

CLUSTERPRO

MC StorageSaver 1.2 for Linux

はじめての StorageSaver

© 2014(Mar) NEC Corporation

- ☐ はじめに
- ☐ 製品概要
- ☐ 監視の設定と実践
- ☐ おわりに

はしがき

本書は、CLUSTERPRO MC StorageSaver 1.2 for Linux(以後 StorageSaver と記載します)による ディスク装置監視の基本的な動作について説明したものです。

(1) 商標および商標登録

- ✓ 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。
- ✓ SUSE は、米国およびその他の国における Novell, Inc.の登録商標または商標です。
- ✓ 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 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。

目次

1. はじめに	1
2. 製品概要	2
2.1. 製品概要について	2
3. 監視の設定と実践	10
3.1. インストール	10
3.2. 設定ファイルの作成	11
3.3. 監視の開始	12
3.4. 監視状態の確認	13
3.5. ディスク障害検知の確認	15
3.6. 監視の復旧	18
3.7. 監視の終了	20
4. おわりに	21

1. はじめに

本書では、StorageSaver による ディスク装置監視の基本的な動作について説明します。
ディスク装置監視の設定および監視手順を一連の流れで確認することで、StorageSaver の動作を理解することができます。

本書では、以下のような監視例を使用します。

- ・ StorageSaver によるディスク装置監視
- ・ 疑似障害コマンドによる障害検知

なお、各パラメータの意味や詳細につきましては、同梱の「CLUSTERPRO MC StorageSaver 1.2 for Linux ユーザーズガイド」をご覧ください。

本書では以下の手順で ディスク装置の監視を行います。

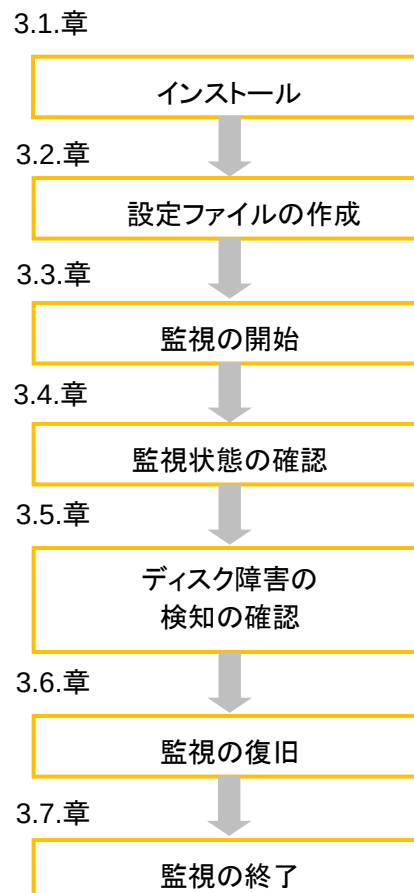


図 1 ディスク装置監視手順

2. 製品概要

2.1. 製品概要について

本製品には以下の機能があります。

◆ 共有ディスクのリソース監視機能

FC 接続や SCSI で接続されたディスク装置に対して、
TestI/O によるリソース監視を実現します。

→ ディスクアレイ装置コントローラの障害監視

FC カード、SCSI カードの障害監視

I/O パスの死活監視

I/O リクエストのストール監視

◆ I/O パスの自動閉塞機能

StoragePathSavior(以降、SPS)および PowerPath で冗長化された I/O パスを
有するディスク装置において、リソース監視で異常を検出すると障害の発生した
I/O パスを閉塞し、すみやかに正常なパスへ切り替えます。

→ 間欠的な FC リンクダウン障害に伴う頻繁なパス切り替えによる

I/O パフォーマンスの低下や他のディスク装置への影響を未然に防止します。

注意 : 閉塞機能・復旧機能は SPS または PowerPath を導入している環境でのみ有効です。

◆ クラスタウェア連携機能

ディスク装置へのすべてのインタフェース機構の異常により、ユーザーデータへの
アクセスができなくなると、クラスタウェアと連携し、パッケージの移動や
ノード切り替えにより業務の継続的な運用を実現します。

