
# /etc/init.d/rdmd restart
```

- (2) RootDiskMonitor のプロセス rdmdiagd、rdmping が起動していることを確認します。

```
# ps -ef |grep rdm
root 9351      1  0 15:37:35 ?        0:03 /opt/HA/RDM/bin/rdmdiagd
root 9354  9351  0 15:37:35 ?        0:00 rdmping
```

- (3) rdmadmin コマンドにより擬似的に I/O ストール擬似障害を開始します。

実行後、以下のメッセージが出力されます。

```
# /opt/HA/RDM/bin/rdmadmin -c debug2 -v on
DEBUG: I/O STALL start.
```

- (4) I/O ストールの警告メッセージを確認します。
- ※ (3) の rdmadmin コマンド実行後、設定ファイルに記載の
TIME_VG_STALL(デフォルト 360 秒)の半分(180 秒)経過した場合に
トレースに出力されます。

```
# /opt/HA/RDM/bin/rdmadmin -c trace -l diag
Thu Jul 21 17:22:05 2005: TRACE : I/O stall check start
Thu Jul 21 17:22:05 2005: TRACE : I/O stall find .(sf=/dev/sda) ★
Thu Jul 21 17:22:05 2005: TRACE : I/O stall check end
```

※ I/O stall find .(sf=/dev/sda) このメッセージは syslog には出力されません。

- (5) I/O ストールのエラーメッセージを確認します。
- ※ (3) の rdmadmin コマンド実行後、設定ファイルに記載の
