

# CLUSTERPRO システム構築ガイド

---

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

---

## システム設計編(基本/共有ディスク)

第3版 2002.07.23

改版履歴

| 版 数 | 改版年月日      | 改版ページ         | 内 容                             |
|-----|------------|---------------|---------------------------------|
| 第1版 | 2001.12.14 |               | Ver1.0の第4版をベースに新規作成             |
| 第2版 | 2002.3.5   | -             | IA-64版の差分を吸収                    |
|     |            | 9,16<br>12,13 | Web版マネージャの記述を削除<br>md,LDMの注意を追記 |
| 第3版 | 2002.7.23  |               | 共有ディスクに関するエラーメッセージを削除           |
|     |            |               |                                 |

## はじめに

『CLUSTERPRO システム構築ガイド』は、これからクラスタシステムを設計・導入しようとしているシステムエンジニアや、すでに導入されているクラスタシステムの保守・運用管理を行う管理者や保守員の方を対象にしています。

CLUSTERPROは日本電気株式会社の登録商標です。

Linuxは、Linus Torvalds氏の米国およびその他の国における登録商標あるいは商標です。

Microsoft®、Windows®、およびWindows NT®は、米国Microsoft Corporation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。

Netscape および Netscape Navigatorは、米国およびその他の国におけるNetscape Communicationsの登録商標です。

その他のシステム名、社名、製品名等はそれぞれの会社の商標または登録商標です。

# CLUSTERPROドキュメント体系

CLUSTERPROのドキュメントは、CLUSTERPROをご利用になる局面や読者に応じて以下の通り分冊しています。初めてクラスタシステムを設計する場合は、システム構築ガイド【入門編】を最初に読んでください。

## ■ システム構築ガイド

### 【入門編】

(必須) 設計・構築・運用・保守

クラスタシステムをはじめて設計・構築する方を対象にした入門書です。

### 【システム設計編(基本/共有ディスク)】

(必須) 設計・構築・運用・保守

クラスタシステムを設計・構築を行う上でほとんどのシステムで必要となる事項をまとめたノウハウ集です。構築前に知っておくべき情報、構築にあたっての注意事項などを説明しています。

### 【システム設計編(応用)】

(選択) 設計・構築・運用・保守

設計編(基本)で触れなかったCLUSTERPROのより高度な機能を使用する場合に必要な事項をまとめたノウハウ集です。

### 【クラスタ生成ガイド(共有ディスク)】

(必須) 設計・構築・運用・保守

CLUSTERPROのインストール後に行う環境設定を実際の作業手順に沿って分かりやすく説明しています。

### 【運用/保守編】

(必須) 設計・構築・運用・保守

クラスタシステムの運用を行う上で必要な知識と、障害発生時の対処方法やエラー一覧をまとめたドキュメントです。

### 【GUIリファレンス】

(必須) 設計・構築・運用・保守

クラスタシステムの運用を行う上で必要なCLUSTERPROマネージャなどの操作方法をまとめたリファレンスです。

### 【コマンドリファレンス】

(選択) 設計・構築・運用・保守

CLUSTERPROのスクリプトに記述できるコマンドやサーバから実行できる運用管理コマンドについてのリファレンスです。

### 【トレッキングツール編】

(選択) 設計・構築・運用・保守

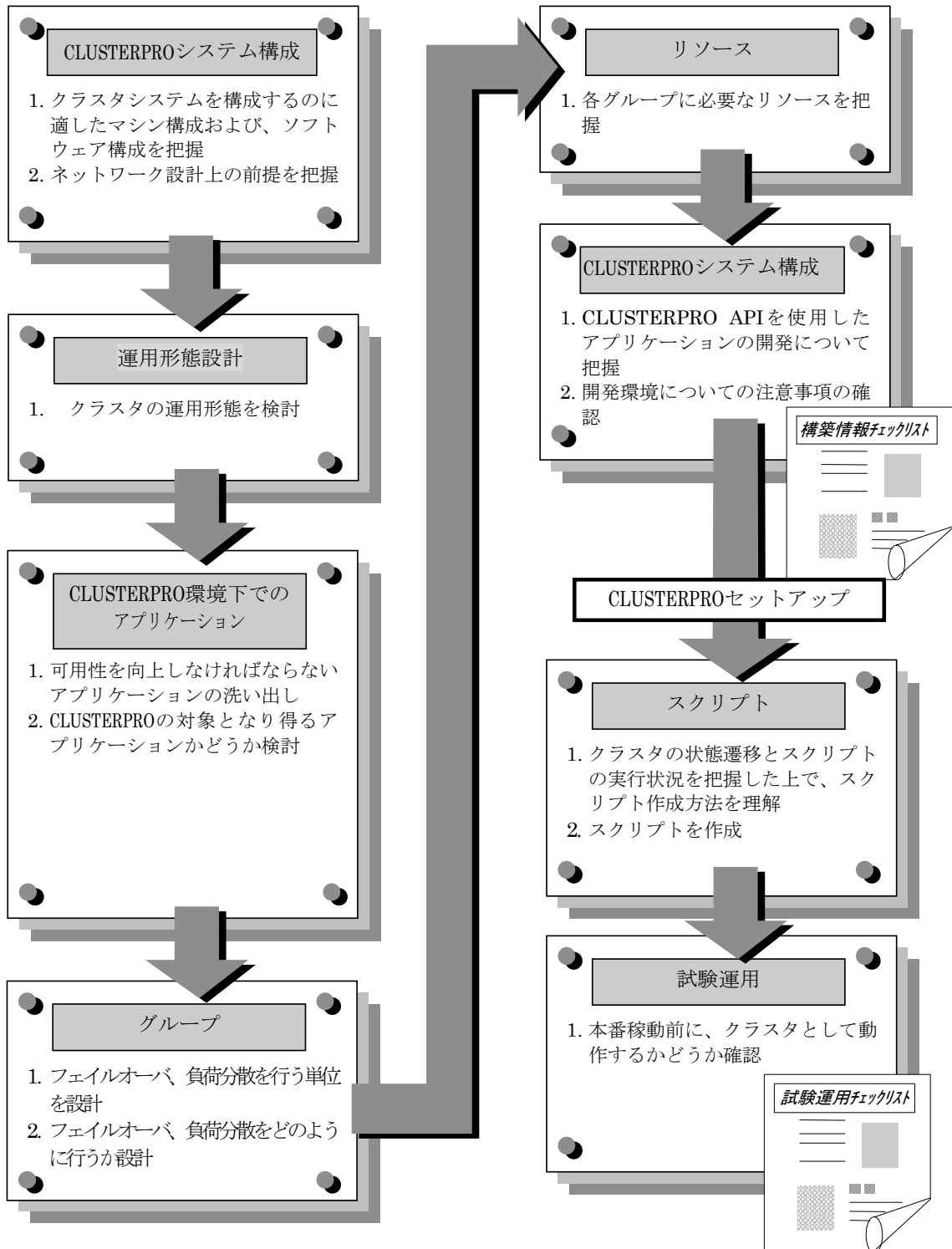
CLUSTERPROトレッキングツールの操作方法を説明したリファレンスです。分冊(GUI、システム構成(共有ディスクシステム))しています。

# 目次

|          |                                      |           |
|----------|--------------------------------------|-----------|
| <b>1</b> | <b>CLUSTERPROシステム設計手順</b> .....      | <b>7</b>  |
| <b>2</b> | <b>CLUSTERPROの構成概要</b> .....         | <b>8</b>  |
| 2.1      | スーパークラスタ .....                       | 8         |
| 2.2      | サブクラスタ .....                         | 8         |
| 2.3      | マスタクラスタ .....                        | 8         |
| <b>3</b> | <b>CLUSTERPROシステム構成</b> .....        | <b>9</b>  |
| 3.1      | マシン構成 .....                          | 9         |
| 3.1.1    | クラスタサーバ .....                        | 10        |
| 3.1.2    | 管理クライアント.....                        | 10        |
| 3.1.3    | 共有ディスク .....                         | 10        |
| 3.1.4    | インタコネクト .....                        | 10        |
| 3.1.5    | COMポート.....                          | 10        |
| 3.1.6    | NFSサーバ.....                          | 10        |
| 3.2      | ソフトウェア構成 .....                       | 11        |
| 3.2.1    | 動作環境.....                            | 12        |
| 3.2.2    | クラスタサーバ .....                        | 12        |
| 3.2.3    | NFSサーバ.....                          | 12        |
| 3.2.4    | クラスタ設定に関する注意事項.....                  | 13        |
| 3.2.5    | 管理クライアント.....                        | 14        |
| 3.3      | ネットワーク設計 .....                       | 15        |
| 3.3.1    | ネットワークの概要.....                       | 15        |
| 3.3.2    | クラスタサーバ .....                        | 16        |
| 3.3.3    | 管理クライアント.....                        | 16        |
| 3.3.4    | ルータ.....                             | 16        |
| <b>4</b> | <b>運用形態設計</b> .....                  | <b>17</b> |
| 4.1      | フェイルオーバー型クラスタ .....                  | 17        |
| 4.1.1    | 片方向スタンバイクラスタ .....                   | 18        |
| 4.1.2    | 双方向スタンバイクラスタ.....                    | 19        |
| <b>5</b> | <b>CLUSTERPRO環境下でのアプリケーション</b> ..... | <b>20</b> |
| 5.1      | 業務の洗い出し .....                        | 20        |
| 5.2      | CLUSTERPRO環境下でのアプリケーション .....        | 20        |
| 5.2.1    | サーバアプリケーション.....                     | 20        |
| 5.2.2    | サーバアプリケーションについての注意事項.....            | 20        |
| 5.2.3    | 注意事項に対する対策 .....                     | 22        |
| 5.3      | 業務形態の決定 .....                        | 22        |
| <b>6</b> | <b>フェイルオーバーグループ</b> .....            | <b>23</b> |
| 6.1      | クラスタリソース .....                       | 24        |
| 6.2      | 属性.....                              | 24        |
| 6.2.1    | フェイルオーバーグループ名 .....                  | 24        |
| 6.2.2    | 起動属性.....                            | 24        |
| 6.2.3    | フェイルオーバー属性.....                      | 24        |
| 6.2.4    | フェイルバック属性.....                       | 25        |
| 6.2.5    | フェイルオーバーポリシー .....                   | 25        |

|          |                             |           |
|----------|-----------------------------|-----------|
| 6.2.6    | アプリケーション.....               | 28        |
| 6.3      | フェイルオーバー要因.....             | 29        |
| <b>7</b> | <b>リソース.....</b>            | <b>30</b> |
| 7.1      | ディスクリソース.....               | 30        |
| 7.1.1    | 切替パーティション.....              | 30        |
| 7.1.2    | <i>CLUSTER</i> パーティション..... | 31        |
| 7.2      | フローティングIPアドレス.....          | 32        |
| 7.2.1    | アドレスの割り当て.....              | 32        |
| 7.2.2    | 環境設定.....                   | 33        |
| 7.2.3    | 経路制御.....                   | 33        |
| 7.2.4    | 使用条件.....                   | 33        |
| 7.2.5    | 接続形態.....                   | 34        |
| 7.3      | スクリプト.....                  | 38        |
| 7.4      | リソース監視.....                 | 38        |
| <b>8</b> | <b>付録.....</b>              | <b>39</b> |
| 8.1      | サーバダウン時の切替時間.....           | 39        |
| 8.2      | 待機系からのディスクチェック.....         | 40        |

