

CLUSTERPRO[®] X 2.0 *for Linux*

スタートアップガイド

2009.03.31
第3版

CLUSTERPRO

改版履歴

版数	改版日付	内 容
1	2008/04/25	新規作成
2	2008/10/15	内部バージョン2.0.2-1に対応
3	2009/03/31	内部バージョン2.0.3-1に対応

免責事項

本書の内容は、予告なしに変更されることがあります。

日本電気株式会社は、本書の技術的もしくは編集上の間違い、欠落について、一切責任をおいせん。

また、お客様が期待される効果を得るために、本書に従った導入、使用および使用効果につきましては、お客様の責任とさせていただきます。

本書に記載されている内容の著作権は、日本電気株式会社に帰属します。本書の内容の一部または全部を日本電気株式会社の許諾なしに複製、改変、および翻訳することは禁止されています。

商標情報

CLUSTERPRO[®] X は日本電気株式会社の登録商標です。

FastSync[™]は日本電気株式会社の商標です。

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。

RPMの名称は、Red Hat, Inc.の商標です。

Intel、Pentium、Xeonは、Intel Corporationの登録商標または商標です。

Microsoft、Windowsは、米国Microsoft Corporationの米国およびその他の国における登録商標です。

Turbolinuxおよびターボリナックスは、ターボリナックス株式会社の登録商標です。

VERITAS、VERITAS ロゴ、およびその他のすべてのVERITAS 製品名およびスローガンは、VERITAS Software Corporation の商標または登録商標です。

本書に記載されたその他の製品名および標語は、各社の商標または登録商標です。

目次

はじめに	vii
対象読者と目的	vii
本書の構成	vii
CLUSTERPRO マニュアル体系	viii
本書の表記規則	ix
最新情報の入手先	x
セクション I CLUSTERPROの概要	13
第 1 章 クラスタシステムとは?	15
クラスタシステムの概要	16
HA (High Availability) クラスタ	16
共有ディスク型	17
データミラー型	19
障害検出のメカニズム	20
共有ディスク型の諸問題	20
ネットワークパーティション症状(Split-brain-syndrome)	21
クラスタリソースの引き継ぎ	21
データの引き継ぎ	21
アプリケーションの引き継ぎ	22
フェイルオーバー総括	23
Single Point of Failureの排除	24
共有ディスク	24
共有ディスクへのアクセスパス	25
LAN	26
可用性を支える運用	26
運用前評価	26
障害の監視	27
第 2 章 CLUSTERPRO の使用方法	29
CLUSTERPRO とは?	30
CLUSTERPRO の製品構成	30
CLUSTERPRO のソフトウェア構成	30
CLUSTERPRO の障害監視のしくみ	31
サーバ監視とは	31
業務監視とは	32
内部監視とは	32
監視できる障害と監視できない障害	33
サーバ監視で検出できる障害とできない障害	33
業務監視で検出できる障害とできない障害	33
ネットワークパーティション解決	34
フェイルオーバーのしくみ	34
フェイルオーバーリソース	35
フェイルオーバー型クラスタのシステム構成	36
共有ディスク型のハードウェア構成	39
ミラーディスク型のハードウェア構成	40
ハイブリッドディスク型のハードウェア構成	41
ハイブリッドディスク型のハードウェア構成	41
クラスタオブジェクトとは?	42
リソースとは?	43

