

# CLUSTERPRO システム構築ガイド

---

CLUSTERPRO<sup>®</sup> for Linux Ver 2.0

---

GFS編

第2版 2002.3.6

改版履歴

版 数	改版年月日	改版ページ	内 容
第1版	2002.2.19	-	先行リリース用.
第2版	2002.3.6	-	IA-64固有部分を削除. GFSに依存した記述のみに変更.

# 1 概要

本書はAzusa IA-64 Linux用のCLUSTERPRO for Linux Ver2.0でGFSの二重化運用の注意事項ををまとめたものです。

インストール及び構築作業前に必ずお読みください。

## 2 注意制限事項

GFSサーバ機能を二重化する際に留意して頂きたい事項です。

- \* 共有ディスク上の物理的に同一のLUNが両ノードから同じメジャー・マイナー番号で参照できるようにしてください。
- \* フェイルオーバー発生時にクライアントからのNFS経由のアクセスがタイムアウトする可能性があります。  
タイムアウトが発生してもクライアント側から再マウントの必要はありません。  
2/19現在 NFSクライアントのタイムアウトのパラメータの調整可否については調査中です。
- \* フェイルオーバー発生時にクライアントからのGFS経由のアクセスがタイムアウトする場合にはGFSのタイムアウト値を変更してください。  
フェイルオーバーに必要な時間はユーザ環境によって異なりますので必ず実環境でフェイルオーバーを発生させて問題がないことを確認してください。
- \* オンラインフェイルバックを実行しないでください。  
オンラインフェイルバックを実行すると共有ディスク上のファイルシステムの切り離しに失敗してサーバがシャットダウンする場合があります。  
サーバダウンが発生してフェイルオーバーグループがフェイルオーバーし、ダウンしたサーバが復旧した後にフェイルオーバーグループをフェイルバック(本来動作すべきサーバへ戻すこと)を行う場合には以下の手順に従ってください。
  1. CLUSTERPROマネージャ(管理用GUI)から復旧したサーバを復帰する
  2. CLUSTERPROマネージャ上で上記のサーバが緑色(正常状態)に変わるのが確認する
  3. CLUSTERPROマネージャから(サブ)クラスタまたはスーパクラスタのシャットダウン、再起動を選択する
  4. 両サーバがシャットダウンし、再起動する
  5. 各フェイルオーバーグループがプライマリサーバで動作する
- \* フローティングIP(クライアントからシームレスにアクセスするためのIP)の設定はフェイルオーバーグループのリソースに登録しないでください。  
CLUSTERRPOからGFSを動作/停止させるためのスクリプト上に記述します。本書4章を参照ください。