# 1 CLUSTERPROシステム設計手順



## 2 CLUSTERPROの構成概要

### 2.1 スーパークラスタ

スーパークラスタは、CLUSTERPROで管理を行う最上位の単位です。スーパークラスタは、フェイルオーバ型クラスタを束ねた構成となります。

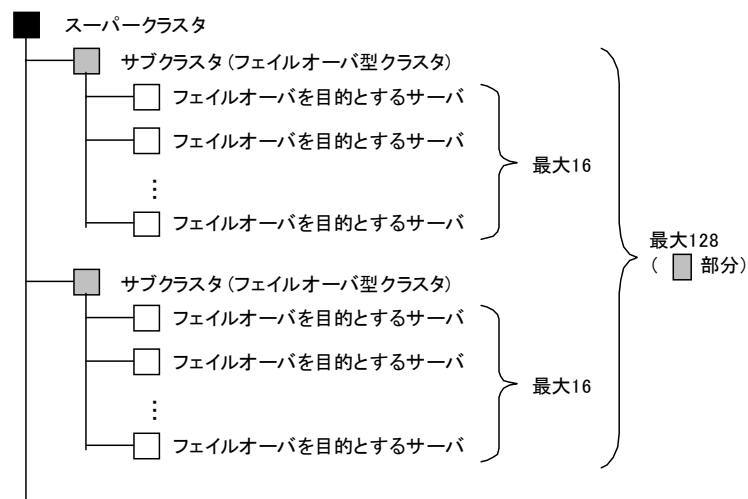
スーパークラスタの配下には、次のものを作ることができます。

\* サブクラスタ

フェイルオーバ型クラスタを構成します。

スーパークラスタは、最大8個作成することができます。

また、一つのスーパークラスタに作成できるサブクラスタとサーバは、最大128となります。以下の図を参照してください。



### 2.2 サブクラスタ

スーパークラスタの下に所属するクラスタを、サブクラスタと呼びます。

サブクラスタは、最大16台のサーバを構成することができます。ただし、共有ディスクを接続する場合には、サブクラスタ配下のサーバ数は共有ディスクの仕様に依存します。

IA-64版の場合にはサブクラスタは、最大2台のサーバを構成することができます。

また、一つのサブクラスタ内に、共有ディスクを接続するサーバと、接続しないサーバとを混在させることはできません。

サブクラスタを構成できるのは、フェイルオーバ型クラスタです。

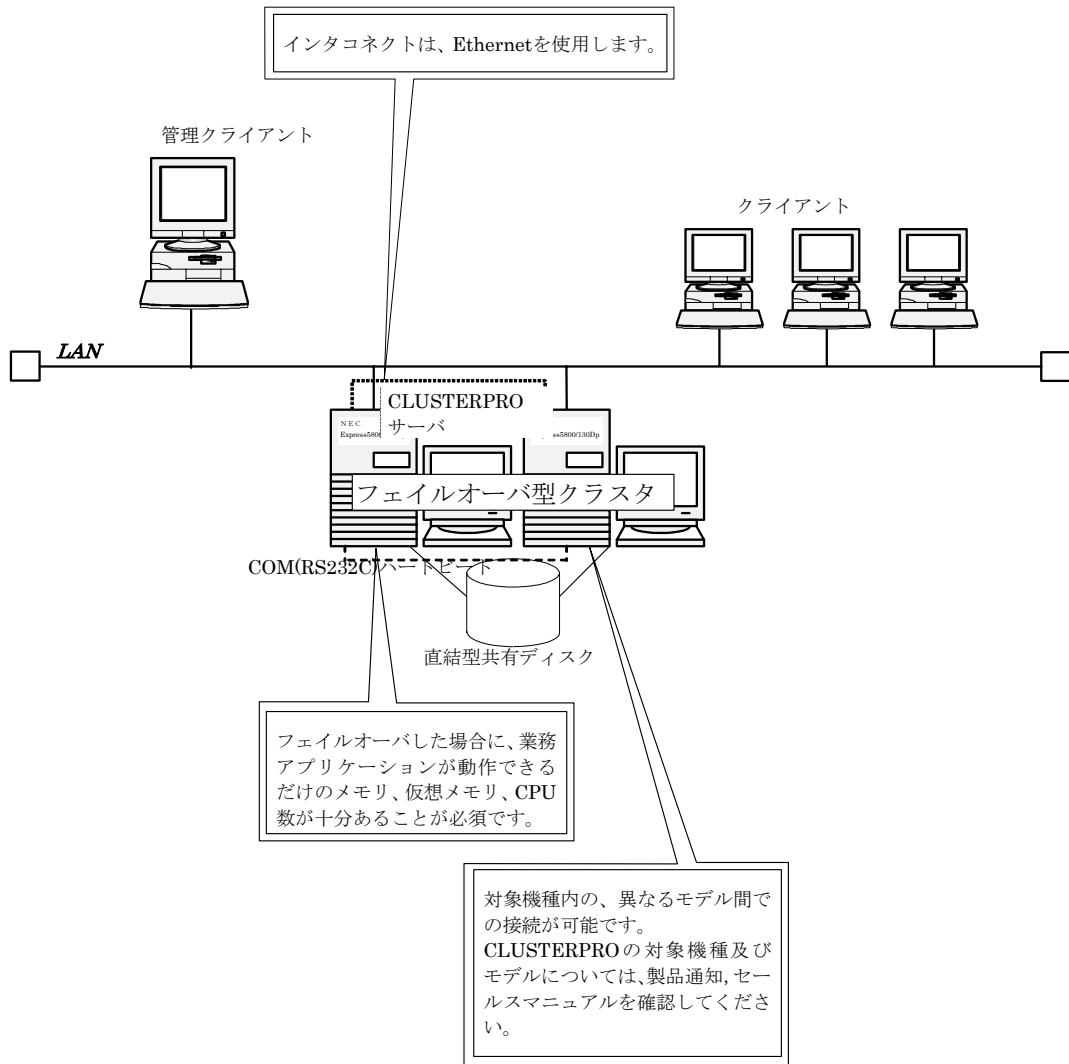
### 2.3 マスタクラスタ

マスタクラスタは、スーパークラスタの管理情報を持つサーバが所属するクラスタで、スーパークラスタ内に必ず一つなければなりません。マスタクラスタについては、「システム構築ガイド システム設計編(応用)」をご覧ください。



### 3 CLUSTERPROシステム構成

#### 3.1 マシン構成



### 3.1.1 クラスタサーバ

- \* 対象機種内の異なるモデル間での接続が可能です。CLUSTERPROの対象機種およびモデルについては、製品通知を確認してください。
- \* フェイルオーバーした場合に、業務アプリケーションが動作できるだけのメモリ、仮想メモリ、CPU数が充分あることが必須です。
- \* インタコネクには、以下の規則があります。
  - + 1 クラスタシステムに対して、最小2、最大16です。
  - + プライマリインタコネクはパブリックLANとの共用できません。

### 3.1.2 管理クライアント

- \* CLUSTERPROマネージャをインストールするマシンを管理クライアントと呼びます。
- \* 管理クライアントのOSは、Windows 95、Windows 98、Windows NT 4.0、Windows 2000、Windows XPのいずれかが必要です。

### 3.1.3 共有ディスク

- \* 必要に応じて使用するアレイディスクに対応するアレイディスク管理ツールをインストールする必要があります。RaidPlusのマニュアルで、ご確認ください。

### 3.1.4 インタコネク

- \* 100BASE-TXのEthernetを使用します。

### 3.1.5 COMポート

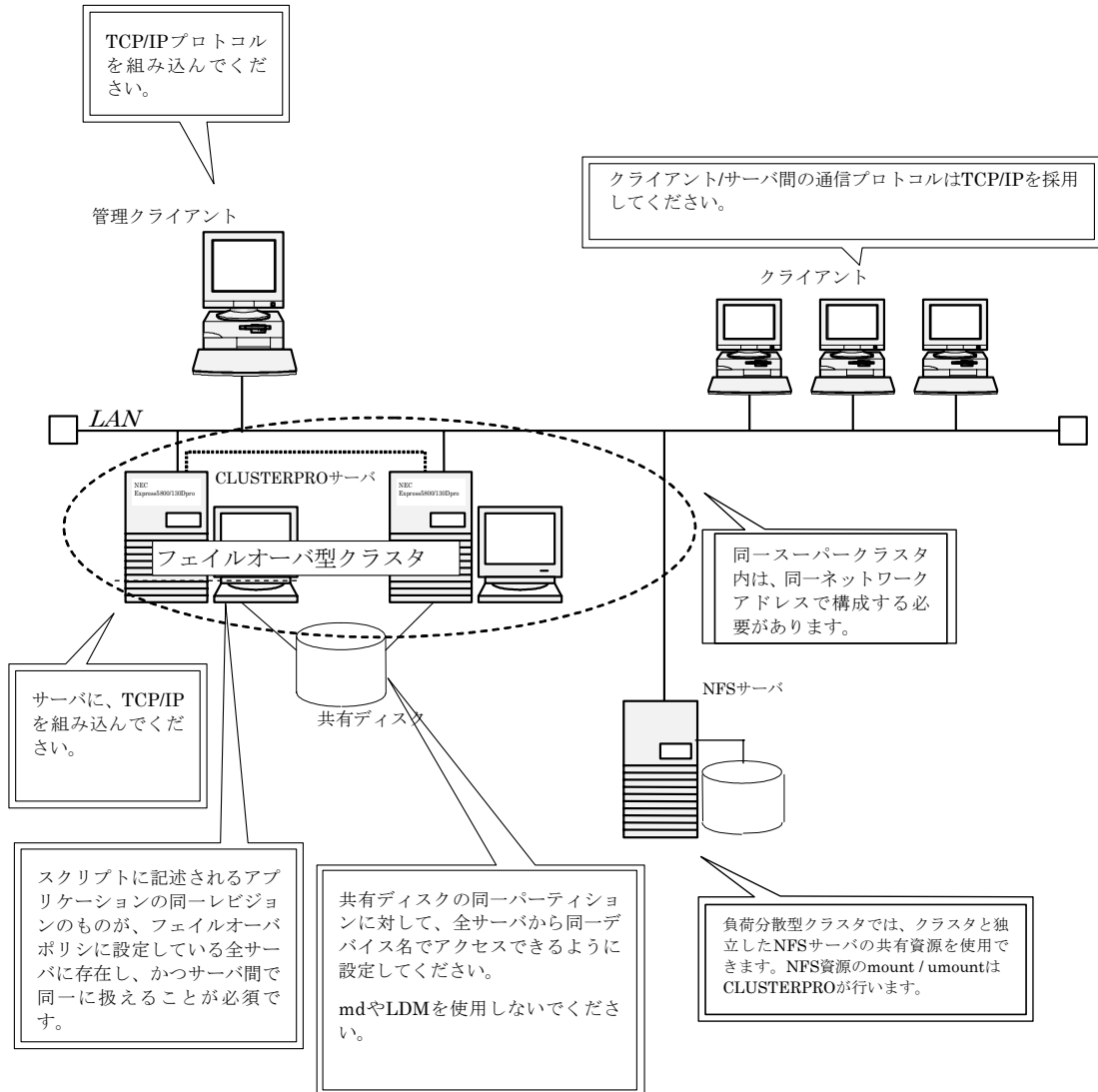
- \* IA-64版では使用しないでください。
- \* COMポート方式のハートビートを行うために、両サーバの同一デバイス名をRS-232Cのクロスケーブルで接続します。
- \* 2ノード構成のクラスタでは共有ディスクの有無に関わらず、COMポート方式のハートビートを使用してください。

### 3.1.6 NFSサーバ

- \* クラスタ構成のサーバと独立したNFSサーバを使用します。(NFSサーバをクラスタに参加させることはできません。)
- \* NFSサーバで、クラスタ構成のサーバへNFS資源のアクセスを許可する設定をしておきます。

## 3.2 ソフトウェア構成

下図は、CLUSTERPROを導入する場合のソフトウェア構成の一例です。



### 3.2.1 動作環境

- \* CLUSTERPROサーバのインストールが必要です。
- \* スーパークラスタ直下には、サブクラスタまたはサーバを、合わせて128個配置することができます。
- \* サブクラスタは、フェイルオーバー型クラスタを構成します。
- \* サブクラスタは、最大16台のサーバを構成することができます。共有ディスクを接続する場合は、共有ディスクの仕様に依存します。
- \* IA-64版では サブクラスタは、最大2台のサーバを構成することができます。
- \* 一つのサブクラスタに、共有ディスクを接続するサーバと、接続しないサーバとを混在させることはできません。