注意 : 本機能は、クラスタウェアが導入されたクラスタシステムで使用可能です。

◆ オンライン保守機能

FC 接続のディスク装置に対して、SPS および PowerPath における I/O パスの一括閉塞、一括復旧を実現します。

これにより、FC 上でリンクダウン時の障害が発生した際に、障害装置の特定化と保守員によるシステム無停止保守を可能とします。

◆ 運用管理機能

I/O パスの監視状態の表示や手動による閉塞、復旧といった運用管理機能をコマンドインタフェースで提供します。

注意:閉塞機能・復旧機能は SPS または PowerPath を導入している環境でのみ有効です。

◆ オートコンフィグレーション機能

設定ファイルの自動生成機能です。

アクションに関しては、ユーザーカスタマイズ機能によりリソース単位での定義が可能です。

◆ プロセス監視機能

本製品で提供するデーモンプロセスやリソース監視コマンドの動作状態を監視し、異常を検出すると自動的に再起動します。これにより、継続的なリソース監視を実現します。

StorageSaver によるディスク装置監視の仕組みは以下の図のようになります。

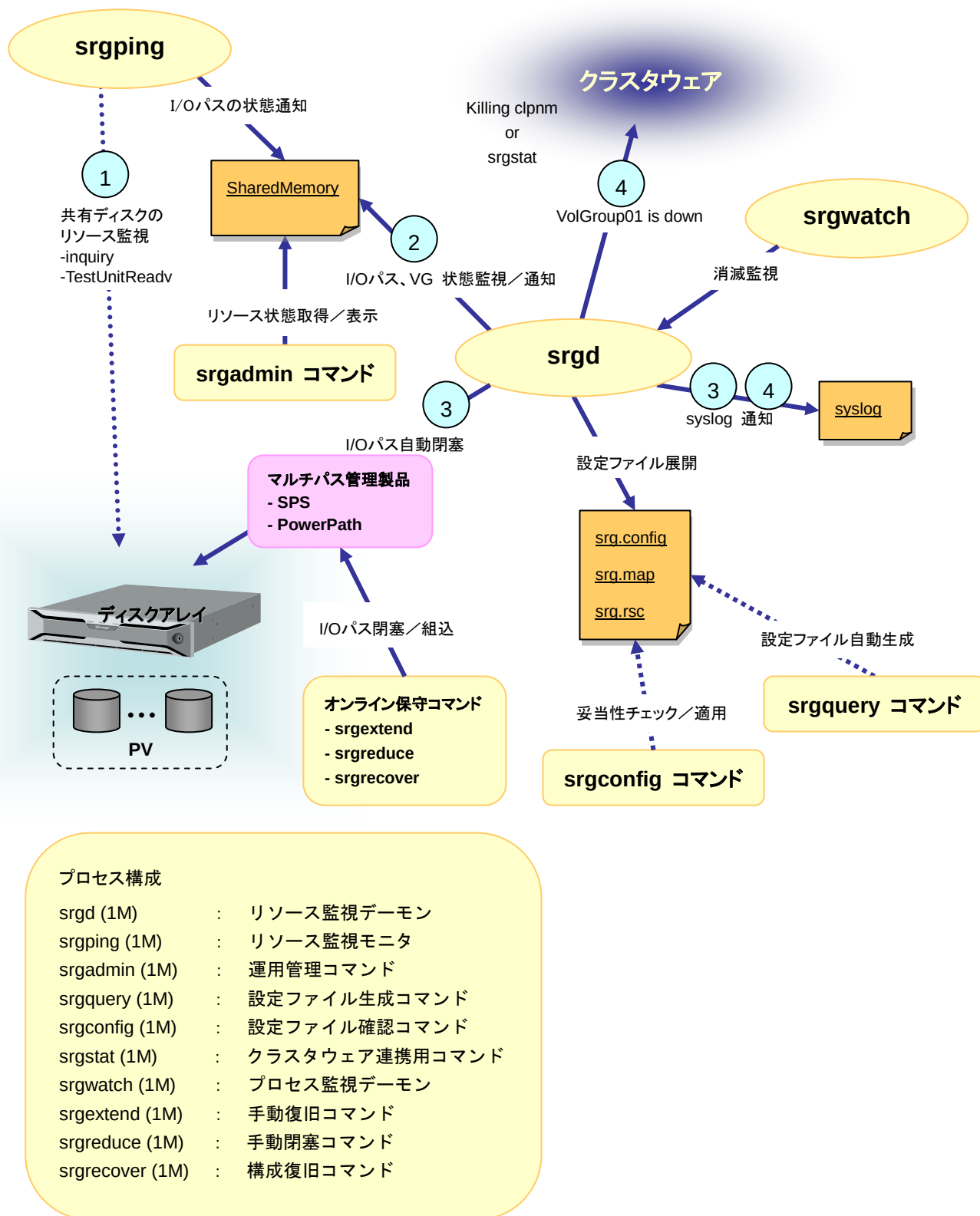


図 2 StorageSaver のプロセスモデル

1

ディスクアレイ装置を構成する I/O パスに対して定期的に TestI/O を発行します。

※Test I/O は SCSI パススルードライバ経由で SCSI コマンド(Inquiry、TestUnitReady)を発行することで行われます。

Test I/O で監視対象となる項目は、I/O パスの死活監視と、I/O リクエストのストール監視になります。

2

I/O パス、VG の状態を監視/更新します。

3

Test I/O で片系コントローラ異常を検出すると、syslogに障害レポートを通知します。

さらに、設定ファイルの TESTIO_FAULT_ACTION 値を BLOCK_PATH に設定している場合、I/O パスの自動閉塞を実行します。

4

Test I/O で両系コントローラ異常を検出すると、syslogに障害レポートを通知します。

さらに、LUNへのアクセスが不可能になると、ノード切り替え等のコンフィグレーションで指定されたアクションを実行します。

I/O パスの死活管理について

TestI/O の実行結果として、以下の状態をレポートします。

- ・UP : TestI/O が正常終了し I/O パスが正常に動作している状態です。
- ・DOWN : TestI/O が異常終了し I/O パスが利用不可な状態です。

VG レベル (I/O パスをSPS 等のマルチパス管理製品で冗長化したGROUP 単位) のステータス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 を構成する片系の I/O パスに異常を検出した状態です。
- ・DOWN : VG に異常があり、利用不可な状態です。