### 3 GFSの二重化について

#### 3.1 動作概要

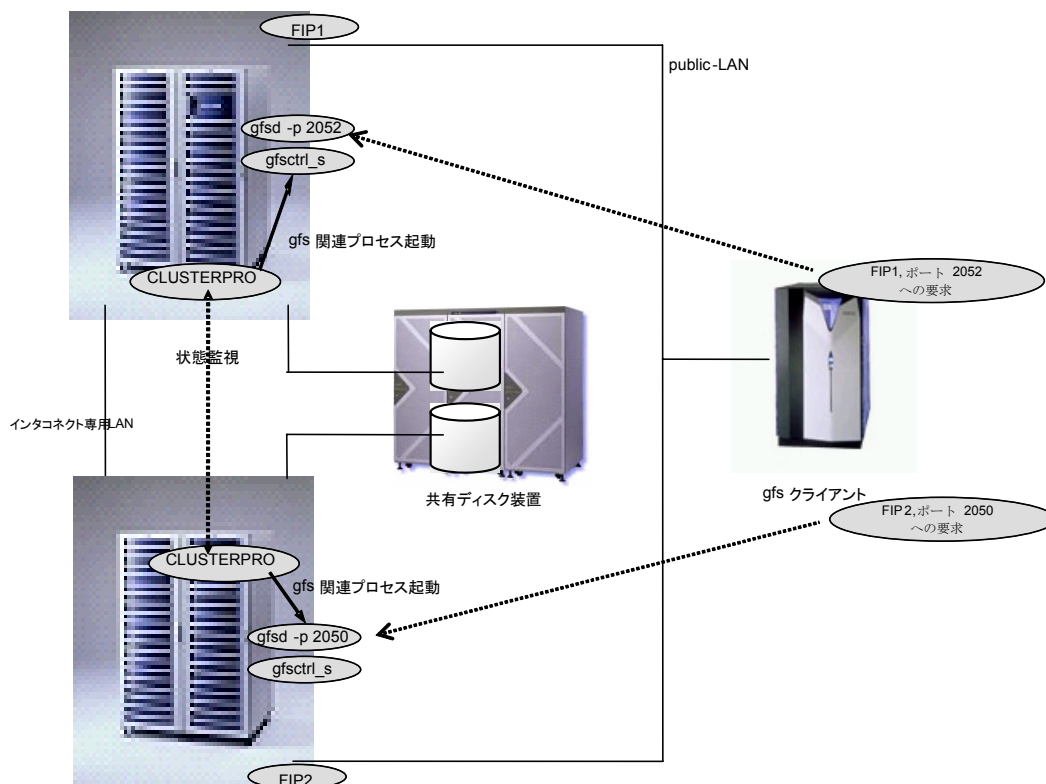
\* 正常運用時

GFSのクライアント側からはCLUSTERPROが提供するフローティングIPでアクセスします。

GFS関連のプロセスの起動/停止とフローティングIPの活性化/非活性化はフェイルオーバーグループのstartスクリプト/stopスクリプトから行います<sup>1</sup>。共有ディスク上のファイルシステムの制御(マウント/アンマウント)はCLUSTERPROが内部で行います。

双方向スタンバイの場合にはインスタンス毎にGFSのポート番号を変更するのでGFSクライアント側からは該当するインスタンスのポート番号を指定してアクセスをしてください。

#### ■正常運用時



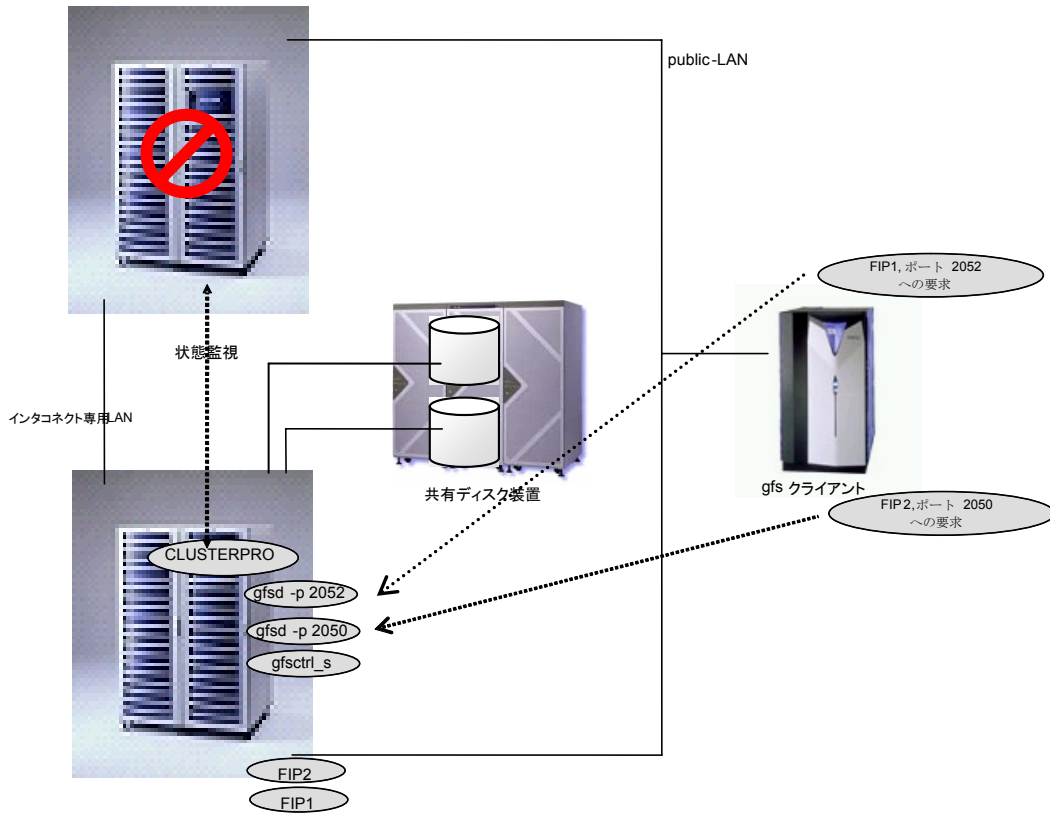
<sup>1</sup> gfsiod\_sのみ後述のように/etc/rc.d/init.d/gfsにて起動/停止をしてください。