### 3.2.2 クラスタサーバ

#### 3.2.2.1 共有ディスクに関する注意事項

- \* 同一パーティションに対して、同一デバイス名でアクセスできるように設定してください。
- \* 作成できるパーティション数は、1LUNあたり最大14（基本3+拡張11）です。
- \* 共有ディスクに対してLinuxのmdやLVMによるストライプセット、ボリュームセット、ミラーリング、パリティ付ストライプセットの機能はサポートしていません。
- \* 共有ディスクには、以下の規則があります。
  - + 1クラスタシステムに対して、クラスタパーティション数は最大16までです。CLUSTERパーティションについては、「7.1.2 CLUSTERパーティション」を参照してください。
  - + 1クラスタシステムに対して、切替パーティションは最大256までです。
  - + クラスタパーティションはサーバの資源ツリーには表示しません。クラスタパーティションの接続に失敗した場合にはサーバを黄色表示し、サーバ状態に「接続に失敗した共有資源があります」を表示します。
- \* ファイルシステムのアクセス制御（mount/umount）は、CLUSTERPROが行いますので、OS側でmount/umountする設定を行わないでください。

#### 3.2.2.2 ネットワーク環境に関する注意事項

- \* TCP/IPを組み込む必要があります。
- \* IPアドレスには、以下の規則があります。
  - + 1サーバに対して最大16（フローティングIPアドレスを除く）までです。
  - + 1サーバ内に同一ネットワークアドレスに属するIPアドレスが複数存在してはいけません。  
また、以下のように包含関係にあってもいけません。  
IPアドレス：10.1.1.10、サブネットマスク：255.255.0.0  
IPアドレス：10.1.2.10、サブネットマスク：255.255.255.0

### 3.2.3 NFSサーバ

- \* クラスタ構成のサーバでは、NFS資源へのアクセス制御(mount / umount)はCLUSTERPROが行いますので、OS側でmount / umountする設定を行わないでください。

### 3.2.4 クラスタ設定に関する注意事項

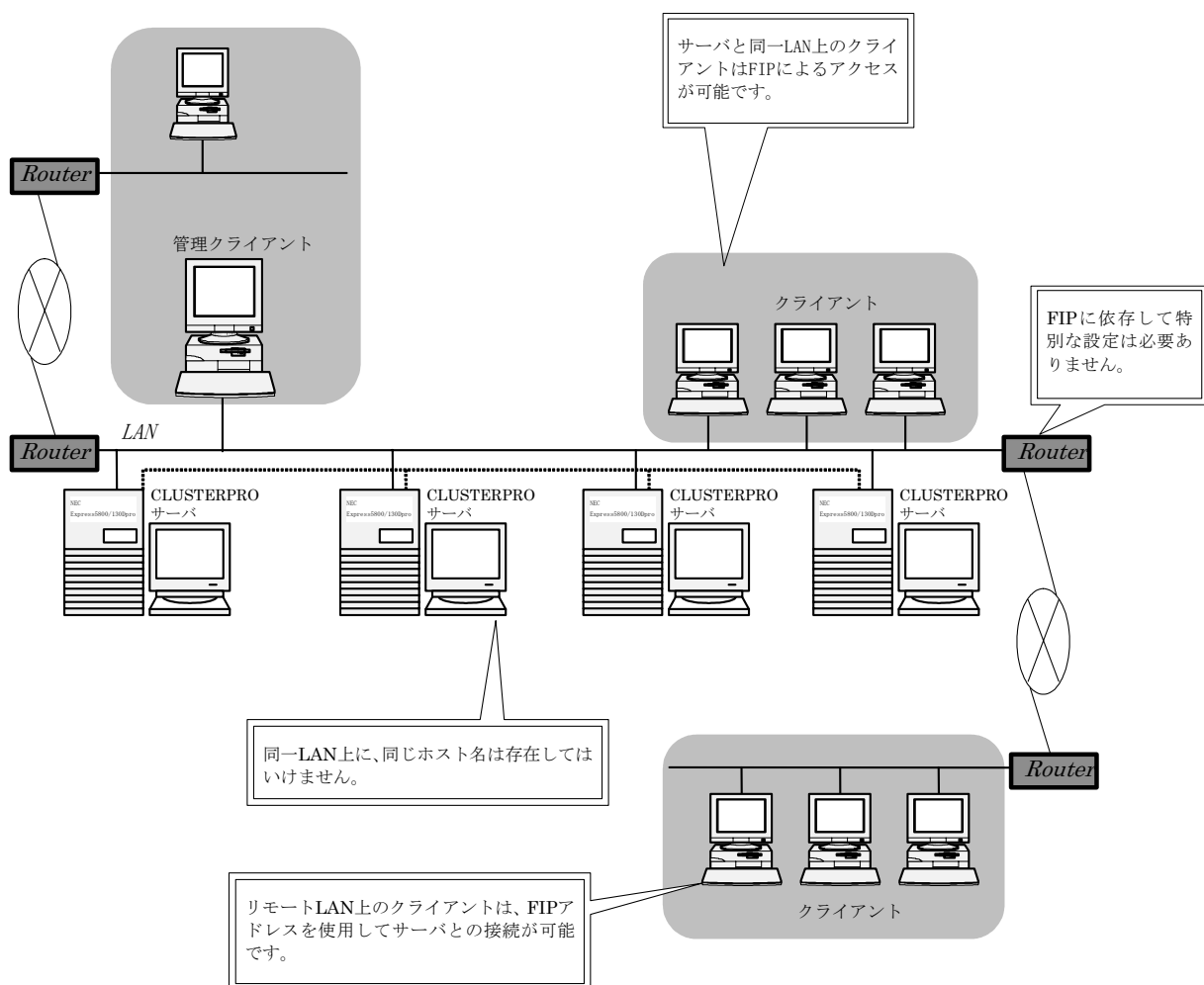
- \* スーパークラスタ名、サブクラスタ名には、以下の規則があります。
  - + 1バイトの英大文字・小文字、数字、ハイフン(-)、アンダーバー(\_)のみ使用可能です。
  - + 英大文字・小文字の区別はありません。
  - + 最大15文字(15バイト)までです。
  - + 各クラスタシステムに対して、一意な名前であればなりません。
- \* サーバ名には、以下の規則があります。
  - + OSで設定可能なコンピュータ名と同じ規則があります。
  - + 大文字・小文字は区別します。
  - + 最大15文字(15バイト)までです。
- \* グループには、以下の規則があります。
  - + フェイルオーバーグループ数は1サブクラスタに対して最大64までです。
  - + グループ名には、以下の規則があります。
    - = 1バイトの英大文字・小文字、数字、ハイフン(-)、アンダーバー(\_)のみ使用可能です。
    - = 大文字・小文字の区別はありません。
    - = 最大15文字(15バイト)までです。
    - = クラスタシステム内で一意な名前であればなりません。
    - = PRNなどのDOS入出力デバイス名は使用できません。
  - + スーパークラスタ内で、一意な名前であればなりません。
- \* クラスタパスワードについては、以下の規則があります。
  - + パスワード長は最大15バイトまでです。
- \* スクリプトに記述されるアプリケーションの同一レビジョンのものが、フェイルオーバーポリシーに設定されている全サーバに存在し、かつサーバ間で同一に扱えることが必須です。スクリプトの詳細については、「システム構築ガイド システム設計編(応用)」を、フェイルオーバーポリシーについては「6.2.5 フェイルオーバーポリシー」を参照してください。
- \* フローティングIPアドレスには、以下の規則があります。
  - + 1クラスタシステムに対して最大64までです。
  - + サーバのパブリックLANと同一ネットワークアドレス内で使用していないホストアドレスを割り当てる必要があります。
- \* リソース監視については、以下の原則があります。
  - + 1つのグループ内のリソース監視で監視できるLANの数は、最大16までです。  
(フローティングIPアドレスは、IPアドレス毎の個別の監視設定ができませんので上記の数には含めません。)
- \* COMポートによるハートビートについては、以下の規則があります。
  - + 1つのサブクラスタ内で設定できる、COMポートによるハートビートのデバイス数は、最大1です。
- \* 切替ディスクについては、以下の規則があります。
  - + 1つのサブクラスタ内で使用できるディスクリソースは、最大256までです。
  - + サブクラスタ間で同一名の別ディスクリソースを使用することは可能です。
  - + mdやLVMの併用はできません。

### 3.2.5 管理クライアント

- \* TCP/IPを組み込む必要があります。
- \* 1つのCLUSTERPROマネージャが管理できるスーパークラスタは最大8です。
- \* 1つのCLUSTERPROマネージャが管理できるサブクラスタは、全スーパークラスタを合計して最大128までです。
- \* 1つのクラスタシステムに接続できるCLUSTERPROマネージャ数は、クラスタシステム内の1サーバ当たり最大32までです。

### 3.3 ネットワーク設計

#### 3.3.1 ネットワークの概要



フローティングIP(FIP)については「7.2 フローティングIPアドレス」を参照してください。

### 3.3.2 クラスタサーバ

- \* クラスタを構成するサーバは、同一LAN上に存在し、同一ネットワークアドレスで構成していることが必須です。
- \* インタコネクタLANとパブリックLANは、異なるネットワークアドレスである必要があります。  
インタコネクタLANのIPアドレスは、プライベートIPアドレスでも可能です。

### 3.3.3 管理クライアント

- \* CLUSTERPROマネージャのインストールが必要です。
- \* クラスタサーバと同一LAN上に存在する必要はありません。
- \* 管理クライアントのサポートOSは、Windows 95、Windows 98、Windows NT 4.0、Windows 2000、Windows XPです。

### 3.3.4 ルータ

- \* FIPのために特別な設定は必要ありません。

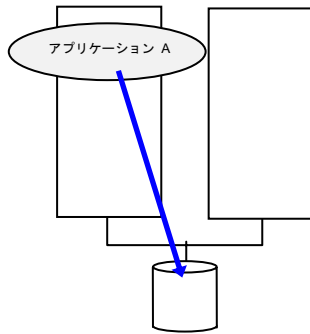


## 4 運用形態設計

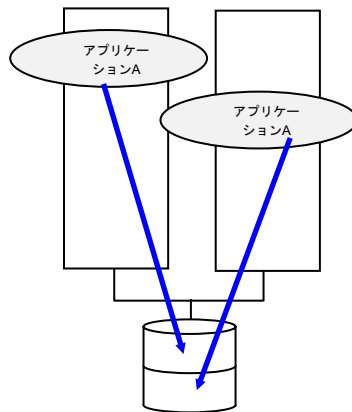
### 4.1 フェイルオーバー型クラスタ

CLUSTERPROでは、以下の運用形態をサポートしています。

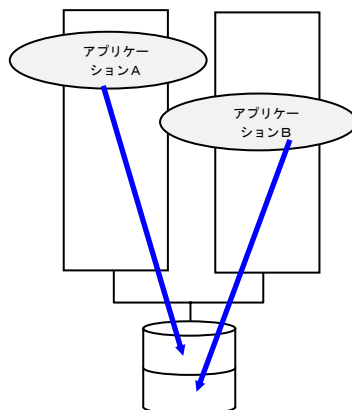
- \* 片方向スタンバイクラスタ  
クラスタシステム全体で同一の業務アプリケーションが1つしか動作しないシステム形態