TIME_VG_STALL(デフォルト 360 秒)経過した場合に syslog に出力されます。

```
# vi /var/log/messages
Jul 21 17:25:06 host1 rdm[11930]: I/O stall find , timeover occurred.(sf=/dev/sda) ★
Jul 21 17:25:06 host1 rdm[11930]: Action is not define.(VG STALL)
```

評価完了後、I/O ストール擬似障害を復旧します。

- (6) I/O ストール擬似障害の復旧を行います。

実行後、以下のメッセージが出力されます。

```
# /opt/HA/RDM/bin/rdmadmin -c debug2 -v off
DEBUG: I/O STALL stop.
```

また syslog に以下のメッセージが出力されます。

```
# vi /var/log/messages
Jul 21 17:26:06 host1 rdm[11930]: DEBUG:I/O stall DEBUG mode end. I/O restore.
```

8.2. カスタムモニタリソースによる CLUSTERPRO との連携手順

8.2.1.CLUSTERPRO 連携設定

本製品は、カスタムモニタリソースによる CLUSTERPRO との連携を行うことが可能です。

以下の作業は、CLUSTERPRO WebManager にて実施します。

本書は、CLUSTERPRO Server をインストールしたサーバの実 IP アドレスを「192.168.11.100」、ポート番号を「29003(デフォルト値)」とした場合の例です。

接続例) `http://192.168.11.100:29003/`

また、本書では CLUSTERPRO WebManager のモニタリソースの設定を以下としています。

プロパティ	設定値
タイプ (モニタリソースのタイプ)	custom monitor