\* 片サーバダウン時

相手サーバの**CLUSTERPRO**から応答がなくなることをトリガとして**FIP**と共有ディスク上のファイルシステムを正常なサーバ側へ移動します。

さらにダウンしたサーバで動作していた**GFS**関連のプロセスと**FIP**を起動します。**GFS**クライアント側からは正常時と同じ**FIP**とポート番号でアクセスできます。

■片サーバダウン時



## 3.2 GFSリソースの設計

双方向スタンバイにする場合には、GFSに依存して以下のリソースの計画を立ててください。

共有ディスク上にはCLUSTERパーティションも必要です。クラススタ生成ガイドを参照して各LUNにCLUSTERパーティションも確保してください。

LUNを複数使用する構成でもCLUSTERパーティションはクラススタ全体で1つで構いません。

下表のリソースをグループ1用、グループ2用にアサインしてください

	フローティングIPアドレス	<p>GFSクライアントからシームレスにアクセスするために使用するIPアドレス (サーバの実IPと同じセグメントで他のIPアドレスと重複しないIPアドレスが必要)</p> <p>GFS用のスクリプトで使用します</p> <p>例) group1用:192.168.1.3 group2用:192.168.1.4</p>
	共有ディスクのLUN全体を示すデバイス名	<p>CLUSTERPRO側のフェイルオーバーグループの切替単位はファイルシステム単位ですが、<code>gfsetrl_s</code>のインスタンスがLUN単位なのでフェイルオーバーグループ間でLUNが重複しないようにしてください</p> <p>GFS用のスクリプトで使用します</p> <p>1つのフェイルオーバーグループ内で複数のLUNを使用することも可能です。</p> <p>例) group1用:/dev/scsi/host2/bus0/target16/lun0/disc group2用:/dev/scsi/host2/bus0/target17/lun1/disc</p>
	GFSクライアント(SX)から見たスペシャルファイル名	<p>GFSのクライアント(SX)側から見たGFS資源のデバイス名です</p> <p>GFS用のスクリプトで使用します。</p> <p>LUN全体を示すデバイス名毎に決まります。使用するLUN数分アサインしてください。</p> <p>例) group1用: /dev/scd/c012t016d000 group2用: /dev/scd/c013t017d001</p>
	共有ディスク上のパーティション	<p>GFSでshareするファイルシステムを構築するパーティションデバイスのスペシャルファイル</p>

	(切替パーティション)	<p>名</p> <p>CLUSTERPROの/etc/clusterpro/fstabで使用します。</p> <p>使用するLUN上でファイルシステムを構築するパーティションです。 1つのフェイルオーバグループに複数の切替パーティションを使用することが可能です。</p> <p>例) group1用:/dev/scsi/host2/bus0/target16/lun0/part5 group2用:/dev/scsi/host2/bus0/target17/lun1/part5</p>
	マウントポイント	<p>上記ファイルシステムをマウントするマウントポイント</p> <p>GFS用のスクリプトで使用します。 CLUSTERPROの/etc/clusterpro/fstabでも使用します。</p> <p>切替パーティションの数分をアサインしてください。</p> <p>例) group1用:/mnt/lun0/part5 group2用:/mnt/lun1/part5</p>
	GFSのポート番号	<p>gfsdが使用するポート番号</p> <p>例) group1用:2050番 group2用:2052番</p>