I/O パスの組み込み状態を示すOnline status として、以下の状態をレポートします。

- ・extended : I/O パスが組み込まれた状態です。
- ・reduced : I/O パスが閉塞された状態です。
- ・unknown、alive : I/O パスの状態が不明です。

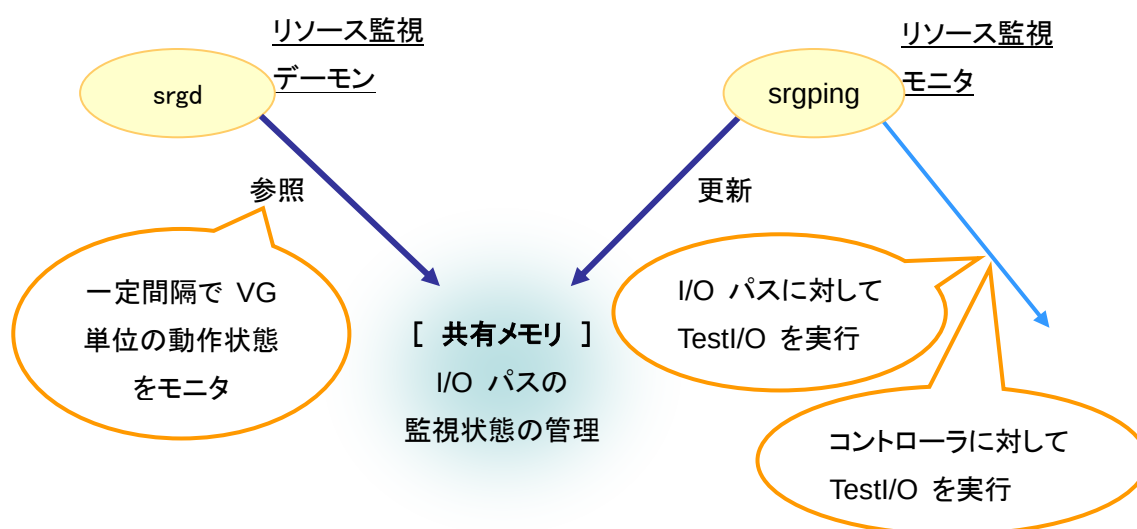
I/O パス監視機能でおこなっている死活監視および、障害レポート機能の動作は以下となります。

ディスク装置コントローラおよびLUN を構成する I/O パスに対して

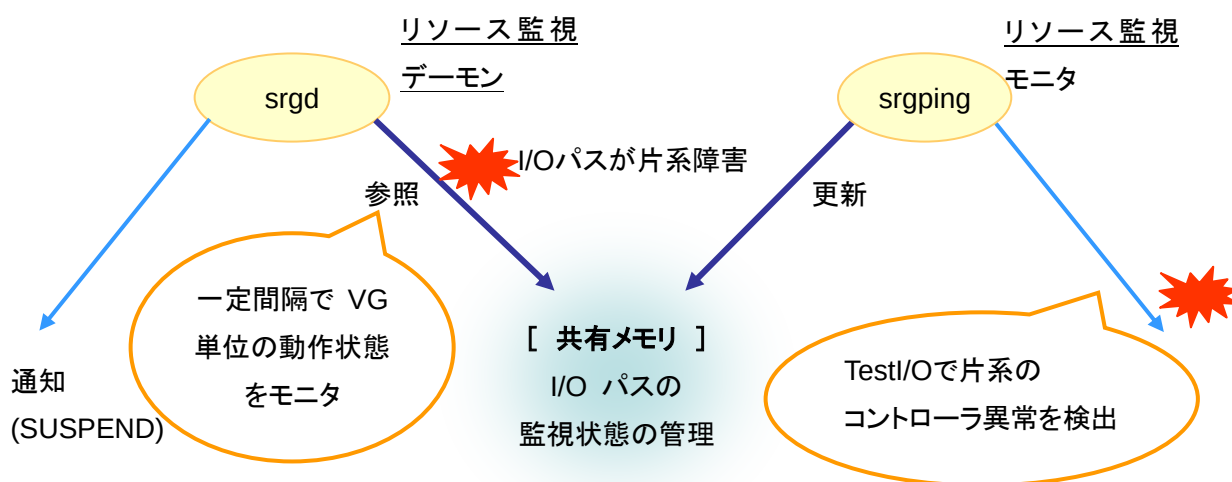
定期的に SCSI パススルーコマンドを利用して TestI/O を発行します。