名前 (カスタムモニタリソース名)	rdmstat_mon
監視タイプ	非同期
回復対象	LocalServer
回復動作	最終動作のみ実行
最終動作	クラスタサービス停止と OS シャットダウン

上記設定を行うことによって、VG 障害発生時にカスタムモニタリソース(rdmstat_mon) が rdmstat の消滅を検知し、現用系ノードを shutdown させた後、待機系へノード切り替えを行います。

※ 本書で設定している各種プロパティの値は一例です。構築時にはそれぞれの環境に応じた値を設定してください。

※ 以下は、CLUSTERPRO X 3.1 の場合の設定手順となります。

1. 設定ファイルの確認

RootDiskMonitor の設定を確認します。

- (1) RootDiskMonitor の設定ファイル (rdm.config) にて OVER_ACTION が SERVICE_CMD_ENABLE になっていることを確認します。

```
# /bin/cat /opt/HA/RDM/conf/rdm.config | /bin/grep "OVER_ACTION"
OVER_ACTION          SERVICE_CMD_ENABLE
```

※ 旧バージョンとの互換を保証するため、OVER_ACTION が ACTION_NONE の場合にも CLUSTERPRO 連携は可能です。

※ OVER_ACTION が SERVICE_CMD_ENABLE または ACTION_NONE になっていない場合は、以下 (2) ~ (4) の手順で設定変更およびデーモンプロセス(rdmd)の再起動を行ってください。

- (2) 設定ファイル (rdm.config) を編集します。

※ 以下は vi コマンドを使用して編集を行う場合の例です。

```
# /bin/vi /opt/HA/RDM/conf/rdm.config
```

- (3) 設定ファイル (rdm.config) が正しく変更されていることを確認します。