### 3.3 GFSスクリプトの修正

start.batとstop.batはスクリプトはユーザの実環境に合わせて作成をしてください。

標準のinitスクリプト(/etc/rc.d/init.d/gfs)の修正も必要です。

下記のスクリプトのH/C中の番号①②...などは説明上、付加したものです。リリースしたスクリプトの雛形には付加されていません。



### 3.3.1 構成例 1

フェイルオーバーグループ毎に共有ディスク装置のLUNがアサインできる時の例です。  
フェイルオーバーグループが2つ必要ない場合(片方向スタンバイ形式での運用の場合)にはフェイルオーバーグループを1つだけ作成してください。

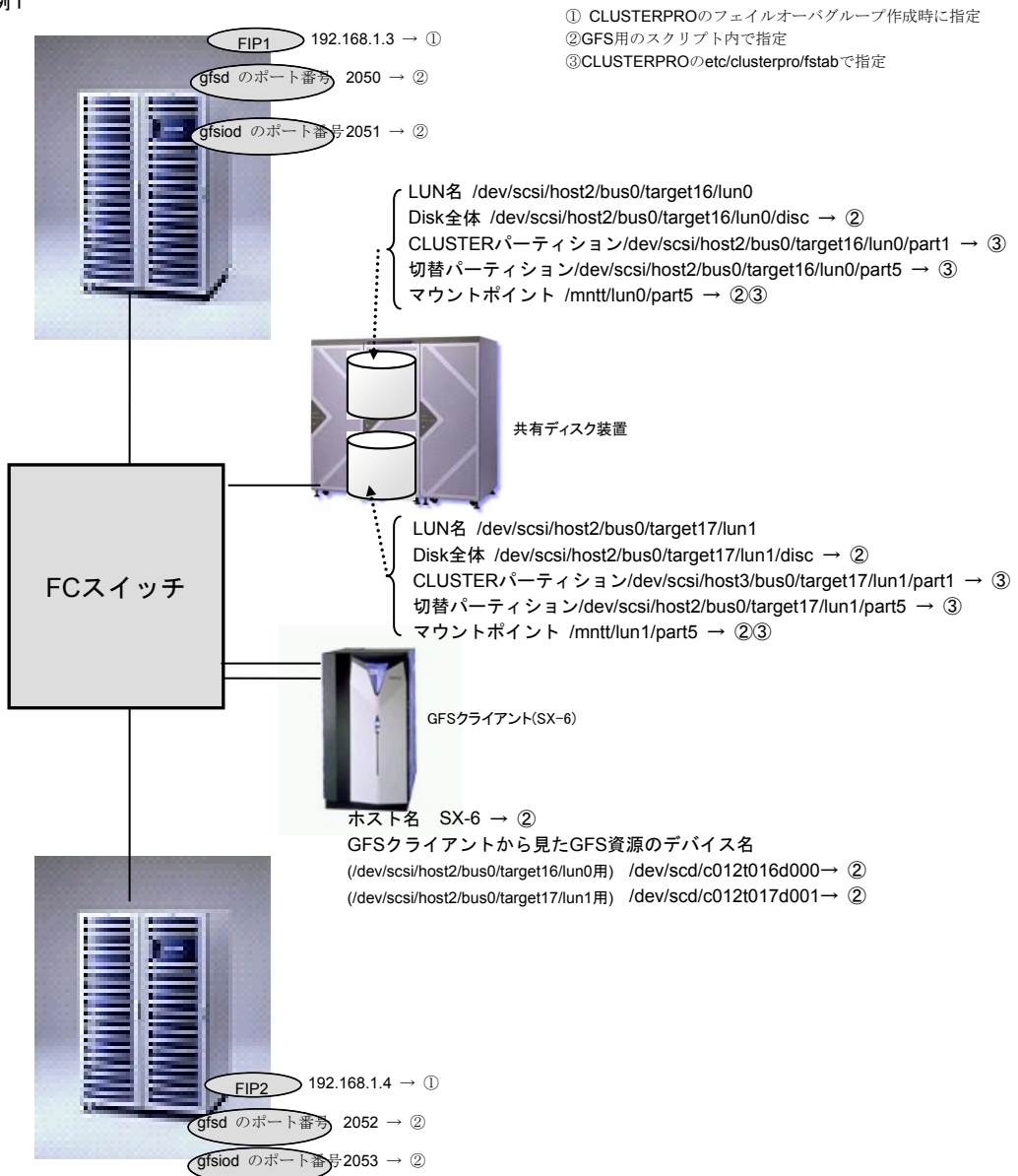
#### フェイルオーバーグループ1

FIP	192.168.1.3
LUN全体のデバイス名	/dev/scsi/host2/bus0/target16/lun0/disc
切替パーティション	/dev/scsi/host2/bus0/target16/lun0/part5
マウントポイント	/mnt/lun0/part5
SXから見たスペシャルファイル名	/dev/scd/c012t016d000
gfsdが使用するポート番号	2050
gfsiod_sが使用するポート番号	2051

#### フェイルオーバーグループ2

FIP	192.168.1.4
LUN全体のデバイス名	/dev/scsi/host2/bus0/target17/lun1/disc
切替パーティション	/dev/scsi/host2/bus0/target17/lun1/part5
マウントポイント	/mnt/lun1/part5
SXから見たスペシャルファイル名	/dev/scd/c012t017d001
gfsdが使用するポート番号	2050
gfsiod_sが使用するポート番号	2053

■構成例1



### 3.3.1.1 標準のinitスクリプト

Linux側のOSのGFSのinitスクリプトのstart(), stop()のルーチンを以下のように修正してください。

start()では/usr/sbin/gfsiod\_sだけを起動するようにします。

stop()ではgfsiod\_sを停止するようにします。下記の修正例を参考にしてください。