TestI/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。

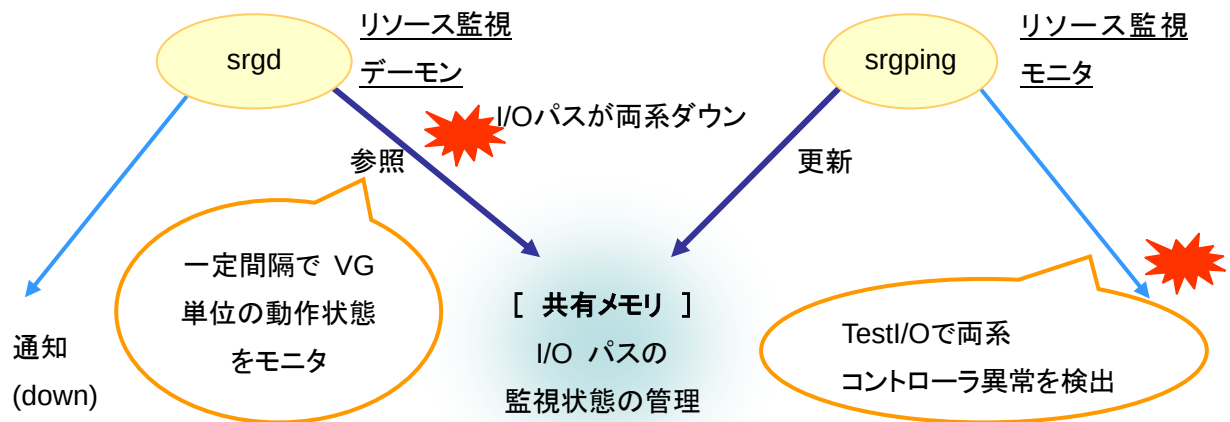
【TestI/O のフレームワーク】



【TestI/O で片系コントローラ異常を検出】

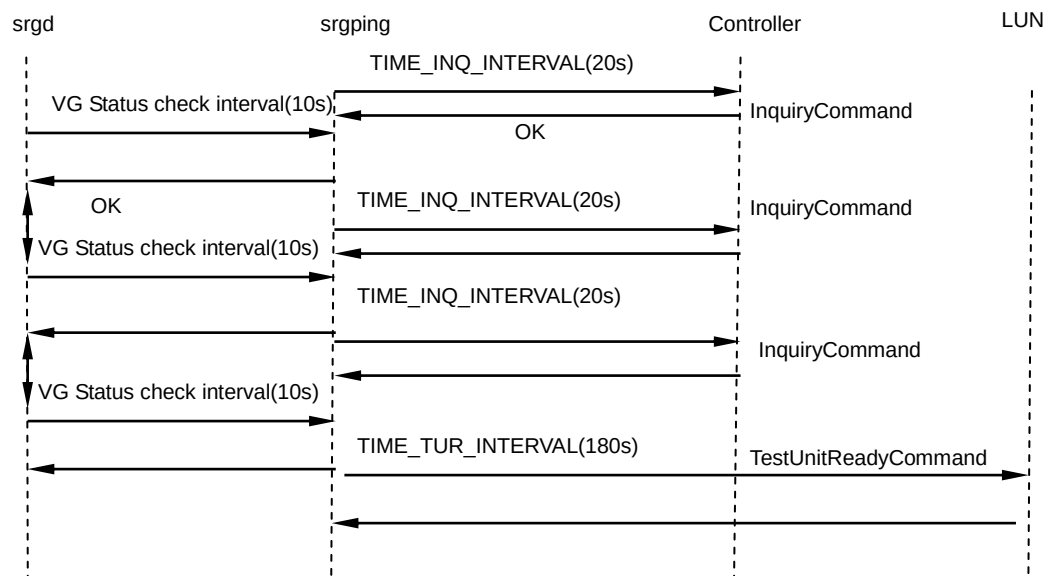


【TestI/O で両系コントローラ異常を検出】



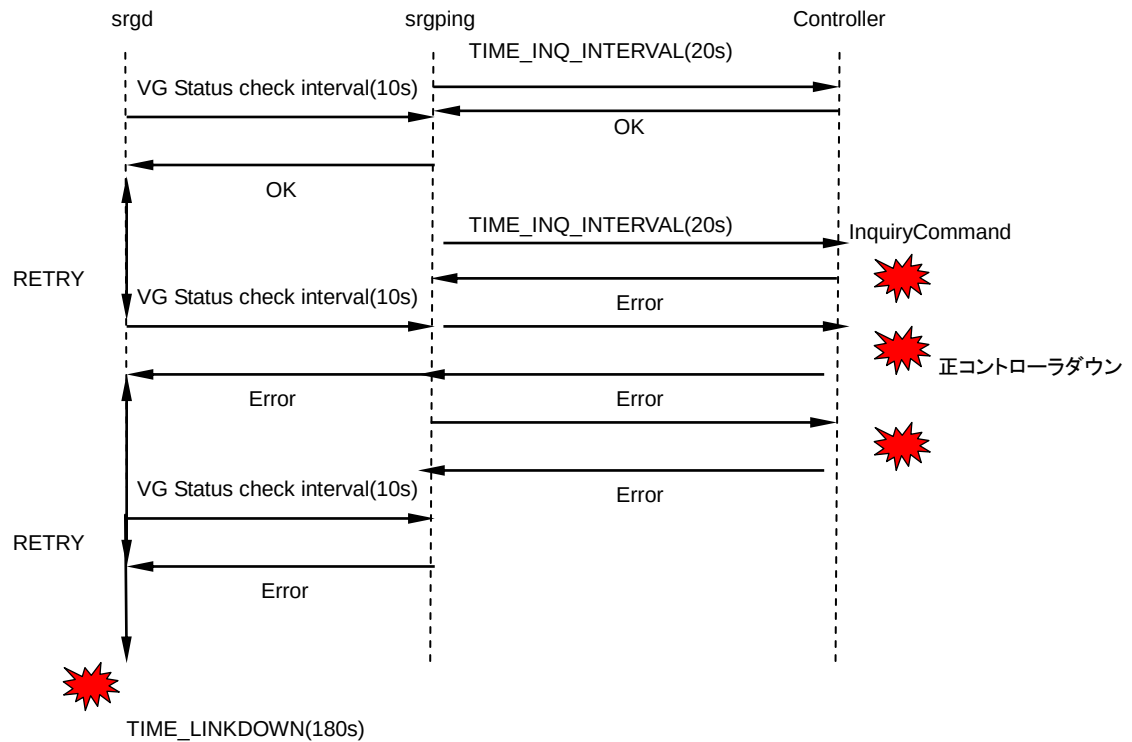
(1) Test I/O の正常なシーケンスは、以下のような動作になります。

VG 障害検出時間	: TIME_VG_FAULT	: 60(秒)
I/O パスの障害検出時間	: TIME_LINKDOWN	: 180(秒)
コントローラ監視間隔	: TIME_INQ_INTERVAL	: 20(秒)
LUN 監視間隔	: TIME_TUR_INTERVAL	: 180(秒)



(2) TestI/O で片系コントローラの異常を検出すると以下のような動作になります。

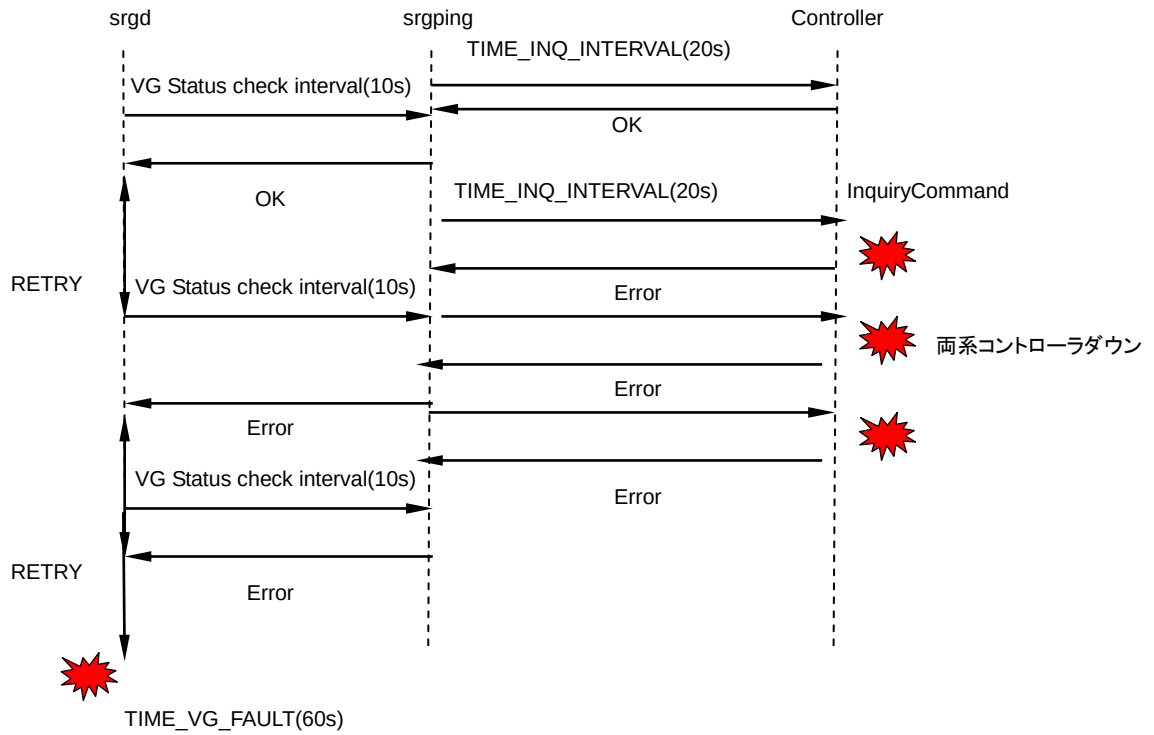
VG 障害検出時間	:	TIME_VG_FAULT	:	60(秒)
I/O パスの障害検出時間	:	TIME_LINKDOWN	:	180(秒)
コントローラ監視間隔	:	TIME_INQ_INTERVAL	:	20(秒)



タイムオーバ
障害の発生した FC 配下 VG suspend 通知

(3) TestI/O で両系コントローラの異常を検出すると以下のような動作になります。

VG 障害検出時間	:	TIME_VG_FAULT	:	60(秒)
I/O パスの障害検出時間	:	TIME_LINKDOWN	:	180(秒)
コントローラ監視間隔	:	TIME_INQ_INTERVAL	:	20(秒)



タイムオーバ
VG down 通知。クラスタウェアと連携して
予備系への切り替えを実施

3. 監視の設定と実践

それでは、一緒に StorageSaver を使用してみましょう！

3.1. インストール

まずは StorageSaver をインストールしましょう。

本製品は以下の手順でインストールします。

コードワードおよびインストールに関する詳細につきましては同梱の「CLUSTERPRO MC StorageSaver 1.1 for Linux リリースメモ」をご覧ください。