```
# /bin/cat /opt/HA/RDM/conf/rdm.config | /bin/grep "OVER_ACTION"
OVER_ACTION          SERVICE_CMD_ENABLE
```

- (4) デーモンプロセス (rdmd) を再起動します。

【Red Hat Enterprise Linux 7.x】

【Oracle Linux 7.x】

```
# systemctl restart rdmd
```

【Red Hat Enterprise Linux 6.x と 5.x】

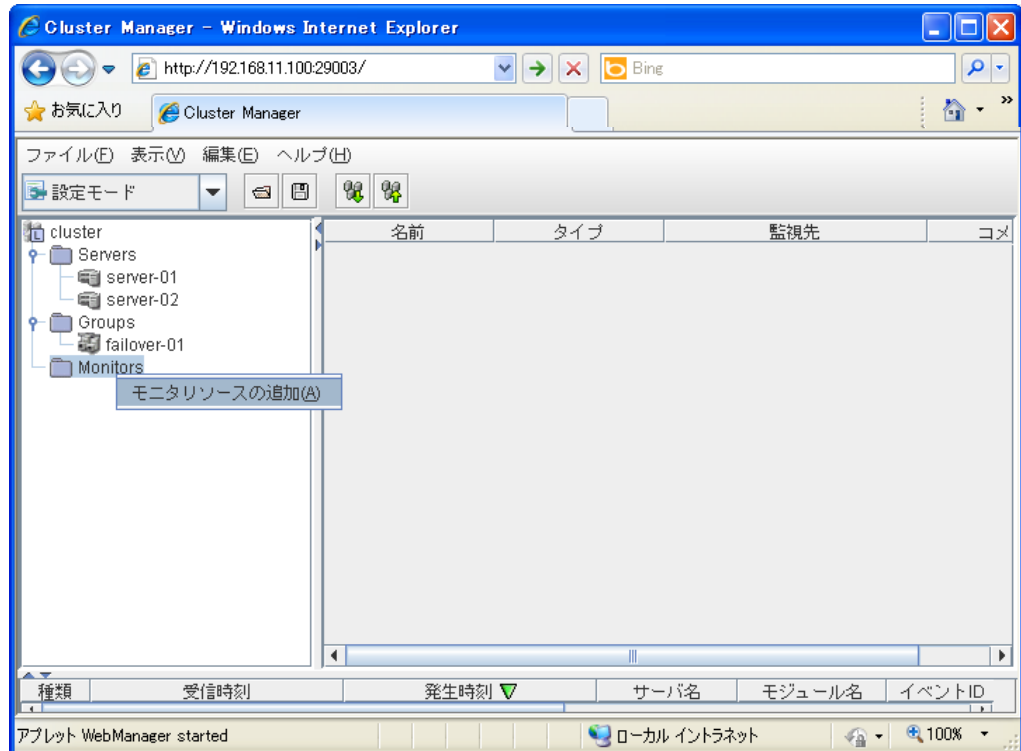
【Oracle Linux 6.x】

```
# /etc/init.d/rdmd restart
```

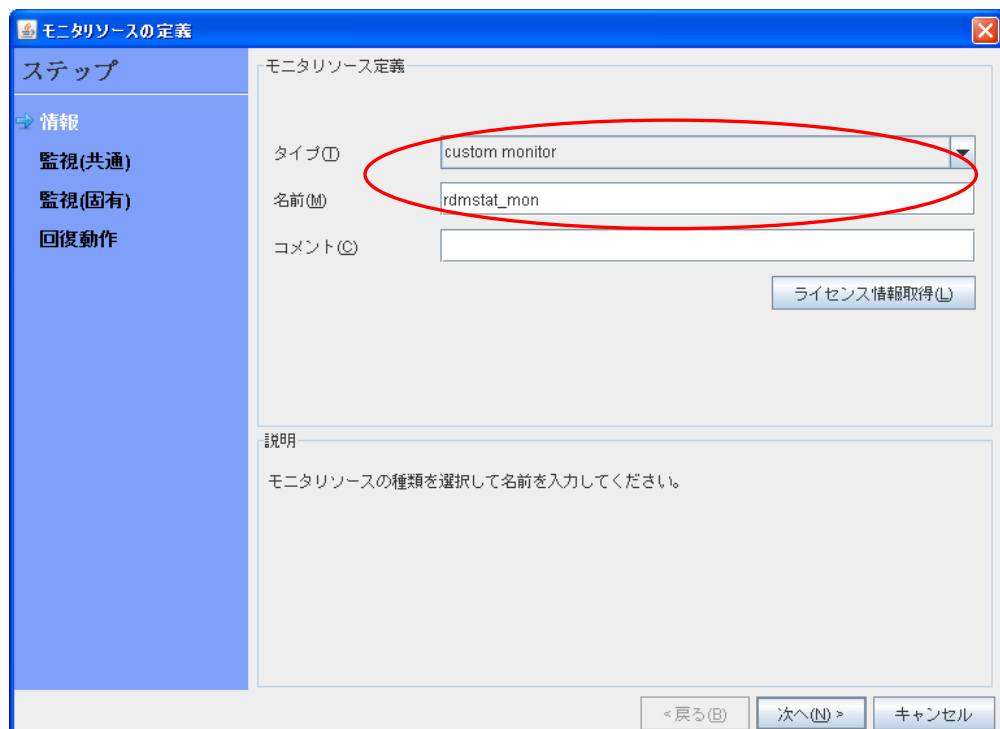
2. カスタムモニタリソースの作成

rdmstat を監視するカスタムモニタリソースを作成します。
CLUSTERPRO WebManager の「表示」メニューより「設定モード」を選択し、設定します。

(1) 「Monitors」を右クリックし、「モニタリソースの追加」を選択します。



(2) 「モニタリソースの定義」ダイアログボックスが開きます。
「タイプ」で「custom monitor」を選択し、「名前」にカスタムモニタリソース名「rdmstat_mon」を入力し、「次へ」をクリックします。



(3) 監視条件を設定します。

本書ではデフォルトのまま変更しません。「次へ」をクリックします。

モニタリソースの定義

ステップ

- 情報
- 監視(共通)
- 監視(固有)
- 回復動作

インターバル(I) 60 秒

タイムアウト(O) 120 秒

リトライ回数(R) 0 回

監視開始待ち時間(S) 0 秒

監視タイミング

- ☒ 常時(L)
- ☐ 活性時(C)

対象リソース 参照(W)