(オリジナルのGFSのinitスクリプトは起動しないようにして、オリジナルの/etc/rc.d/init.d/gfsを元にコピーを作成して、使用するLinuxのランレベルの/etc/rc.d/rc.X/から新たにシンボリックリンクを貼ることをお奨めします。)

#### \* サンプルスクリプトの説明

下記のstart(), stop()ルーチンは**あくまでサンプルですので実環境に合わせて作成**をしてください。

下記を環境に合わせて修正してください。

#### ①gfsiodのポート番号

```
start() {
    if [ $GFSSRV = 1 ]; then
        # Start daemons.
        echo -n "Starting GFS IO daemon (server) : "
        daemon /usr/sbin/gfsiod_s -s ${GFSDSCOUNT}
        echo
        sleep 1
        echo -n "Starting GFS IO daemon (receiver) : "
        daemon /usr/sbin/gfsiod_s -r -p 2051 ${GFSDSCOUNT}
        echo
        fi

        touch /var/lock/subsys/gfs
    }
stop() {
    # Stop daemons.
    echo -n "Shutting down GFS daemon: "
    PID=`ps -e | egrep 'gfsiod' | sed -e 's/^ *//' -e 's/ .*//'`
    if [ "$PID" != "" ]; then
        kill -9 ${PID}
    fi
    echo
}
}
```

### 3.3.1.2 start.bat(起動スクリプト)

start.batではGFSに依存して基本的には下記の記述が必要です。

- + GFS で使用するディスク装置の設定 (gfsctl\_s)
- + GFS デーモンの起動 (gfsd)
- + クライアントがマウントするディレクトリのエクスポート(exports)
- + FIPの活性化