※ すでに本製品をインストール済みの場合は、以下の手順を行う必要はありません。

次のステップ(3.2 章)へ進んでください。

(1) rpm(8) コマンドにより、本製品のパッケージをインストールします。

※ 以下の手順は、CD-R 媒体を CD-ROM(DVD) 装置にマウントした場合の例です。

```
# rpm -ivh clusterpro-mc-ss-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：

- ◆ StorageSaver を使用する場合、sg3_utils がインストールされている必要があります。
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。
- ◆ Red Hat Enterprise Linux 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.el6.i686.rpm)がインストールされている必要があります。
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。

(2) rpm(8) コマンドにより、本製品が正しくインストールされたことを確認します。

```
# rpm -qa | grep clusterpro-mc-ss  
clusterpro-mc-ss-w.x.y-z
```

※w, x, y, z は任意のバージョン番号が入ります

以上で、本製品のインストールは終了です。

3.2. 設定ファイルの作成

StorageSaver で使用する設定ファイルを作成しましょう。

StorageSaver は以下のファイルを使用します。

- ◆ /var/opt/HA/SrG/conf/srg.config
- ◆ /var/opt/HA/SrG/conf/srg.map
- ◆ /var/opt/HA/SrG/conf/srg.rsc

設定ファイルはノード単位に定義するファイルで、監視間隔や障害検出時のアクション、監視対象のデバイス情報等を定義します。

設定ファイル自動生成コマンド /opt/HA/SrG/bin/srgquery(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。すでに設定ファイルが存在する場合は、自動で上書きします。

```
# /opt/HA/SrG/bin/srgquery -s <格納ディレクトリ>
```

注意：

- ◆ 設定ファイル自動生成機能は設定作業軽減のため設定ファイルのテンプレートを作成する機能であり、すべてのディスク構成をサポートしているわけではありません。したがって、構成によっては作成できない場合があります。必ず作成された設定ファイルを確認し、実際の環境と差異がある場合には手動で修正してください。
- ◆ SANboot 環境において設定ファイル自動生成コマンド (srgquery) にて設定ファイルを生成した場合、OS 情報が格納されたディスク装置も FC 接続構成であるため、他のデータディスク装置同様、監視対象として定義します。設定ファイルから OS 情報が格納されたディスク装置の記述を手動で削除してください。
- ◆ IDE 接続のデバイスが接続されている環境において設定ファイル自動生成コマンド(srgquery) にて設定ファイルを生成した際、IDE 接続のデバイスを監視対象として定義する場合があります。IDE 接続のデバイスが設定ファイルに含まれている場合は手動で削除してください。

3.3. 監視の開始

StorageSaver によるディスク装置監視を開始しましょう。

- (1) StorageSaver を起動して、監視をはじめます。

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。

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

- (2) ps(1) コマンドにより StorageSaver のプロセス(srgping, srgd, srgwatch)が起動していることを確認します。

```
# ps -ef | grep srg
root 8720 8098 0 10:44:40 ? 0:00 srgping
root 8098      1 0 10:43:40 ? 0:00 /opt/HA/SrG/bin/srgd
root 8111      1 0 10:43:40 ? 0:00 /opt/HA/SrG/local/bin/srgwatch
```

注意:srgping は srgd を起動後、約 1 分後に起動されます。

3.4. 監視状態の確認

srgadmin コマンドにてディスク装置の監視状態を確認しましょう。

監視リソースおよび各リソースの監視状態を表示します。

```
# /opt/HA/SrG/bin/srgadmin
(monitor status = TRUE)
```

type	device	HostBusAdapter	L status	P status	Online status
VG	VolGroup01	---	up		
PV	/dev/sdj	pci-0000:13:00.0	up	up	extended
PV	/dev/sdh	pci-0000:13:00.0	up	up	extended
PV	/dev/sdf	pci-0000:13:00.1	up	up	extended
PV	/dev/sdd	pci-0000:13:00.1	up	up	extended
VG	VolGroup02	---	up		
PV	/dev/sdk	pci-0000:13:00.0	up	up	extended
PV	/dev/sdi	pci-0000:13:00.0	up	up	extended
PV	/dev/sdg	pci-0000:13:00.1	up	up	extended
PV	/dev/sde	pci-0000:13:00.1	up	up	extended

# /opt/HA/SrG/bin/srgadmin (monitor status = TRUE)					
リソース監視の有効／無効を表示します					
=====					
type	: device	: HostBusAdapter	: L status	: P s	
=====					
VG	: VolGroup01	: ---	: up		
PV	: /dev/sdj	: pci-0000:13:00.0	: up	: up	: extended
PV	: /dev/sdh	: pci-0000:13:00.0	: up	: up	: extended
PV	: /dev/sdf	: pci-0000:13:00.1	: up	: up	: extended
PV	: /dev/sdd	: pci-0000:13:00.1	: up	: up	: extended
VG	: VolGroup02	: ---			
PV	: /dev/sdk	: pci-0000:13:00.0	: up	: up	: extended
PV	: /dev/sdi	: pci-0000:13:00.0	: up	: up	: extended
PV	: /dev/sdg	: pci-0000:13:00.1	: up	: up	: extended
PV	: /dev/sde	: pci-0000:13:00.1	: up	: up	: extended

① VG の監視状態を表示します。

up	VG または PKG を構成するすべての I/O パスが正常に動作している状態です。
suspend	VG または PKG を構成する I/O パスの一部に異常があり、片パスで運用されている状態、または、保守員によるオンライン保守実施中のため、監視を停止している状態です。
down	VG または PKG を構成する I/O パスに異常があり、利用不可の LUN がある状態です。

② I/O パスの論理ステータス(管理状態)を L status として表示します。

表示	意味
up	正常動作中
down	障害状態
-	監視停止中

③ I/O パスの物理ステータス(TestI/O の実行状態)を P status として表示します。

表示	意味
up	正常動作中
down	障害状態
-	監視停止中

④ I/O パスの組み込み状態(Online status)を表示します。

表示	意味
extended	組み込み済み
reduced	閉塞状態
alive	状態不明
unknown	状態不明

3.5. ディスク障害検知の確認

srgadmin コマンドにてディスク障害を擬似的に発生させ、障害が検出されることを確認しましょう。

本章では、I/O パスの片系故障と両系故障における障害検出の動作を確認します。

◆ I/O パスの片系故障

(1) 障害前に現在の状態をモニタリングします。