- \* 同一アプリケーション双方向スタンバイクラスタ  
クラスタシステム全体で同一の業務アプリケーションが複数動作するシステム形態

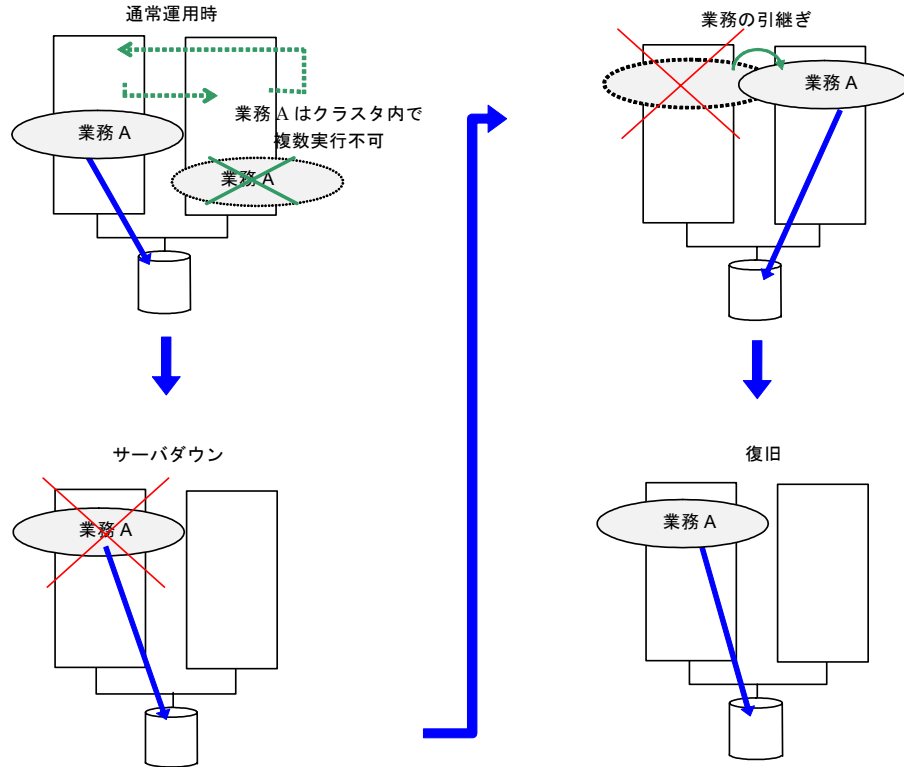


- \* 異種アプリケーション双方向スタンバイクラスタ  
複数の種類の業務アプリケーションが、それぞれ異なるサーバで稼動し、相互に待機するシステム形態



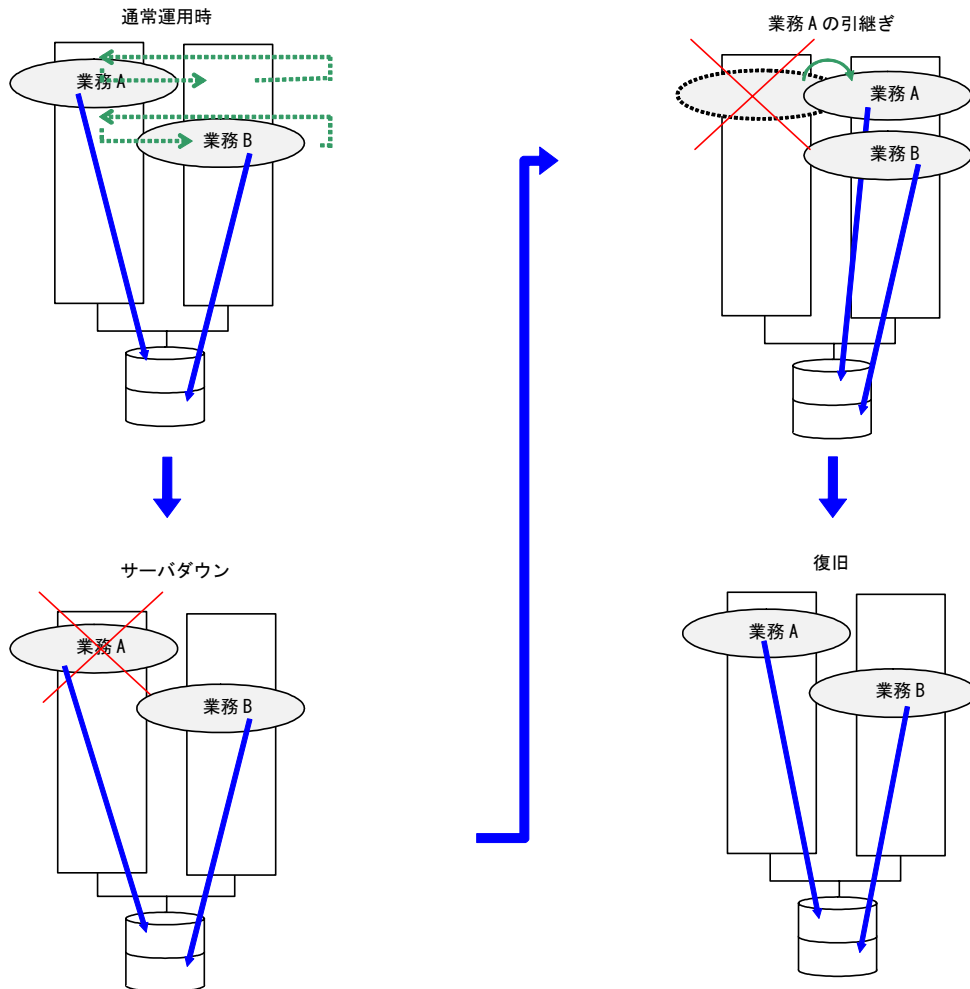
### 4.1.1 片方向スタンバイクラスタ

片方向スタンバイクラスタとは、ある業務についてフェイルオーバーグループを1グループに制限したクラスタシステムです。



## 4.1.2 双方向スタンバイクラスタ

双方向スタンバイクラスタとは、ある業務が複数のサーバ上で同時に動作することが可能なクラスタシステムです。



## 5 CLUSTERPRO環境下でのアプリケーション

ここでは、CLUSTERPRO環境下で動作できるアプリケーションについて、留意すべき事項を述べます。

### 5.1 業務の洗い出し

CLUSTERPROを導入する場合、まず可用性を向上しなければならないアプリケーション/を、洗い出す必要があります。また、洗い出したアプリケーションが、CLUSTERPROの環境下で動作するのに適しているかどうかを、見極めなければなりません。

洗い出したアプリケーションが、CLUSTERPROでのクラスタ対象として適しているかどうかは、次節からの内容を十分検討して判断してください。

### 5.2 CLUSTERPRO環境下でのアプリケーション

#### 5.2.1 サーバアプリケーション

##### (1) フェイルオーバー型クラスタ

対象アプリケーションがどのようなスタンバイ形態で実行するかで注意事項が異なります。

- \* 片方向スタンバイ[運用-待機] 注意事項: 1 2 3 4 5  
クラスタ内で、あるアプリケーションの稼動サーバが常に一台である運用形態です。
- \* 双方向スタンバイ[運用-運用] 注意事項: 1 2 3 4 5  
クラスタ内で、あるアプリケーションの稼動サーバが複数台である運用形態です。
- \* 共存動作 注意事項: 1 2 3 4 5  
クラスタシステムによるフェイルオーバーの対象とはせず、共存動作する運用形態です。

#### 5.2.2 サーバアプリケーションについての注意事項

##### (1) 障害発生後のデータ修復

障害発生時にアプリケーションが更新していたファイルは、待機系にてアプリケーションがそのファイルにアクセスするときデータとして完結していない状態にある場合があります。

非クラスタ(単体サーバ)での障害後のリブートでも同様のことが発生するため、本来アプリケーションはこの状態に備えておく必要があります。クラスタシステム上ではこれに加え人間の関与なしに(スクリプトから)復旧が行える必要があります。

共有ディスクのファイルシステムにfsckが必要な場合には、CLUSTERPROがfsckを行います。

##### (2) アプリケーションの終了

CLUSTERPROが業務グループを停止・移動(オンラインフェイルバック)する場合、その業務グループが使用していたファイルシステムをアンマウントします。このため、アプリケーションへの終了指示にて、共有ディスク上の全てのファイルに対するアクセスを停止する必要があります。

通常は終了スクリプトでアプリケーション終了指示コマンドを実行しますが、終了指示コマンドが(アプリケーションの終了と)非同期で完了してしまう場合注意が必要です。(例えばarmsleepコマンドによって一定時間待ち合わせするなど)

### (3) データ格納位置

CLUSTERPROがサーバ間で引き継ぐことのできるデータは次の通りです。

- + 共有ディスク上のデータ

アプリケーションはサーバ間で引き継ぎたいデータと引き継ぎたくないデータを分離できる必要があります。

| データの種類      | (例)            | 配置場所         |
|-------------|----------------|--------------|
| 引き継ぎたいデータ   | (ユーザデータなど)     | 共有ディスク       |
| 引き継ぎたくないデータ | (プログラム、設定情報など) | サーバのローカルディスク |

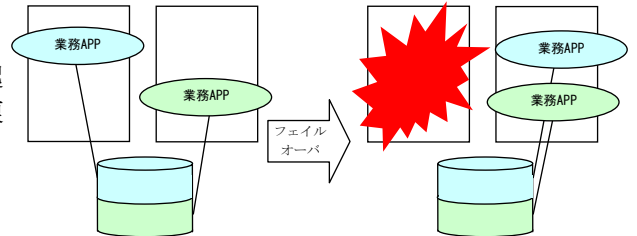
### (4) 複数業務グループ

双方向スタンバイの運用形態では、(障害による縮退時)、1つのサーバ上で同一アプリケーションによる複数業務グループが稼動することを想定しなくてはなりません。

アプリケーションは次のいずれかの方法で引き継がれた資源を引き取り、単一サーバ上で複数業務グループを実行できなければなりません。

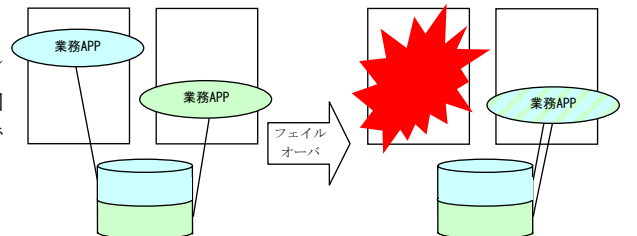
#### \* 複数インスタンス起動

新たに別インスタンス(プロセス)を起動する方法です。アプリケーションが複数動作できる必要があります。



#### \* アプリケーション再起動

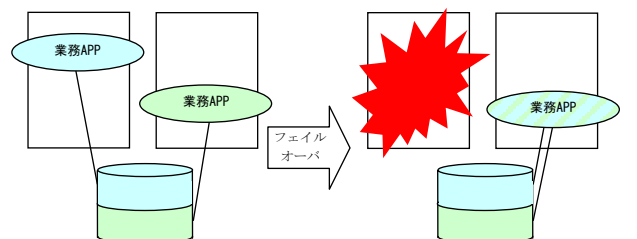
もともと動いていたアプリケーションを一旦停止し、再起動することで、追加された資源を扱えるようにする方法です。



業務APPを再起動することでデータを引き継ぐ

#### \* 動的追加

動作中のアプリケーションに対して、自動またはスクリプトからの指示により資源を追加する方法です。



実行中の業務APPに動的にデータを追加することでデータを引き継ぐ

### (5) アプリケーションとの相互干渉(相性問題)

CLUSTERPROの機能や動作に必要なOS機能との相互干渉によってアプリケーションまたはCLUSTERPROが動作できない場合があります。

\* I/Oフィルタリング

CLUSTERPROはサーバ間でのディスクパーティションの切り替えのために、I/Oフィルタリングを行い非活性状態のパーティションへのI/O要求に対しては"EIO"のエラーを返却します。

アプリケーションは非活性状態の(つまりアクセス権利のない)共有ディスクにアクセスしてはいけません。

通常クラスタスクリプトから起動されるアプリケーションは、それが起動された時点でアクセスすべき共有ディスクパーティションが既にアクセス可となっていることを想定してかまいません。

\* マルチホーム環境およびIPアドレスの移動

クラスタシステムでは、通常、一つのサーバが複数のIPアドレスを持ち、あるIPアドレス(フローティングIPアドレスなど)はサーバ間で移動します。

問題点の多くはアプリケーションがWinsockのgethostbynameで返却されるIPアドレスが一つしかないことを前提に作成されている場合に起こります。

\* アプリケーションの共有ディスクへのアクセス

共存動作アプリケーションには、業務グループの停止が通知されません。もし、業務グループの停止のタイミングでそのグループが使用している共有ディスクパーティションにアクセスしている場合、アンマウントに失敗してしまいます。

システム監視サービスを行うようなアプリケーションの中には、定期的に全てのディスクパーティションをアクセスするようなものがあります。この場合、監視対象パーティションを指定できる機能などが必要になります。

### 5.2.3 注意事項に対する対策

| 5.2.2の注意事項に対応する番号                           |  |     |
|---|--|-----|
| 問題点   | 対策   |     |
| データファイル更新中に障害が発生した場合、待機系にてアプリケーションが正常に動作しない | プログラム修正  | (1) |
| アプリケーションを停止しても一定時間の間、共有ディスクへのアクセスしつづける      | 停止スクリプト中にarmsleepコマンドを使用し待ち合わせる                | (2) |
| 一台のサーバ上で同一アプリケーションを複数起動できない                 | 双方向スタンバイ運用では、フェイルオーバー時にアプリケーションを再起動し共有データを引き継ぐ | (4) |