nice値(E) 0

監視を行うサーバを選択する サーバ(V)

< 戻る(B) 次へ(N) > キャンセル

(4) 監視条件を設定します。

「ユーザアプリケーション」を選択し、「ファイル」に rdmstat 起動処理

(/opt/HA/RDM/bin/rdmstat -w 30) を入力します。

監視タイプで「非同期」を選択し、「次へ」をクリックします。

モニタリソースの定義

ステップ

- 情報
- 監視(共通)
- 監視(固有)
- 回復動作

☒ ユーザアプリケーション(U)

☐ この製品で作成したスクリプト(W)

ファイル(F) /opt/HA/RDM/bin/rdmstat -w 30

表示(V) 編集(E) 置換(P)

監視タイプ

- ☐ 同期(S)
- ☒ 非同期(A)

ログ出力先(L)

☐ ロータートする(R)

ローテートサイズ(Z) 1000000 バイト

正常な戻り値(M) 0

表示及び編集するツールは変更できます。 変更(C)

< 戻る(B) 次へ(N) > キャンセル

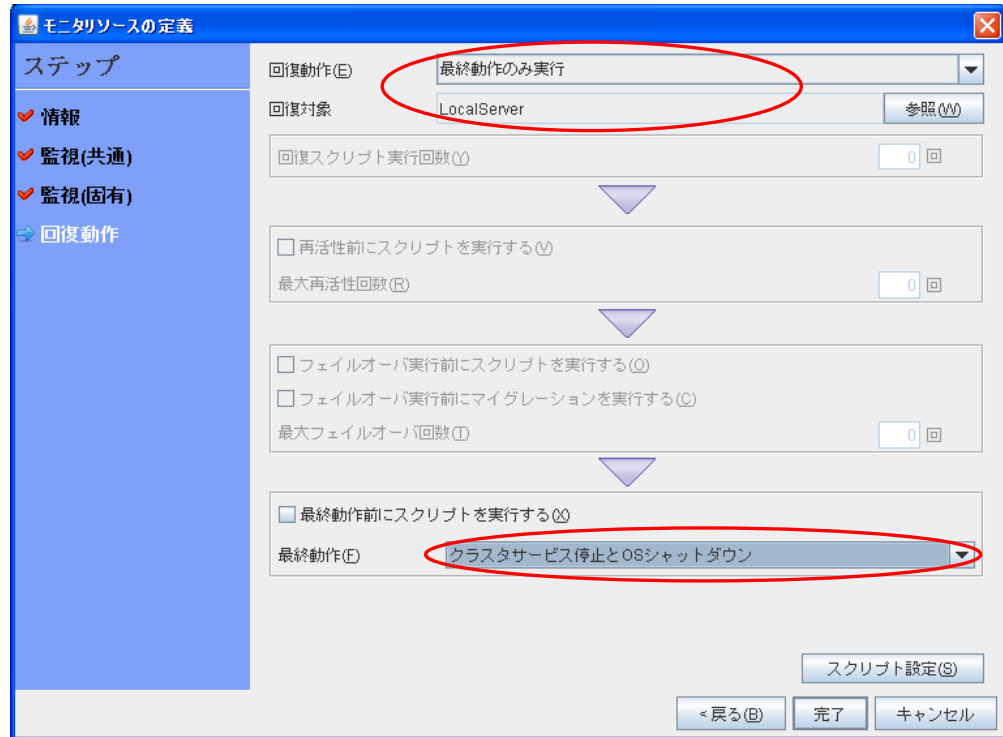
(5) 回復動作を設定します。

「回復動作」で「最終動作のみ実行」を選択します。

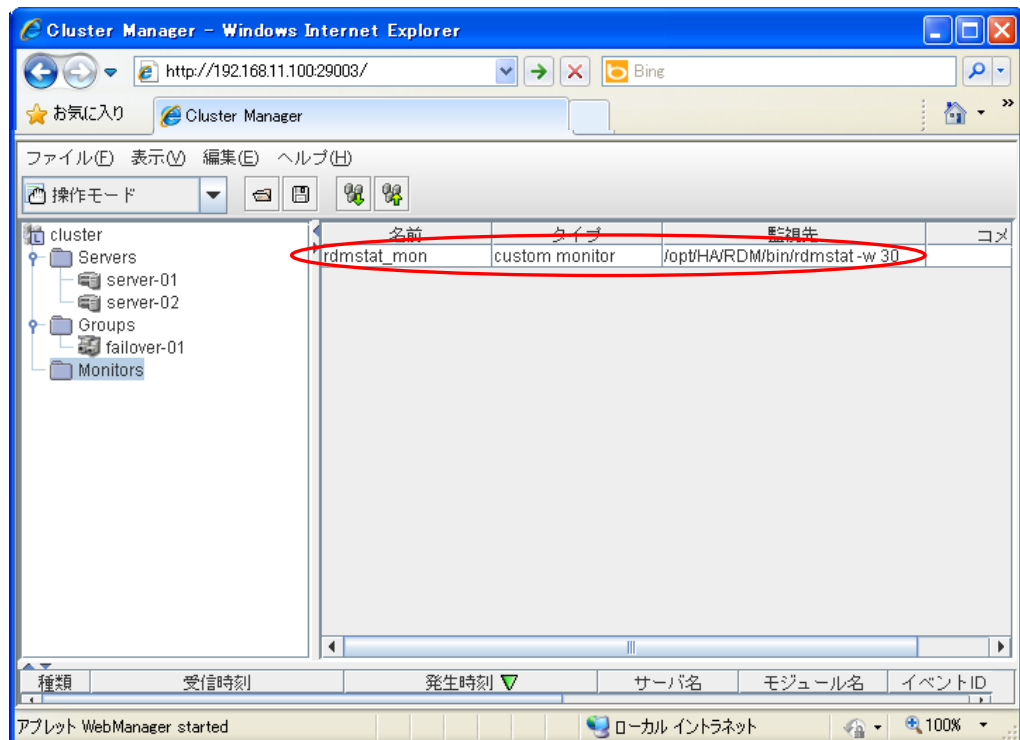
「回復対象」の「参照」をクリックし、表示されるツリービューで“LocalServer”を選択して

「OK」をクリックします。「回復対象」に“LocalServer”が追加されたことを確認します。

「最終動作」で「クラスタサービス停止と OS シャットダウン」を選択し、「完了」をクリックします。



(6) 「Monitors」をクリックし、カスタムモニタリソースが登録されていることを確認します。

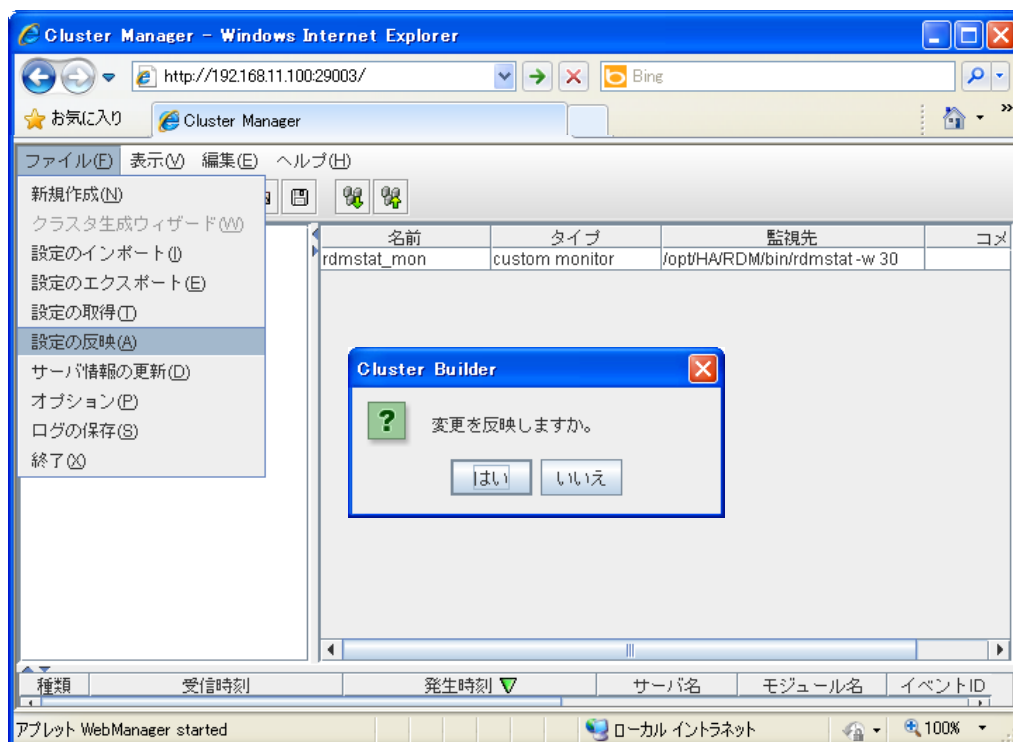


3. クラスタ構成情報のアップロード