\* サンプルスクリプトの説明

次ページのstart.batは**あくまでサンプルですので実環境に合わせて作成**をしてください。

/usr/sbin/gfsd から /usr/sbin/gfsctl\_s までがGFS依存の起動処理です。下記を環境とフェイルオーバーグループによって修正してください。

- ①gfsdのポート番号 (\*)
- ②GFSクライアント(SX)のホスト名
- ③共有ディスク上のファイルシステムのマウントポイント (\*)
- ④FIPのIPアドレス (\*)
- ⑤GFSでshareするLUN全体を示すデバイス名 (\*)
- ⑥SX側から見たスペシャルファイル名 (\*)

\* 次ページのstart.bat内の上記(\*)は構成例1のフェイルオーバーグループ1を想定した値になっています。フェイルオーバーグループ2用のスクリプトを作成するときにはフェイルオーバーグループ間で重複しないように予め計画をしてください。

\* 1つのフェイルオーバーグループ内に複数の切替パーティションを設定するときには★の行を切替パーティションごとに追加してください。(1つのフェイルオーバーグループ内で複数LUNを使用する時には下記の\*も参照してください)

\* 1つのフェイルオーバーグループ内に複数のLUNを設定する時には☆の行をLUNごとに追加してください。

```

#!/bin/sh
#*****
#*           START.BAT           *
#*****

if [ "$ARMS_EVENT" = "START" ]
then
  if [ "$ARMS_DISK" = "SUCCESS" ]
  then
    armlog "NORMAL1"
    ↓①
    /usr/sbin/gfsd -p 2050 4                ↓② ↓③
    /usr/sbin/exportfs -o rw,insecure,no_root_squash SX-6/mnt/lun0/part5
    /bin/chmod 777 /mnt/lun0/part5                ☆
    ↑③
    # start FIP                                ↓④
    /usr/clusterpro/armmontr/bin/fipcmd64 -a 192.168.1.3

    if [ "$ARMS_SERVER" = "HOME" ]
    then
      armlog "NORMAL2"
    else
      armlog "ON_OTHER1"
    fi

    ↓⑤
    /usr/sbin/gfsctl_s -o /dev/scsi/host3/bus0/target16/lun0/disc SX-6 /dev/scd/c015t016d001 ☆
  else
    armlog "ERROR_DISK from START"
  fi
elif [ "$ARMS_EVENT" = "RECOVER" ]
then
  armlog "RECOVER"
elif [ "$ARMS_EVENT" = "FAILOVER" ]
then
  if [ "$ARMS_DISK" = "SUCCESS" ]
  then
    armlog "FAILOVER1"
    ↓①
    /usr/sbin/gfsd -p 2050 4                ↓② ↓③
    /usr/sbin/exportfs -o rw,insecure,no_root_squash SX-6/mnt/lun0/part5                ☆
    /bin/chmod 777 /mnt/lun0/part5                ☆
    ↑③
    # start FIP                                ↓④
    /usr/clusterpro/armmontr/bin/fipcmd64 -a 192.168.1.3

    if [ "$ARMS_SERVER" = "HOME" ]
    then
      armlog "FAILOVER2"
    else
      armlog "ON_OTHER2"
    fi

    ↓⑤
    /usr/sbin/gfsctl_s -o /dev/scsi/host3/bus0/target16/lun0/disc SX-6 /dev/scd/c015t016d001 ☆
  else
    armlog "ERROR_DISK from FAILOVER"
  fi
else
  armlog "NO_ARM"
fi
armlog "EXIT"
exit 0

```

### 3.3.1.3 stop.bat(停止スクリプト)

stop.batではGFSに依存して基本的には下記の記述が必要です。

- + FIPの非活性化
- + GFS で使用するディスク装置の解除 (gfsctl\_s)
- + GFS デーモンの終了 (gfsd)
- + クライアントがマウントするディレクトリのアンエクスポート(exports)