ハートビートリソース	43
ネットワークパーティション解決リソース	43
グループリソース	43
モニタリソース	44
CLUSTERPRO を始めよう!	47
最新情報の確認	47
クラスタシステムの設計	47
クラスタシステムの構築	47
クラスタシステムの運用開始後の障害対応	47
セクション II リリースノート (CLUSTERPRO 最新情報)	49
第 3 章 CLUSTERPRO の動作環境	51
ハードウェア	52
スペック 52	
動作確認済ディスクインターフェイス	52
動作確認済ネットワークインターフェイス	53
ソフトウェア	54
CLUSTERPRO Serverの動作環境	54
動作可能なディストリビューションとkernel	54
監視オプションの動作確認済アプリケーション情報	60
必要メモリ容量とディスクサイズ	64
Builderの動作環境	65
動作確認済OS、ブラウザ	65
Java実行環境	65
必要メモリ容量/ディスク容量	65
オフライン版Builderが対応するCLUSTERPROのバージョン	66
WebManagerの動作環境	67
動作確認済OS、ブラウザ	67
Java実行環境	67
必要メモリ容量/ディスク容量	67
第 4 章 最新バージョン情報	69
CLUSTERPRO とマニュアルの対応一覧	70
機能強化	71
修正情報	72
第 5 章 注意制限事項	73
システム構成検討時	74
Builder、WebManagerの動作OSについて	74
ミラーディスクの要件について	74
共有ディスクの要件について	75
ハイブリッドディスクとして使用するディスクの要件について	77
NIC Link Up/Downモニタリソース	79
ミラーリソース、ハイブリッドディスクリソースのwrite性能について	80
OSインストール前、OSインストール時	81
/opt/nec/clusterproのファイルシステムについて	81
ミラー用のディスクについて	81
ハイブリッドディスクリソース用のディスクについて	84
依存するライブラリ	85
依存するドライバ	85
ミラードライバ	85
カーネルモードLANハートビートドライバ、キープアライブドライバ	85
RAWモニタリソース用のパーティション確保	85
SELinuxの設定	85

OSインストール後、CLUSTERPROインストール前	86
通信ポート番号	86
通信ポート番号の自動割り当て範囲の変更	88
時刻同期の設定	88
NICデバイス名について	89
共有ディスクについて	89
ミラー用のディスクについて	89
ハイブリッドディスクリソース用のディスクについて	90
OS起動時間の調整	90
ネットワークの確認	90
ipmiutil, OpenIPMIについて	90
ユーザ空間モニタリソース(監視方法softdog)について	91
ログ収集について	91
CLUSTERPROの情報作成時	92
サーバのリセット、パニック、パワーオフ	92
グループリソースの非活性異常時の最終アクション	93
execリソースから起動されるアプリケーションのスタックサイズについて	93
VxVMが使用するRAWデバイスの確認	94
ミラーディスクリソース、ハイブリッドディスクリソースの復帰回数制限について	94
ミラーディスクのファイルシステムの選択について	95
ハイブリッドディスクのファイルシステムの選択について	95
RAWモニタリソースについて	95
遅延警告割合	95
ディスクモニタリソースの監視方法TURについて	95
WebManagerの画面更新間隔について	97
LANハートビートの設定について	97
カーネルモードLANハートビートの設定について	97
COMハートビートの設定について	97
CLUSTERPRO運用後	98
hotplugサービスについて	98
udev環境でのミラードライバロード時のエラーメッセージについて	98
X-Window上のファイル操作ユーティリティについて	98
ドライバロード時のメッセージについて	98
ipmiのメッセージについて	100
複数のミラーディスクリソース、ハイブリッドディスクリソースを使用している場合の活性時のsyslogのメッ セージについて	100
回復動作中の操作制限	100
コマンド編に記載されていない実行形式ファイルやスクリプトファイルについて	101
kernelページアロケートエラーのメッセージについて	101
ログ収集時のメッセージ	101
クラスタシャットダウン・クラスタシャットダウンリポート	102
特定サーバのシャットダウン、リポート	102
サービス起動／停止用スクリプトについて	102
EXECリソースで使用するスクリプトファイルについて	104
活性時監視設定のモニタリソースについて	104
WebManagerについて	105
Builder について	106
ミラーディスク、ハイブリッドディスクリソースについて	106
第 6 章 アップデート手順	107
CLUSTERPRO X 1.0からのアップデート手順	108
CLUSTERPRO Server RPMのアップデート	108
CLUSTERPRO Server RPMのアップデート	110
付録	113
付録 A 用語集	115