(1) クラスタ構成情報の内容を、CLUSTERPRO 本体の環境に反映します。

「ファイル」メニューから「設定の反映」を選択します。

確認ダイアログが表示されます。「OK」をクリックします。

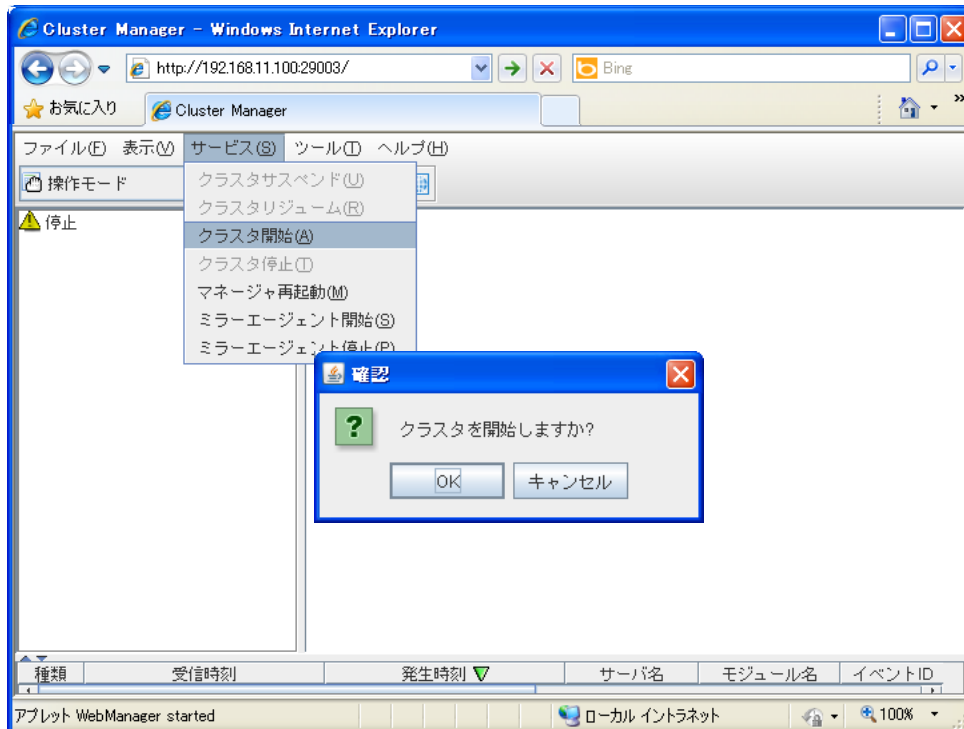


※ 変更した情報によって表示されるメッセージが異なりますので、表示されたメッセージにしたがって操作を行ってください。

詳細は、CLUSTERPRO X 3.x for Linux のマニュアルを参照してください。

反映に成功すると確認ダイアログが表示されます。「了解」をクリックしてダイアログを閉じます。

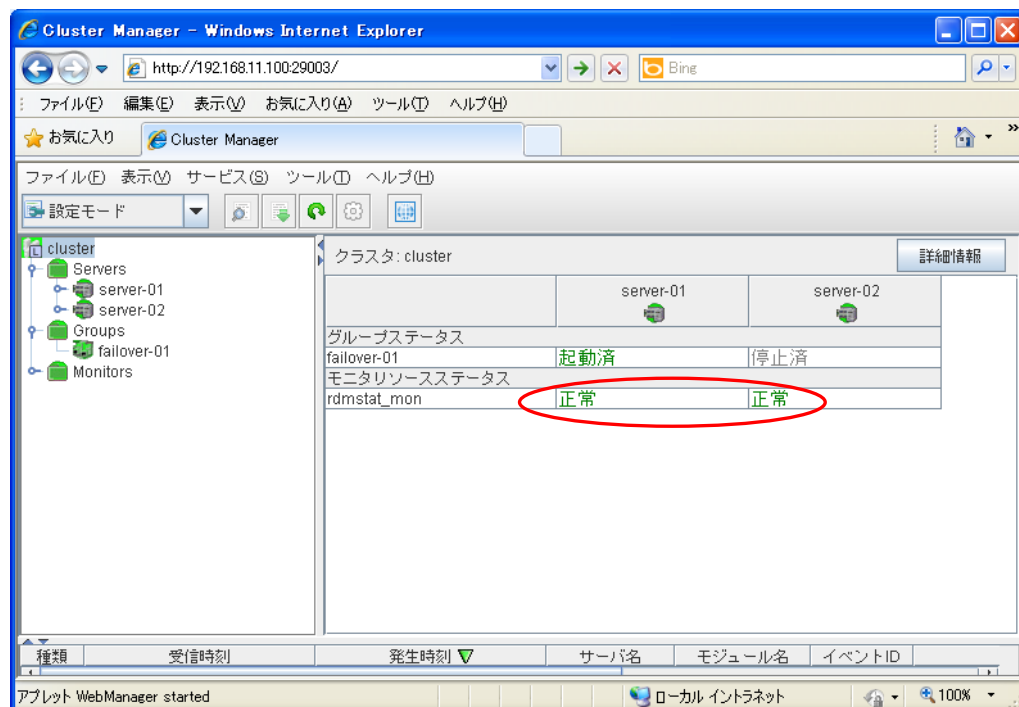
クラスタが停止状態の場合は、クラスタを開始します。
WebManager の「表示」メニューより「操作モード」を選択し、「サービス」メニューから「クラスタ開始」を選択し、クリックしてください。



(2) 設定が反映されていることを確認します。

WebManager の「表示」メニューより「操作モード」を選択し、以下の項目を確認してください。

- ・ 現用系サーバ、待機系サーバにて rdmstat 監視用のカスタムモニタリソース「rdmstat_mon」のステータスが「正常」であることを確認してください。



以上で、CLUSTERPRO の設定は終了です。

8.2.2.動作確認

以降の手順で RootDiskMonitor および CLUSTERPRO の設定の動作確認を行います。
コマンドオペレーションで擬似的にディスク障害を発生させ、ディスクにアクセスができなくなった際にフェイルオーバーが発生することを確認します。

1. 現用系サーバでの RootDiskMonitor 動作確認

コマンドオペレーションでディスク障害を擬似的に発生させます。
擬似障害テスト手順については、「8.1 本製品のテスト手順について」の「物理ディスク擬似障害」の章を参照してください。