### 5.3 業務形態の決定

5章全体を踏まえた上で、業務形態を決定してください。

- \* どのアプリケーションをいつ起動するか
- \* 起動時やフェイルオーバー時に必要な処理は何か
- \* 共有ディスクに置くべき情報は何なのか

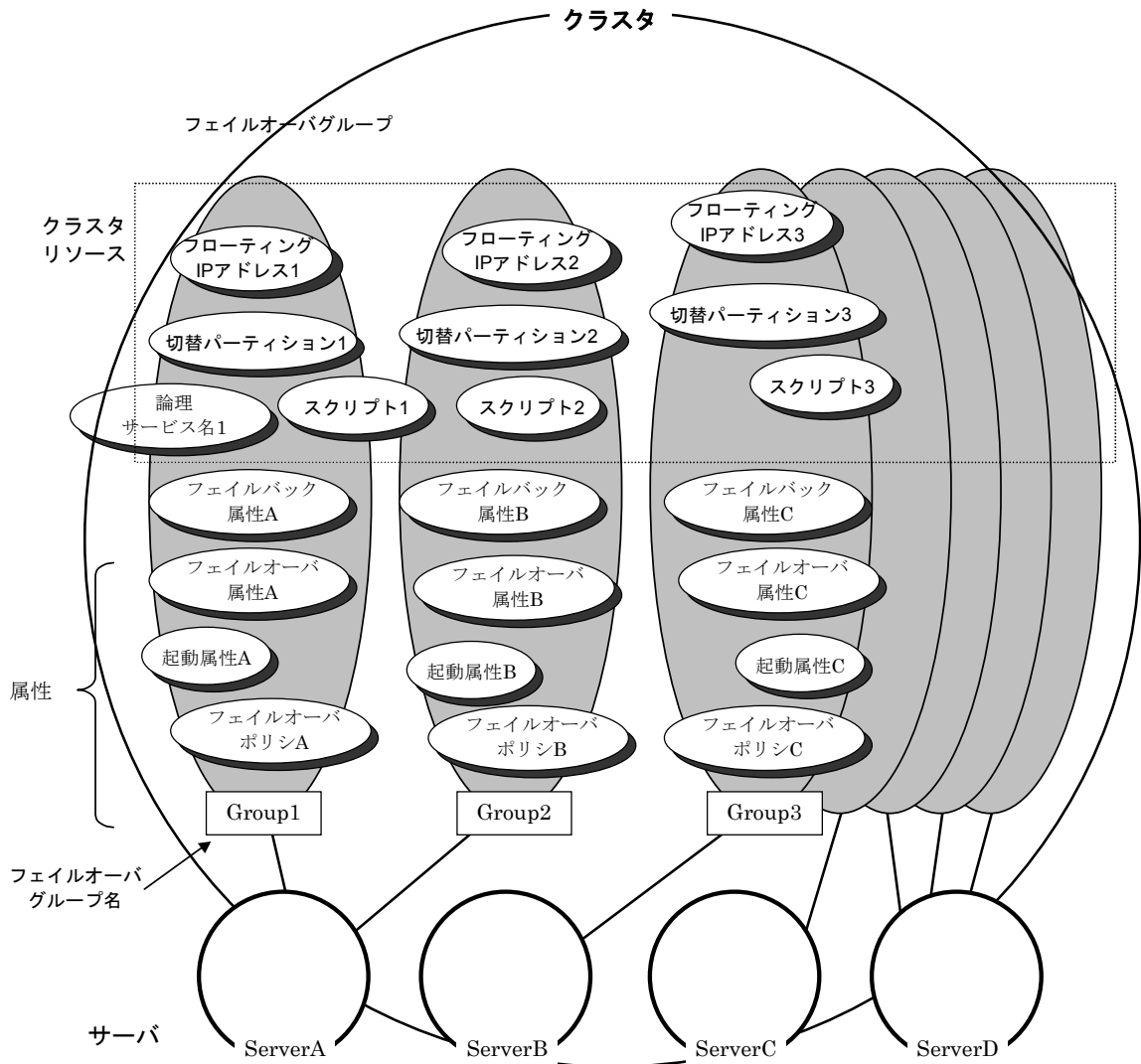
また、以下を運用の中に必ず組み込んでください。

- \* 共有ディスクの定期的なバックアップ

## 6 フェイルオーバーグループ

フェイルオーバーグループとは、クラスタシステム内のある1つの独立した業務を実行するために必要な資源の集まりのことで、フェイルオーバーを行なう単位になります。

フェイルオーバーグループは、フェイルオーバーグループ名、クラスタリソース、属性を持ちます。1サブクラスタに対して64フェイルオーバーグループまで作成することができます。



各フェイルオーバーグループのクラスタリソースは、それぞれひとまとまりのグループとして処理されます。すなわち、切替パーティション1とフローティングIPアドレス1を持つGroup1においてフェイルオーバーが発生した場合、切替パーティション1とフローティングIPアドレス1がフェイルオーバーすることになります(切替パーティション1のみが、フェイルオーバーすることはありません)。

また、切替パーティション1は、他のフェイルオーバーグループ(たとえばGroup2)に含まれることはありません。

## 6.1 クラスタリソース

フェイルオーバーグループは以下のクラスタリソースを所有することができます。  
詳細は「7 リソース」を参照してください。

- \* 切替パーティション
- \* フローティングIPアドレス
- \* リソース監視
- \* スクリプト

## 6.2 属性

フェイルオーバーグループは以下の属性を所有します。

- \* 起動属性
- \* フェイルオーバー属性
- \* フェイルバック属性
- \* フェイルオーバーポリシー

### 6.2.1 フェイルオーバーグループ名

フェイルオーバーグループの名前です。

以下の規則があります。

- \* 1バイトの英大文字/小文字、数字、ハイフン(-)、アンダーバー(\_)のみ使用可能
- \* 大文字/小文字の区別なし
- \* 最大15文字(15バイト)
- \* スーパークラスタ内で一意な名前

### 6.2.2 起動属性

クラスタ起動時にCLUSTERPROによりフェイルオーバーグループを自動的に起動するか（自動起動）、もしくはCLUSTERPROマネージャからユーザが操作して起動するか（手動起動）、の属性を指定します。

- \* 自動起動  
CLUSTERPROにより自動的に起動される。  
クラスタの起動時、フェイルオーバーグループは自動的に起動される(活性状態)。
- \* 手動起動  
CLUSTERPROからは起動されず、ユーザによるCLUSTERPROマネージャからの起動指示により起動される。  
クラスタの起動時、フェイルオーバーグループは、起動されない（非活性状態）。その後、CLUSTERPROマネージャから、ユーザが操作して起動される（活性状態）。

### 6.2.3 フェイルオーバー属性

フェイルオーバー先の決定規則を指定します。

決定規則として、常に一番優先順位の高いサーバにフェイルオーバーするか（通常）、常にグループが起動されていないサーバにフェイルオーバーするか（排他）、を選択できます。

- \* 通常  
CLUSTERPROにより自動的にフェイルオーバーされる。フェイルオーバー先の決定規則は、常に一番優先順位の高いサーバ。同一サーバで複数のグループが起動されることがある。オフラインフェイルバックはあり。
- \* 排他  
CLUSTERPROにより自動的にフェイルオーバーされる。フェイルオーバー先の決定規則は、排他のグループが起動されていないサーバのうち一番優先順位の高いサーバ。このときグループが起動されていないサーバが存在しなければ、フェイルオーバーしない。サーバで複数のグループが起動されることはない。オフラインフェイルバックはなし。



## 6.2.4 フェイルバック属性

フェイルオーバーポリシーで設定した、最高プライオリティサーバが正常状態に戻ったとき、自動的に最高プライオリティサーバへフェイルバックするかどうかを指定します。

以下のどちらかを選択します。

- \* 自動フェイルバックする
- \* 自動フェイルバックしない

既定値は、「自動フェイルバックしない」となります。

## 6.2.5 フェイルオーバーポリシー

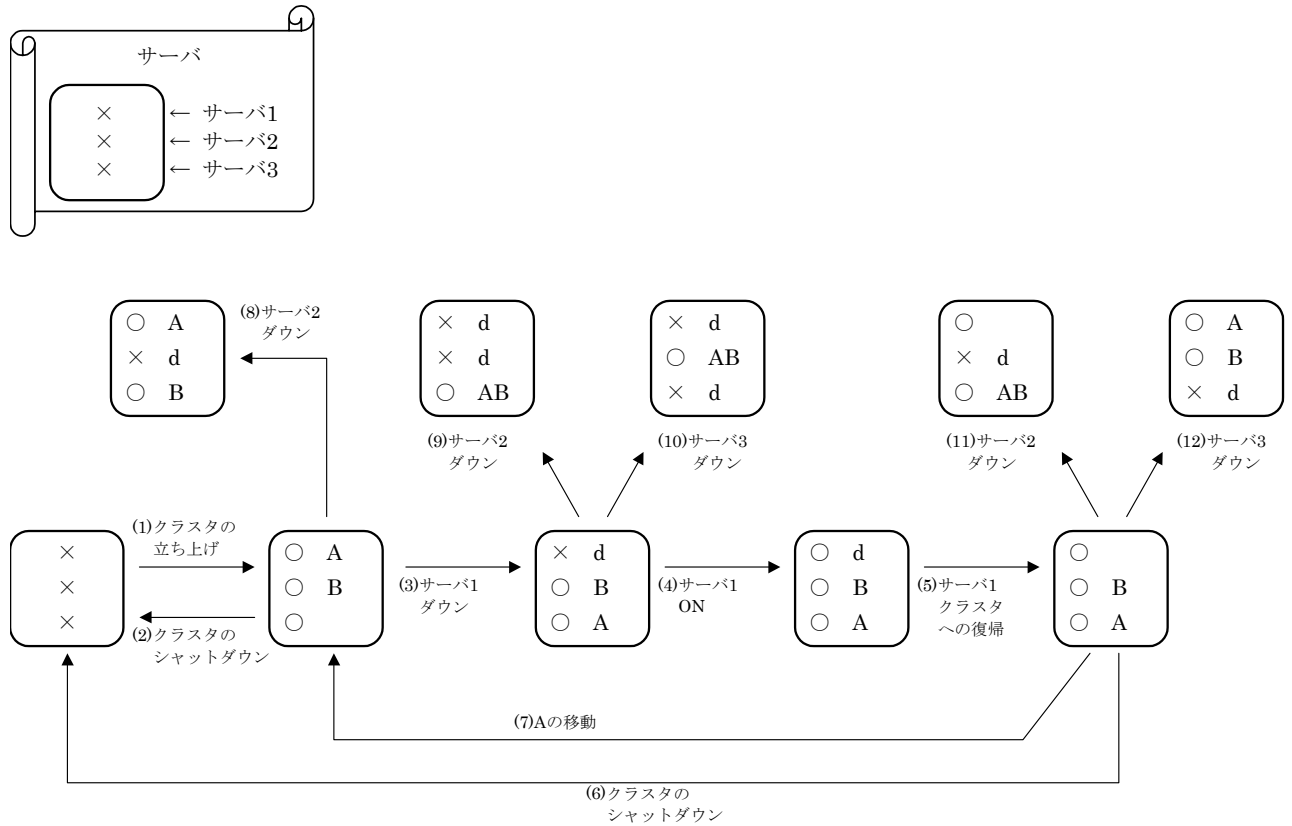
フェイルオーバー可能なサーバリストとそこでのフェイルオーバー優先順位です。フェイルオーバー発生時のフェイルオーバーポリシーによる動作の違いを説明します。

< 図中記号の説明 >

| サーバ状態 | 説明                         |
|-------|----------------------------|
| ○     | 正常状態 (クラスタとして正常に動作している)    |
| ×     | 停止状態 (クラスタシャットダウンによる停止状態)  |
| × d   | ダウン状態 (サーバダウンによる停止状態)      |
| ○ d   | ダウン後再起動状態 (クラスタから切り離されている) |