```
# /opt/HA/SrG/bin/srgadmin -c status
(monitor status = TRUE)
```

type	device	HostBusAdapter	L status	P status	Online status
VG	VolGroup01	---	up		
PV	/dev/sdj	pci-0000:13:00.0	up	up	extended
PV	/dev/sdh	pci-0000:13:00.0	up	up	extended
PV	/dev/sdf	pci-0000:13:00.1	up	up	extended
PV	/dev/sdd	pci-0000:13:00.1	up	up	extended
VG	VolGroup02	---	up		
PV	/dev/sdk	pci-0000:13:00.0	up	up	extended
PV	/dev/sdi	pci-0000:13:00.0	up	up	extended
PV	/dev/sdg	pci-0000:13:00.1	up	up	extended
PV	/dev/sde	pci-0000:13:00.1	up	up	extended

(2) srgadmin -c debug にて、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/SrG/bin/srgadmin -c debug -v on -F 0000:13:00.1
Change debug value.
FC devfile = 0000:13:00.1
0 -> 1
FC devfile = 0000:13:00.1
0 -> 1
FC devfile = 0000:13:00.1
0 -> 1
FC devfile = 0000:13:00.1
0 -> 1
```

約 180 秒後

```
# /opt/HA/SrG/bin/srgadmin -c status
(monitor status = TRUE)
```

type	device	HostBusAdapter	L status	P status	Online status
VG	VolGroup01	---	suspend		
PV	/dev/sdj	pci-0000:13:00.0	up	up	extended
PV	/dev/sdh	pci-0000:13:00.0	up	up	extended
PV	/dev/sdf	pci-0000:13:00.1	down	down	extended
PV	/dev/sdd	pci-0000:13:00.1	down	down	extended
VG	VolGroup02	---	suspend		
PV	/dev/sdk	pci-0000:13:00.0	up	up	extended
PV	/dev/sdi	pci-0000:13:00.0	up	up	extended
PV	/dev/sdg	pci-0000:13:00.1	down	down	extended
PV	/dev/sde	pci-0000:13:00.1	down	down	extended

(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。

```
srgd[xxxxx]: PV status change fail .[hwpath
=pci-0000:13:00.1-fc-0x2100001697120ca7:0x0001000000000000: s.f =/dev/sdg].
srgd[xxxxx]: PV status change fail .[hwpath
=pci-0000:13:00.1-fc-0x2900001697120ca7:0x0001000000000000: s.f =/dev/sde].
srgd[xxxxx]: PV status change fail .[hwpath
=pci-0000:13:00.1-fc-0x2100001697120ca7:0x0000000000000000: s.f =/dev/sdf].
srgd[xxxxx]: PV status change fail .[hwpath
=pci-0000:13:00.1-fc-0x2900001697120ca7:0x0000000000000000: s.f =/dev/sdd].
```

◆ I/O パスの両系故障

(1) 障害前に現在の状態をモニタリングします。

既に片系の I/O パスが異常であることを確認します。

```
# /opt/HA/SrG/bin/srgadmin -c status
(monitor status = TRUE)
```

type	device	HostBusAdapter	L status	P status	Online status
VG	VolGroup01	---	suspend		
PV	/dev/sdj	pci-0000:13:00.0	up	up	extended
PV	/dev/sdh	pci-0000:13:00.0	up	up	extended
PV	/dev/sdf	pci-0000:13:00.1	down	down	extended
PV	/dev/sdd	pci-0000:13:00.1	down	down	extended
VG	VolGroup02	---	suspend		
PV	/dev/sdk	pci-0000:13:00.0	up	up	extended
PV	/dev/sdi	pci-0000:13:00.0	up	up	extended
PV	/dev/sdg	pci-0000:13:00.1	down	down	extended
PV	/dev/sde	pci-0000:13:00.1	down	down	extended

(2) 両系の I/O パスを異常状態にします。

```
# /opt/HA/SrG/bin/srgadmin -c debug -v on -F 0000:13:00.0
Change debug value.
FC devfile = 0000:13:00.0
0 -> 1
FC devfile = 0000:13:00.0
0 -> 1
FC devfile = 0000:13:00.0
0 -> 1
FC devfile = 0000:13:00.0
0 -> 1
```