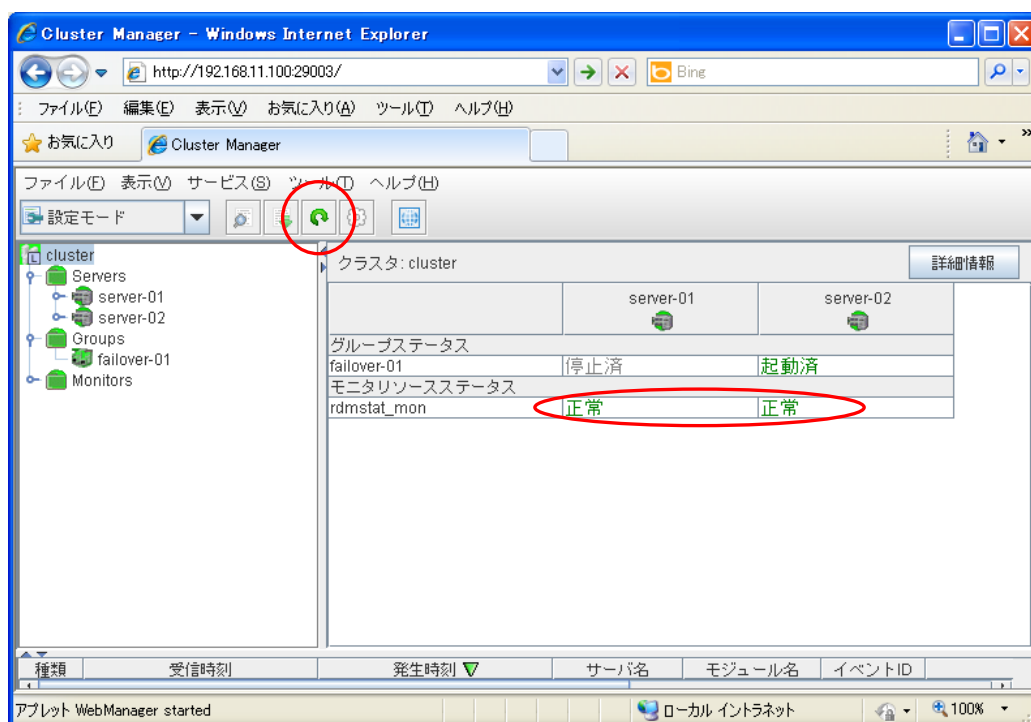
2. 待機系サーバへのフェイルオーバー確認

- (1) 手順 1. で shutdown したサーバを起動させてください。
- (2) syslog に 以下のようなメッセージが出力されていることを確認してください。

```
clusterpro: [I] <type: rm><event: 16> Stopping the system has been required because an error was detected in monitoring rdmstat_mon.  
clusterpro: [I] <type: pm><event: 30> Received a request to stop the system from internal(rc).
```

- (3) CLUSTERPRO WebManager の「リロード」をクリックし、以下の項目を確認してください。

- ・ rdmstat 監視用のカスタムモニタリソース「rdmstat_mon」のステータスが現用系、待機系にて「正常」である事を確認してください。



以上で、動作確認は終了となります。

8.3. HW-RAID 状態表示

RootDiskMonitor HW-RAID 監視機能 をインストールすることで、運用管理コマンドを用いて HW-RAID 構成の論理ディスクおよび物理ディスクの状態表示を行うことが可能となります。

ただし、監視対象ディスクを構成している論理ディスクおよび物理ディスクのみの表示となります。

RootDiskMonitor HW-RAID 監視機能 については別冊『CLUSTERPRO MC RootDiskMonitor 2.0 for Linux HW-RAID 監視機能 ユーザーズガイド』を参照してください。

(1) 運用管理コマンドによる HW-RAID 状態表示

以下のコマンドにて HW-RAID 状態表示を行います。

```