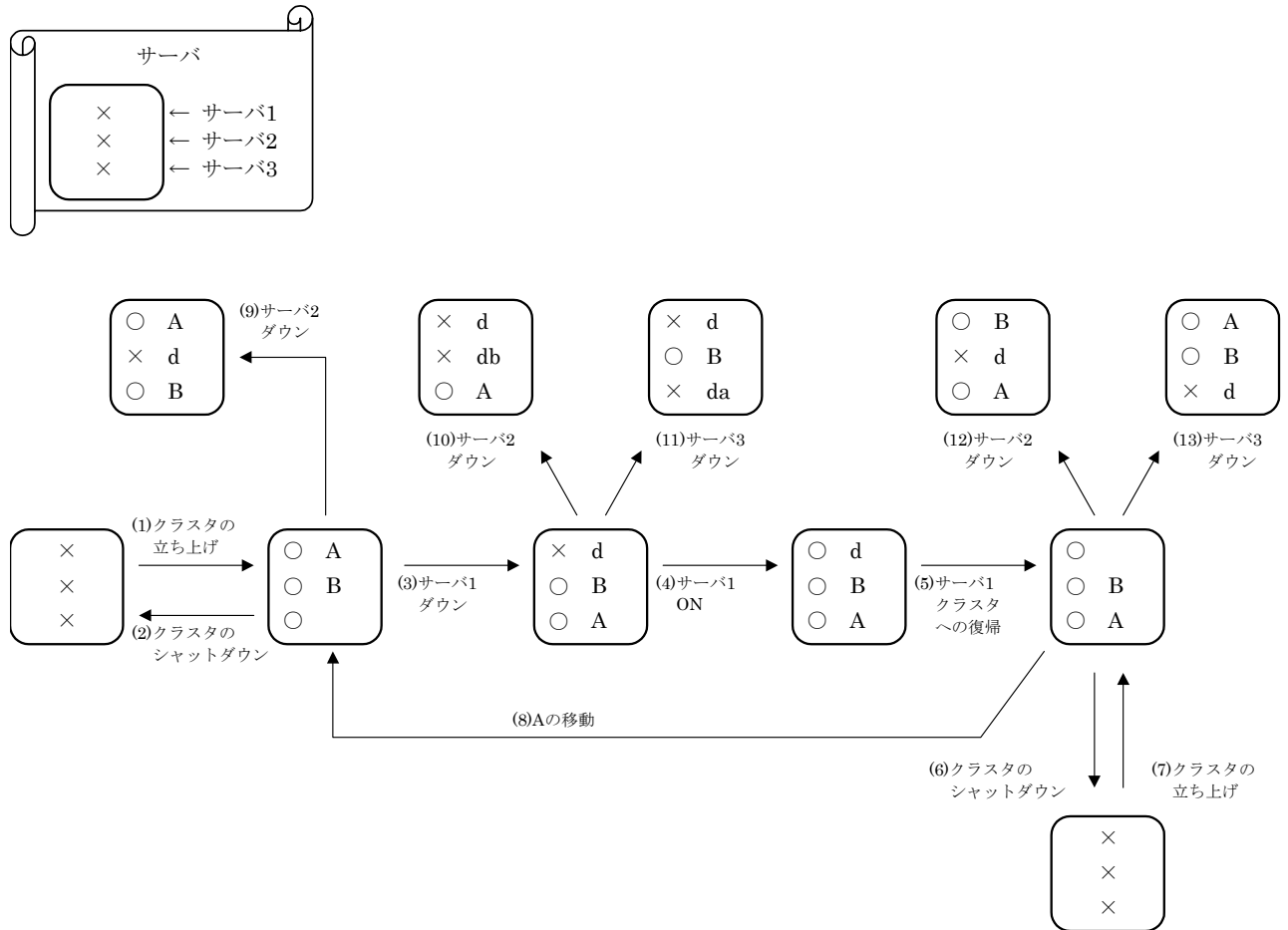
| フェイルオーバーグループ | フェイルオーバーポリシー |         |         |
|--------------|--------------|---------|---------|
|              | 優先度1サーバ      | 優先度2サーバ | 優先度3サーバ |
| A            | サーバ1         | サーバ3    | サーバ2    |
| B            | サーバ2         | サーバ3    | サーバ1    |

## フェイルオーバー属性が通常の場合



- (1) クラスタの立ち上げ
- (2) クラスタのシャットダウン
- (3) サーバ1ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (4) サーバ1の電源on
- (5) サーバ1のクラスタ復帰
- (6) クラスタのシャットダウン
- (7) フェイルオーバーグループAの移動
- (8) サーバ2ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (9) サーバ2ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (10) サーバ3ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (11) サーバ2ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (12) サーバ3ダウン : 次に優先順位の高いサーバへフェイルオーバーする

## フェイルオーバー属性が排他的の場合



\* a,bは停止状態を意味します

- (1) クラスタの立ち上げ
- (2) クラスタのシャットダウン
- (3) サーバ1ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (4) サーバ1の電源on
- (5) サーバ1のクラスタ復帰
- (6) クラスタのシャットダウン
- (7) クラスタの立ち上げ : オフラインフェイルバックしない
- (8) フェイルオーバーグループAの移動
- (9) サーバ2ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (10) サーバ2ダウン : フェイルオーバーしない
- (11) サーバ3ダウン : フェイルオーバーしない
- (12) サーバ2ダウン : 次に優先順位の高いサーバへフェイルオーバーする
- (13) サーバ3ダウン : 次に優先順位の高いサーバへフェイルオーバーする

クラスタ内にフェイルオーバ属性の異なるフェイルオーバグループが混在した場合、以下の規則で使用されます。

#### <クラスタの立ち上げ>

- \* はじめに、通常のフェイルオーバグループについて、起動するサーバを決定します。通常のフェイルオーバグループはそのときの一番優先順位の高いサーバになります。
- \* 次に、排他のフェイルオーバグループについて、起動するサーバを決定します。排他の各フェイルオーバグループについて、前回クラスタシャットダウン時に位置していたサーバより高いプライオリティのサーバを除いて、最も優先順位の高いサーバから優先順位の低いサーバを調査します。このとき、排他のフェイルオーバグループがまだ一つも活性されていないサーバを検索し、そのサーバをフェイルオーバグループの起動するサーバの候補とします。排他のフェイルオーバグループの起動先サーバ候補の検索がすべて終わった後、それぞれの候補サーバで一つのフェイルオーバグループを選択し起動します。このとき、同一サーバに対するフェイルオーバグループ同士の優先順位は、そのサーバに対するフェイルオーバ優先順位に基づきます。また優先順位が同じ場合には、フェイルオーバグループ名のアルファベット順の若い方を優先とします。これを、フェイルオーバグループがまだ一つも活性されていないサーバがなくなるまで繰り返します。

#### <サーバダウン>

- \* はじめに、通常フェイルオーバグループについてフェイルオーバするサーバを決定します。そのときの一番優先順位の高いサーバになります。このとき、仮にフェイルオーバ属性が排他のフェイルオーバグループが既に起動されていても、これは無視されます。
- \* 次に、排他のフェイルオーバグループについてフェイルオーバ先のサーバを決定します。フェイルオーバ対象の各フェイルオーバグループについて、クラスタ起動と同様の方法でフェイルオーバ先のサーバを決定しフェイルオーバします。ただし検索対象サーバは正常状態で動作中の全サーバです。

## 6.2.6 アプリケーション

クラスタに対応したアプリケーションは、“フェイルオーバ”または、“フェイルオーバグループの移動”が発生した場合に、スクリプトにより相手サーバで再起動されます。よって、同一レビジョンのアプリケーションがフェイルオーバポリシーで設定してある全サーバに存在し、かつサーバ間で同一に扱えることが必須です。また、引き継ぐべきデータを共有ディスク上に集められるような性質のものでなくてはなりません。

CLUSTERPRO環境下で動作するアプリケーションは、この他にもいくつかの前提条件をクリアしたものでなければなりません。詳細については、「5. CLUSTERPRO環境下でのアプリケーション」を参照してください。

## 6.3 フェイルオーバー要因

フェイルオーバーを引き起こす要因としては、以下のものがあります。

- \* サーバのシャットダウン
- \* 電源ダウン
- \* OSのパニック
- \* OSのストール
- \* CLUSTERPROサーバの異常
- \* スクリプトからのCLUSTERPROコマンド(`armload`)により起動したアプリケーションの障害
  - + アプリケーションの障害とは、プロセスの消失を示します
  - + `armload`には、下記オプションが指定できます
    - = 監視対象とする/しない
    - = 再起動回数の閾(しきい)値
    - = 再起動回数を0クリアするまでの時間
    - = アプリケーション単体での再起動もしくはスクリプトからの再起動
    - = しきい値を越えた場合の挙動 (サーバシャットダウンもしくはフェイルオーバー)
- \* リソース監視により監視しているリソースおよびパブリックLANで、異常を検出した場合

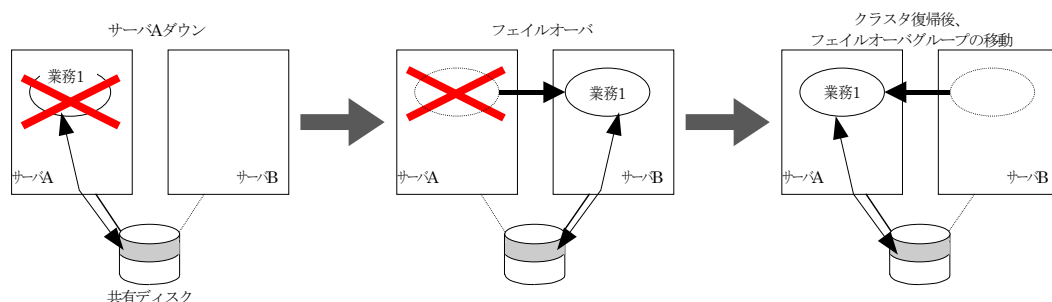
## 7 リソース

### 7.1 ディスクリソース

#### 7.1.1 切替パーティション

##### (1) 共有ディスク

- \* 切替パーティションとは、クラスタを構成する複数台のサーバに接続された共有ディスク上の、プライマリサーバから接続可能なパーティションをいいます。
- \* 切替はフェイルオーバーグループ毎に、フェイルオーバーポリシーにしたがって行われます。業務に必要なデータは、切替パーティション上に格納しておくことで、フェイルオーバー時/フェイルオーバーグループの移動時等に、自動的に引き継がれます。
- \* 切替パーティションのファイルシステムは、必ずext2にしてください。また、全サーバで、同一領域に同じデバイス名でアクセスできるようにしてください。
- \* 各切替パーティションには、ディスクリソースの設定を行う必要があります。
- \* リソース文字は、全サーバで同一でディスクに対して同一文字を割り付けてください。
- \* リソース文字の設定手順については、「システム構築ガイド クラスタ生成ガイド」を参照してください。