約 60 秒後

```
# /opt/HA/SrG/bin/srgadmin -c status
(monitor status = TRUE)
=====
type : device : HostBusAdapter : L status : P status : Online status
=====
VG : VolGroup01 : --- : down
PV : /dev/sdj : pci-0000:13:00.0 : up : down : extended
PV : /dev/sdh : pci-0000:13:00.0 : up : down : extended
PV : /dev/sdf : pci-0000:13:00.1 : down : down : extended
PV : /dev/sdd : pci-0000:13:00.1 : down : down : extended
VG : VolGroup02 : --- : down
PV : /dev/sdk : pci-0000:13:00.0 : up : down : extended
PV : /dev/sdi : pci-0000:13:00.0 : up : down : extended
PV : /dev/sdg : pci-0000:13:00.1 : down : down : extended
PV : /dev/sde : pci-0000:13:00.1 : down : down : extended
```

(3) 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。

```
srgd[xxxxx]: VG status change down .(vg=VolGroup01)
srgd[xxxxx]: VG status change down .(vg=VolGroup02)
```

両系のディスクのステータスが down になったため、VG 状態も down となります。

以上で、ディスク障害の検知の確認は終了です。

3.6. 監視の復旧

3.5 章で発生させた擬似障害状態を復旧させます。

I/O パスの両系故障

(1) srgadmin コマンドにてデバッグモードを off にします。

```
# /opt/HA/SrG/bin/srgadmin -c debug -v off -F 0000:13:00.1
Change debug value.
FC devfile = 0000:13:00.1
1 -> 0
FC devfile = 0000:13:00.1
1 -> 0
FC devfile = 0000:13:00.1
1 -> 0
FC devfile = 0000:13:00.1
1 -> 0
# /opt/HA/SrG/bin/srgadmin -c debug -v off -F 0000:13:00.0
Change debug value.
FC devfile = 0000:13:00.0
1 -> 0
FC devfile = 0000:13:00.0
1 -> 0
FC devfile = 0000:13:00.0
1 -> 0
FC devfile = 0000:13:00.0
1 -> 0
```

(2) srgrecover -v を実行して構成を復旧させます。

```
# /opt/HA/SrG/bin/srgrecover -v
HW-path: pci-0000:13:00.0-fc-0x2100001697120ca7:0x0000000000000000 result: OK
HW-path: pci-0000:13:00.0-fc-0x2100001697120ca7:0x0001000000000000 result: OK
HW-path: pci-0000:13:00.0-fc-0x2900001697120ca7:0x0000000000000000 result: OK
HW-path: pci-0000:13:00.0-fc-0x2900001697120ca7:0x0001000000000000 result: OK
HW-path: pci-0000:13:00.1-fc-0x2100001697120ca7:0x0000000000000000 result: OK
HW-path: pci-0000:13:00.1-fc-0x2100001697120ca7:0x0001000000000000 result: OK
HW-path: pci-0000:13:00.1-fc-0x2900001697120ca7:0x0000000000000000 result: OK
HW-path: pci-0000:13:00.1-fc-0x2900001697120ca7:0x0001000000000000 result: OK
srgrecover complete.
```

(3) 復旧後、srgadmin コマンドにて正常にディスク監視が行われていることを確認します。

```
# /opt/HA/SrG/bin/srgadmin -c status
(monitor status = TRUE)
```

type	device	HostBusAdapter	L status	P status	Online status
VG	VolGroup01	---	up		
PV	/dev/sdj	pci-0000:13:00.0	up	up	extended
PV	/dev/sdh	pci-0000:13:00.0	up	up	extended
PV	/dev/sdf	pci-0000:13:00.1	up	up	extended
PV	/dev/sdd	pci-0000:13:00.1	up	up	extended
VG	VolGroup02	---	up		
PV	/dev/sdk	pci-0000:13:00.0	up	up	extended
PV	/dev/sdi	pci-0000:13:00.0	up	up	extended
PV	/dev/sdg	pci-0000:13:00.1	up	up	extended
PV	/dev/sde	pci-0000:13:00.1	up	up	extended

以上で、監視の復旧は終了です。

3.7. 監視の終了

StorageSaver によるディスク装置監視の終了手順を確認しましょう。

- (1) StorageSaver を停止して、監視を終了します。

以下のコマンドを root 権限で投入することで停止できます。

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

- (2) ps(1) コマンドにより StorageSaver のプロセス(srgd, srgping, srgwatch)が停止していることを確認します。

srgd、srgping および srgwatch が表示されないことを確認します。

```
# ps -ef | grep srg
```

4. おわりに

以上、StorageSaver の基本的な動作となります。

この他にも I/O ストールの監視や、クラスタウェアとの連携など、システムにあわせた監視を行うことができます。

各機能の詳細については「CLUSTERPRO MC StorageSaver 1.2 for Linux ユーザーズガイド」をご覧ください。

CLUSTERPRO
MC StorageSaver 1.2 for Linux
はじめての StorageSaver

2014 年 3 月 第 3 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)

P

© NEC Corporation 2014

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