# /opt/HA/RDM/bin/rdmadmin -c pdstatus
(monitor status = TRUE)
```

=====			
	:	: Logical	: I/O
type	: H/W Path	: status	: status
=====			
VG	: VG_NONE	: up	
MIRROR	:	:	
PV	: /dev/sda	: up	: up
LD	: 1	: Online	
PD	: 1	: Online	
PD	: 2	: Online	
MIRROR	:	:	
PV	: /dev/sdb	: up	: up
LD	: 2	: Degraded	
PD	: 3	: Failed	
PD	: 4	: Online	

論理ディスクの状態を表示します

物理ディスクの状態を表示します

(2) 注意・制限事項

- HW-RAID の冗長性が低下しても OS からは正常に見えるため PV のステータスは up になります。
- 論理ディスクが Offline の場合は論理ディスクおよび物理ディスクの状態は表示されなくなります。
- 本コマンドで障害を検出した場合は HW 部門に確認を取ってください。
- RAID コントローラが複数ある環境はサポート対象外です。
- LD に付加されるデバイス名(/dev/sdX)は補助的な情報のため故障箇所の特特定は HW 構成を確認してください。

CLUSTERPRO
MC RootDiskMonitor 2.0 for Linux

CLUSTERPRO
MC StorageSaver for BootDisk 2.0 (for Linux)

ユーザーズガイド

2015 年 3 月 第 1 版
日本電気株式会社
東京都港区芝五丁目 7 番地 1 号
TEL (03) 3454-1111 (代表)



© NEC Corporation 2015

日本電気株式会社の許可なく複製、改変などを行うことはできません。
本書の内容に関しては将来予告なしに変更することがあります。

保護用紙