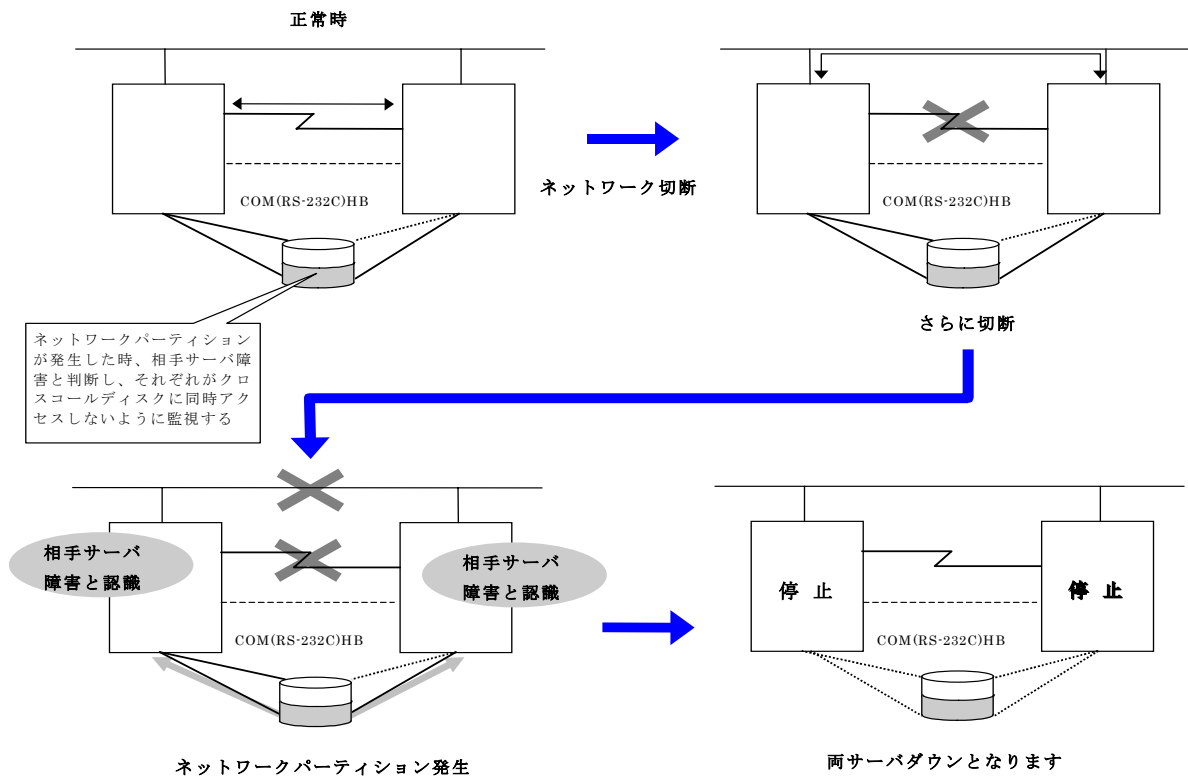


##### (2) NFSサーバ

- \* フェイルオーバーグループで、切替ディスクとして NFSサーバのリソースを使用することができます。
- \* NFSサーバが解放している資源のmount/umountはCLUSTERPROが行いますので、OSの/etc/fstabなどに記述する必要はありません。
- \* 共有ディスク上のパーティションを利用するときと同様に 使用するNFSのリソースごとにディスクリソースの設定を行う必要があります。
- \* リソース文字はスーパークラスタ直下の全サーバで同一のNFSリソースに対して同一文字を割り付けてください。
- \* リソース文字の設定手順については「システム構築ガイド クラスタ生成ガイド」を参照してください。

## 7.1.2 CLUSTERパーティション

- \* ネットワークパーティション検出にディスク方式を使用する場合、CLUSTERパーティションが必要です。
- \* 多数決方式を使用する場合には、CLUSTERパーティションは不要です。
- \* ネットワークパーティション 検出のために、CLUSTERPROサーバが使用する専用パーティションを、CLUSTERパーティションといいます。
- \* CLUSTERパーティションは、1つのLUNにつき必ず1つ確保してください。
- \* CLUSTERパーティションは、ファイルシステムを作成する必要はありません。
- \* 全サーバで同一領域が同じデバイス名でアクセスできるようにしてください。
- \* ディスクリソースの設定によりクラスタ文字を割り当ててください。CLUSTERパーティションを表わすクラスタ文字は、「###NEC\_NP」で始まる13文字以内の文字列です。クラスタ文字は、全サーバで、同一デバイスに対して同一文字列を割り付けてください。
- \* クラスタ文字の設定手順については、「システム構築ガイド クラスタ生成ガイド」を参照してください。

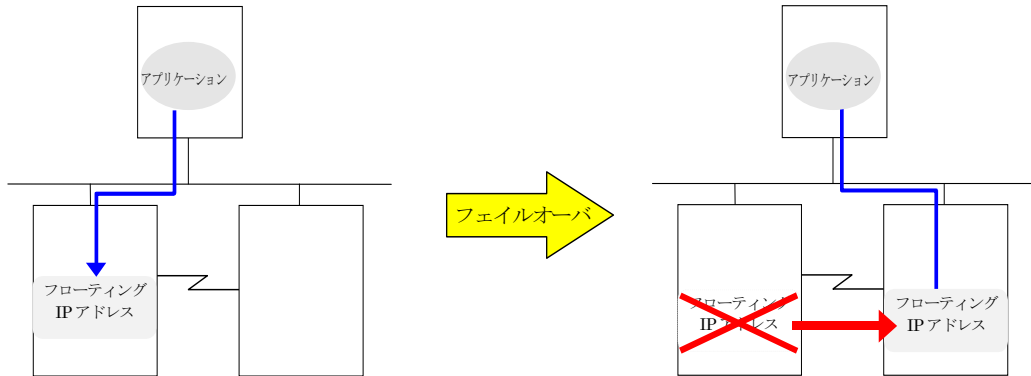


なお、COMポートによるハートビートを使用する場合には、CLUSTERパーティションによるディスク方式との併用を推奨します。

## 7.2 フローティングIPアドレス

クライアントアプリケーションは、フローティングIPアドレスを使用してクラスタサーバに接続することができます。また、サーバ間でも接続可能です。フローティングIPアドレスを使用することにより、“フェイルオーバー”または、“フェイルオーバーグループの移動”が発生しても、クライアントは、接続先サーバの切り替えを意識する必要がありません。

フローティングIPアドレスは、同一LAN上でもリモートLANからでも使用可能です。ARPプロトコルをサポートしているOSであれば使用可能です。



### フローティングIPの概要

|            |   |
|------------|---|
| 使用可能IPアドレス | パブリックLANと同一ネットワークアドレス                       |
| 切替方式       | サーバからのARPブロードキャストにより、ARPテーブル上のMACアドレスが切り替わる |
| クライアントOS   | 選ばない  |
| ルータ設定      | 不要  |
| LAN多重化     | 不可  |
| 潜在リソース     | フローティングIP設定で設定されたIPアドレス一覧                   |
| サーバ間での使用   | 可能(但しパブリックLANのみ)                            |

### 7.2.1 アドレスの割り当て

フローティングIPアドレスに割り当てるIPアドレスは、以下の条件を満たす必要があります。

\* クラスタサーバが所属するLANと同じネットワークアドレス内で かつ 使用していないホストアドレス

この条件内で必要な数(一般的にはフェイルオーバーグループ数分)のIPアドレスを確保してください。

このIPアドレスは一般のホストアドレスが変わらないため、インターネットなどのグローバルIPアドレスから割り当てることも可能です。



## 7.2.2 環境設定

フローティングIPアドレスを使用するには以下の設定が必要です。

- \* CLUSTERPROマネージャ でフェイルオーバーグループへフローティングIPアドレスの割り当て

クラスタ生成後、CLUSTERPROマネージャの「フェイルオーバーグループの追加」→「リソースの設定」→「IPアドレス」→「フローティングIP追加」により選択肢の中から使用するIPアドレスを選択してください。

フローティングIPアドレスの値を変更する場合には、「フェイルオーバーグループのプロパティ」→「リソースの設定」→「IPアドレス」により新しいFIPを追加し古いFIPを削除してください。

## 7.2.3 経路制御

サーバに使用するネットワークIPアドレスの経路制御で フローティングIPアドレスの経路制御も行われますので、フローティングIPアドレスのための特別な経路制御は不要です。

## 7.2.4 使用条件

以下のマシンからフローティングIPアドレスにアクセスできます。

- \* クラスタサーバ自身
- \* 同一クラスタ内の他のサーバ、他のクラスタシステム内のサーバ
- \* クラスタサーバと同一LAN内 及び リモートLANのクライアント

さらに以下の条件であれば上記以外のマシンからでもフローティングIPアドレスが使用できます<sup>1</sup>。

- \* 通信プロトコルがTCP/IPであること
- \* ARPプロトコルをサポートしていること

スイッチングHUBにより構成されたLANであっても、フローティングIPアドレスのメカニズムは問題なく動作します。

サーバダウン時には、接続していたTCP/IPコネクションは切断されます。

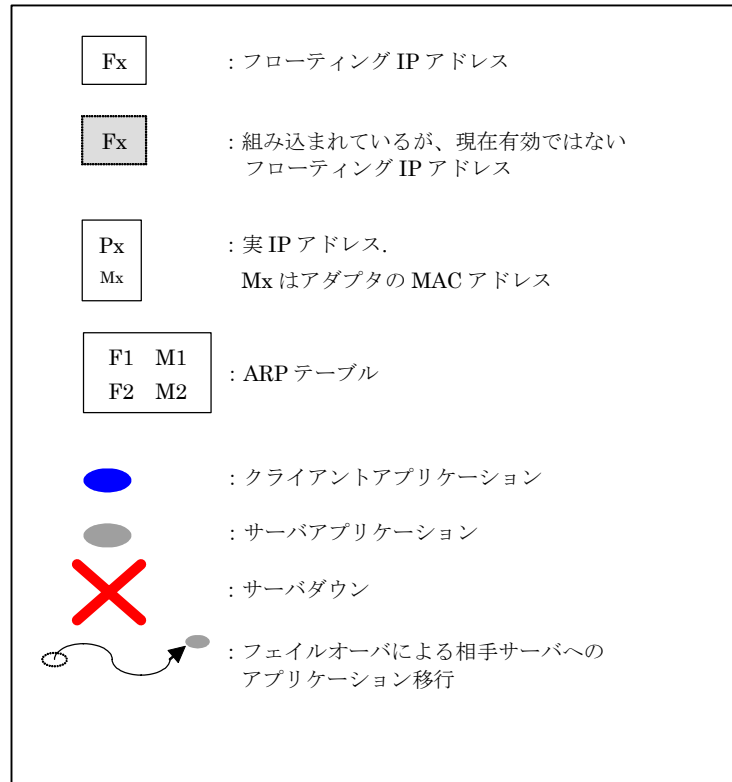
リモートLAN上のマシンからも、フローティングIPアドレスにアクセスできます。

---

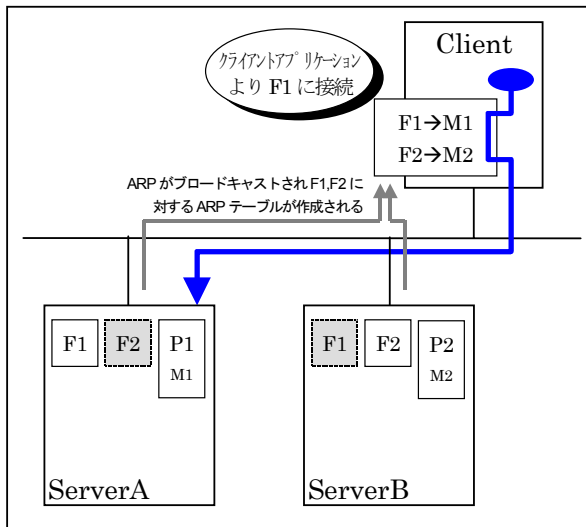
<sup>1</sup> すべてのマシン、アーキテクチャの接続を保証できません。事前に十分に評価をしてください。