\* サンプルスクリプトの説明

次ページのstop.batは **あくまでサンプルですので実環境に合わせて作成**をしてください。

fipcmd64からexportsfs -uまでがGFSに依存した停止処理です。下記を環境とフェイルオーバーグループによって修正してください。

- ①gfsdのポート番号 (\*)
- ②GFSクライアントのホスト名
- ③共有ディスク上のファイルシステムのマウントポイント (\*)
- ④フローティングFIPアドレス (\*)
- ⑤LUN全体を示すデバイス名 (\*)
- ⑥SX側から見たスペシャルファイル名 (\*)
- ⑦GFS,NFS関連プロセスの停止待ちのための待ち時間(秒) <sup>2</sup>

\* 次ページのstop.bat内の上記(\*)は構成例1のフェイルオーバーグループ1を想定した値になっています。フェイルオーバーグループ2用のスクリプトを作成するときにはフェイルオーバーグループ間で重複しないように予め計画をしてください。

\* 1つのフェイルオーバーグループ内に複数の切替パーティションを設定するときには★の行を切替パーティションごとに追加してください。(1つのフェイルオーバーグループ内で複数LUNを使用する時には下記の\*も参照してください)

\* 1つのフェイルオーバーグループ内に複数のLUNを設定する時には★の行をLUNごとに追加してください。

\* ⑦の待ち時間は環境に調整が必要です。

GFSクライアント側からファイルにアクセスした状態でフェイルオーバーグループを停止して“ディスクの切り離し失敗”が発生<sup>3</sup>しないように調整をしてください。

---

<sup>2</sup> 共有ディスク上のファイルシステムを使用しているプロセスが完全に停止するのを待つためです

<sup>3</sup> マネージャのアラートウィンドウに X-CALL disk was not disconnected.x (SCSI-PORT:x SCSI-BUS:x SCSI-ID:x LUN:x DRIVE:x code<x>) ID 4212 が表示されてサーバがシャットダウンします

```

#!/bin/sh
#*****
#*          STOP.BAT          *
#*****
:
: 省略
:
if [ "$ARMS_EVENT" = "START" ]
then
  if [ "$ARMS_DISK" = "SUCCESS" ]
  then
    armlog "NORMAL1"
    if [ "$ARMS_SERVER" = "HOME" ]
    then
      armlog "NORMAL2"
    else
      armlog "ON_OTHER1"
    fi

    # stop FIP
    /usr/clusterpro/armmontr/bin/fipcmd64 -d 192.168.1.3

    /usr/sbin/gfsctl_s -r /dev/scsi/host3/bus0/target16/lun0/disc SX-6 /dev/scd/c015t016d000

    PID=`ps -e | egrep 'gfsd' | egrep '2050' | awk '{print $1}'`
    if [ "$PID" != "" ]; then
      kill -9 ${PID}
    fi

    /usr/sbin/exportfs -u SX-6:/mnt/lun0/part5

    sleep 30
  #
  arm_rel_mntpoint
  else
    armlog "ERROR_DISK from START"
  fi
elif [ "$ARMS_EVENT" = "FAILOVER" ]
then
  if [ "$ARMS_DISK" = "SUCCESS" ]
  then
    armlog "FAILOVER1"
    if [ "$ARMS_SERVER" = "HOME" ]
    then
      armlog "FAILOVER2"
    else
      armlog "ON_OTHER2"
    fi

    # stop FIP
    /usr/clusterpro/armmontr/bin/fipcmd64 -d 192.168.1.3

    /usr/sbin/gfsctl_s -r /dev/scsi/host3/bus0/target16/lun0/disc SX-6 /dev/scd/c015t016d000

    PID=`ps -e | egrep 'gfsd' | egrep '2050' | awk '{print $1}'`
    if [ "$PID" != "" ]; then
      kill -9 ${PID}
    fi

    /usr/sbin/exportfs -u SX-6:/mnt/lun0/part5

    sleep 30
  #
  arm_rel_mntpoint
  else
    armlog "ERROR_DISK from FAILOVER"
  fi
else
  armlog "NO_ARM"
fi
armlog "EXIT"
exit 0

```