## 7.2.5 接続形態

FIPアドレスによる接続形態を説明します。  
図中で使用される記号については、以下のように定義します。



## 7.2.5.1 クライアントからサーバへの接続

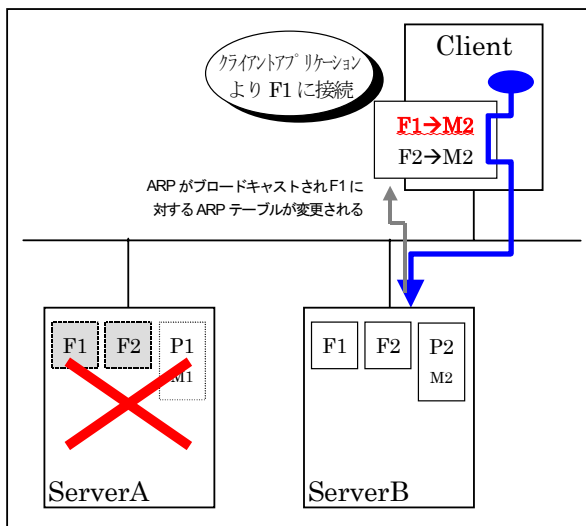


接続形態

\* クライアントからサーバのIPアドレスを指定して接続します。

接続方法

\* 接続先にFIPアドレスを指定します。



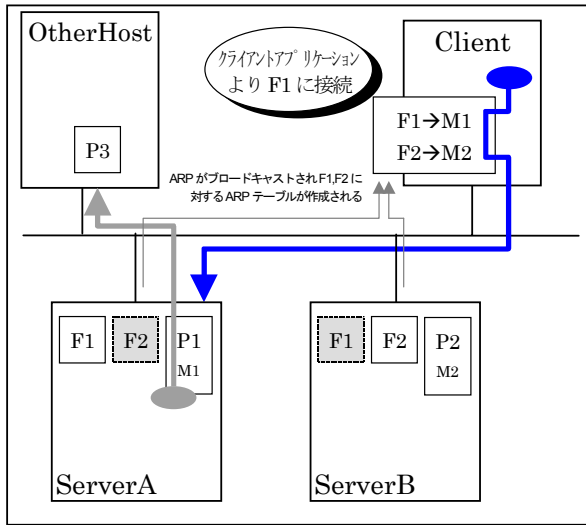
フェイルオーバー時の動作

\* フェイルオーバーが発生すると、FIPアドレスに関するクライアントのARPテーブルが変更されます。

クライアントは、そのままのFIPアドレスを用いてサーバに再接続することができます。

\* クライアントからサーバへ接続する場合に、FIPアドレスを使用すれば、フェイルオーバーの際に接続サーバが変わったことを意識する必要がありません。

### 7.2.5.2 クライアントからの要求を受けて、他ホストへ接続

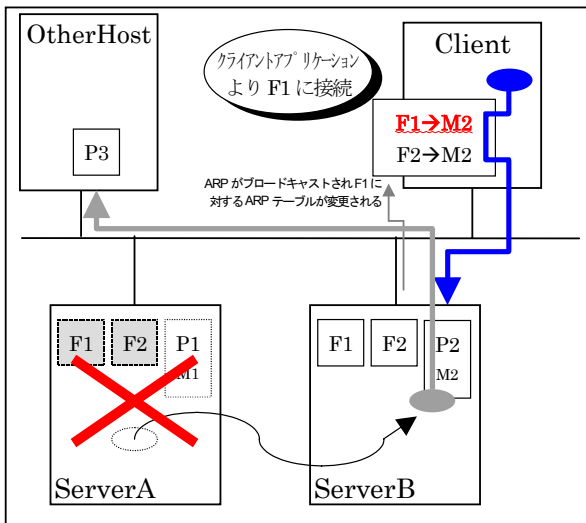


#### 接続形態

- \* クライアントアプリケーションは、サーバアプリケーションに接続します。サーバアプリケーションはクライアントアプリケーションからの要求を受けて、他ホストに接続し、その結果をクライアントアプリケーションに通知します。

#### 接続方法

- \* クライアントアプリケーションは、**FIP**アドレスでサーバアプリケーションに接続します。
- \* サーバアプリケーションが、クライアントからの要求で他ホストに接続する際は、**実IP**アドレスが用いられます。
- \* サーバアプリケーションから接続される他ホストは、どちらのサーバの実IPアドレスからの要求も受け付けるように設定しておきます。
- \* サーバアプリケーションとクライアントアプリケーションとの接続は**FIP**アドレスで、サーバアプリケーションと他ホストとの接続は**実IP**アドレスで行われます。

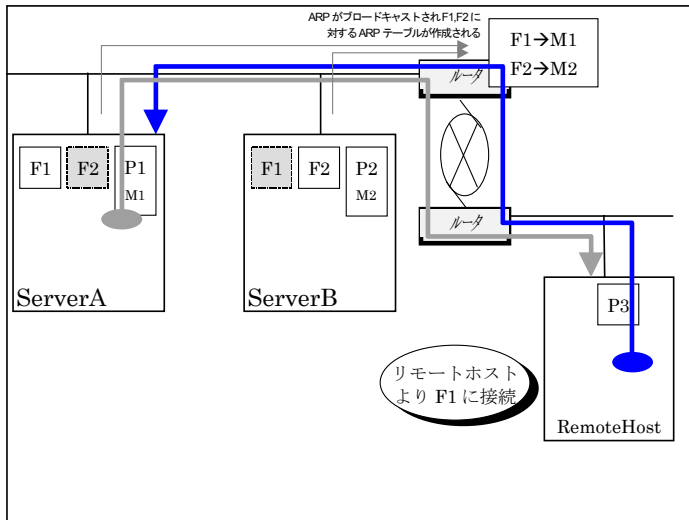


#### フェイルオーバー時の動作

- \* フェイルオーバーが発生すると、**FIP**アドレスに関するクライアントのARPテーブルが変更されます。クライアントは、そのままの**FIP**アドレスを用いてサーバに再接続することができます。フェイルオーバー先のサーバアプリケーションは、クライアントからの要求で他ホストに接続します。

- \* サーバから他ホストへの接続は、**実IP**アドレスで接続してください。サーバから他ホストへの接続に**FIP**アドレスを明示的にbindする必要はありません。

### 7.2.5.3 リモートネットワーク上の非Windowsホストとの接続

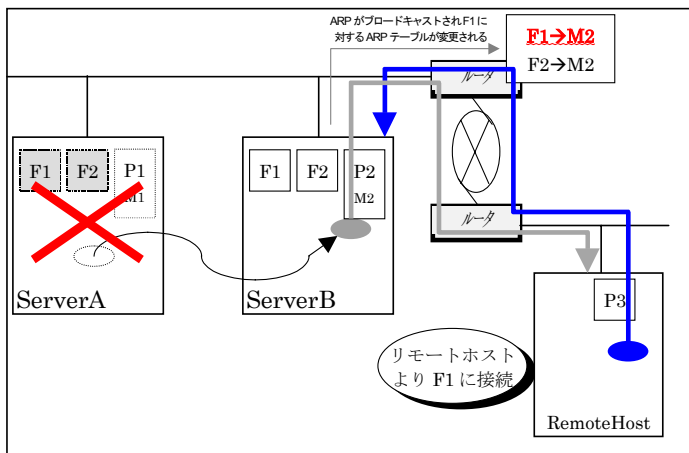


#### 接続形態

\* サーバアプリケーションから、リモートネットワーク上のホスト(以下、リモートホスト)に接続します。また、リモートホストからサーバアプリケーションに接続します。

#### 接続方法

- \* リモートホストは、どちらのサーバの実IPアドレスからの接続要求も受け付けるように設定します。
- \* サーバアプリケーションからリモートホストへの接続は、実IPアドレスでの接続となります。
- \* リモートホストからサーバアプリケーションへの接続は、FIPアドレスを指定します。



#### フェイルオーバー時の動作

\* フェイルオーバーが発生すると、クラスタサーバ側LANのルータで、FIPアドレスに関するARPテーブルが変更されます。このためリモートホストからは元と同じFIPアドレスを用いて新しいサーバに再接続することができます。また、フェイルオーバー先のサーバからも、リモートホストに再接続できます。

## 7.3 スクリプト

CLUSTERPROでは、クラスタ対象アプリケーションは、スクリプトによって制御されます。スクリプトはCLUSTERPROによって管理され、起動時、終了時、フェイルオーバ発生時フェイルオーバ、フェイルオーバグループの移動、およびクラスタ復帰の際に実行されます。

shのシェルスクリプトと同じ書式なので、それぞれのアプリケーションの事情にあわせた処理を記述できます。また、CLUSTERPROコマンドをスクリプト内に記述することで、さらに充実した機能を提供しています。

詳しくは、「システム構築ガイド システム設計編(応用)」を参照してください。

## 7.4 リソース監視

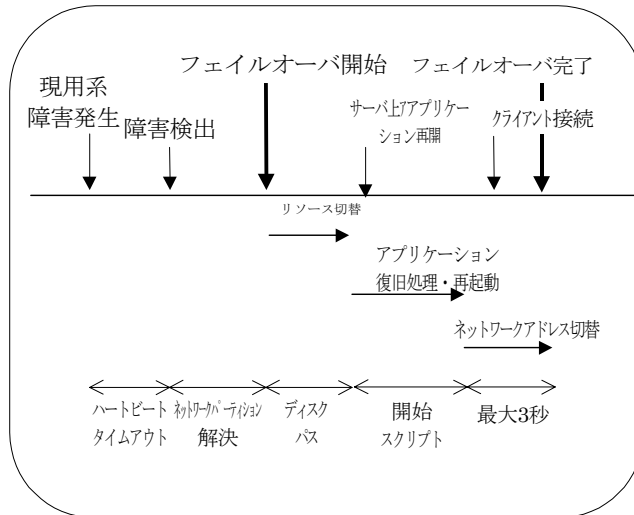
リソース監視は、CLUSTERPROコマンドのarmrspと同等の機能を有しています。

CLUSTERPROのリソースを監視し、異常を検出した場合は、フェイルオーバを発生させるか、またはグループを停止します。

詳しくは、「システム構築ガイド システム設計編(応用)」を参照してください。

## 8 付録

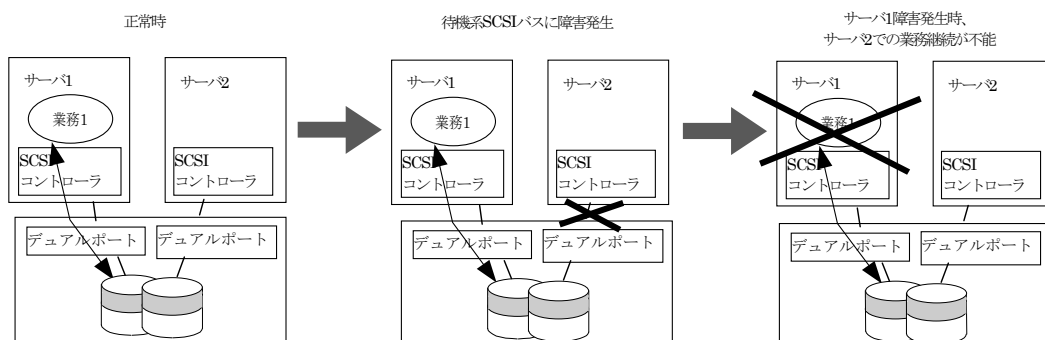
### 8.1 サーバダウン時の切替時間



- \* ハートビートタイムアウト  
プライマリ障害発生後、待機系がその障害を検出するまでの時間で、変更可能です。  
出荷時は、3秒×10回の30秒になっています。
- \* ネットワークパーティション解決  
ネットワークパーティション問題を解決するためには、下記1、2のいずれか大きい方の時間が必要です。
  - (1) 約1~2回のハートビートタイムアウト時間がかかります。(ハートビートタイムアウトを、既定値である30秒に設定している場合、30秒~60秒が目安です)
  - (2) 約1~2回のディスクIO待ち時間がかかります。(既定値である5秒に設定している場合、5秒~10秒が目安です)
- \* リソース切替 (時間は、目安です)  
フェイルオーバを行うリソースが複数ある場合は、目安時間×リソース数で計算してください。
- \* ディスクパス切替(共有ディスク)  
最短の場合約1秒で切替をします。ただしファイルシステムの容量によりmountに必要な時間が異なります。また、fsckの実行時間は除きます。
  - + ディスクパス切替(共有ディスクなし)  
約2~3秒で切り替えます。ただし、NFSサーバの性能に依存します。
  - + フローティングIPアドレス切替  
約6秒で切替えます。
  - + リソース監視  
約1秒で切り替えます。
- \* 開始スクリプト実行時間  
アプリケーションの起動時間、データベースのロールバック時間などが含まれます。ロールバック時間は、チェックポイントインターバルの調整で、ある程度予測可能です。詳しくは、各データベースのドキュメントを参照してください。

## 8.2 待機系からのディスクチェック

SCSIバスが分割されているシステムにおいては、業務が現用系側で動作している時に、待機系側のSCSIバスを確認することにより、可用性を向上させることができます。



片方向スタンバイクラスタで、共有ディスクにDS450(N8541-18、N8590-20、N8590-28、N8590-29)を使用している場合、待機系サーバ側のSCSIバスを確認することができます。待機系サーバからCLUSTERPROコマンドにより、共有ディスクのCLUSTERパーティションを監視し、待機系のSCSIバスの異常を検出させ、早期に対処することにより、可用性を向上させることができます。

- (1) 待機系サーバにおいてフェイルオーバーグループを作成します。
- (2) 起動スクリプトに次のように記述します。  
(例) `armload WatchID -R 9 -H 1 -FOV armrsp -D`
- (3) 終了スクリプトには、次のように記述します。

`armkill WatchID`

上記フェイルオーバーグループがフェイルオーバーした場合、共有ディスク装置の障害ではなく、SCSIコントローラなどの障害であると判断できます。