付録 B	索引	119
------	----------	-----

はじめに

対象読者と目的

『CLUSTERPRO®スタートアップガイド』は、CLUSTERPRO をはじめてご使用になるユーザの皆様を対象に、CLUSTERPRO の製品概要、クラスタシステム導入のロードマップ、他マニュアルの使用方法についてのガイドラインを記載します。また、最新の動作環境情報や制限事項などについても紹介します。

本書の構成

セクション I CLUSTERPRO の概要

- 第 1 章 「クラスタシステムとは?」: クラスタシステムおよびCLUSTERPRO の概要について説明します。
- 第 2 章 「CLUSTERPRO の使用方法」: クラスタシステムの使用方法および関連情報について説明します。

セクション II リリース ノート

- 第 3 章 「CLUSTERPRO の動作環境」: 導入前に確認が必要な最新情報について説明します。
- 第 4 章 「最新バージョン情報」: CLUSTERPRO の最新バージョンについての情報を示します。
- 第 5 章 「注意制限事項」: 既知の問題と制限事項について説明します。
- 第 6 章 「アップデート手順」: 既存バージョンから最新版へのアップデート情報について説明します。

付録

- 付録 A 「用語集」
- 付録 B 「索引」

CLUSTERPRO マニュアル体系

CLUSTERPRO のマニュアルは、以下の 3 つに分類されます。各ガイドのタイトルと役割を以下に示します。

『CLUSTERPRO X スタートアップガイド』(Getting Started Guide)

すべてのユーザを対象読者とし、製品概要、動作環境、アップデート情報、既知の問題などについて記載します。

『CLUSTERPRO X インストール & 設定ガイド』(Install and Configuration Guide)

CLUSTERPRO を使用したクラスタ システムの導入を行うシステム エンジニアと、クラスタ システム導入後の保守・運用を行うシステム管理者を対象読者とし、CLUSTERPRO を使用したクラスタ システム導入から運用開始前までに必須の事項について説明します。実際にクラスタ システムを導入する際の順番に則して、CLUSTERPRO を使用したクラスタ システムの設計方法、CLUSTERPRO のインストールと設定手順、設定後の確認、運用開始前の評価方法について説明します。

『CLUSTERPRO X リファレンス ガイド』(Reference Guide)

管理者を対象とし、CLUSTERPRO の運用手順、各モジュールの機能説明、メンテナンス関連情報およびトラブルシューティング情報等を記載します。『インストール & 設定ガイド』を補完する役割を持ちます。

本書の表記規則

本書では、注意すべき事項、重要な事項および関連情報を以下のように表記します。

注：は、重要ではあるがデータ損失やシステムおよび機器の損傷には関連しない情報を表します。

重要：は、データ損失やシステムおよび機器の損傷を回避するために必要な情報を表します。

関連情報：は、参照先の情報の場所を表します。

また、本書では以下の表記法を使用します。

表記	使用方法	例
[] 角かっこ	コマンド名の前後 画面に表示される語 (ダイアログ ボックス、メニューなど) の前後	[スタート] をクリックします。 [プロパティ] ダイアログ ボックス
コマンドライン中の [] 角かっこ	かっこ内の値の指定が省略可能であることを示します。	clpstat -s[-h host_name]
#	Linux ユーザが、root でログインしていることを示すプロンプト	# clpcl -s -a
モノスペース フォント (courier)	パス名、コマンド ライン、システムからの出力 (メッセージ、プロンプトなど)、ディレクトリ、ファイル名、関数、パラメータ	/Linux/2.0/jpn/server/
モノスペース フォント太字 (courier)	ユーザが実際にコマンドラインから入力する値を示します。	以下を入力します。 # clpcl -s -a
モノスペース フォント (courier) 斜体	ユーザが有効な値に置き換えて入力する項目	rpm -i clusterprobuilder-<バージョン番号>-<リリース番号>.i686.rpm

最新情報の入手先

最新の製品情報については、以下のWebサイトを参照してください。

<http://www.nec.co.jp/clusterpro/>

セクション I CLUSTERPRO の概要

このセクションでは、CLUSTERPRO の製品概要と動作環境について説明します。

- 第 1 章 クラスタシステムとは？
- 第 2 章 CLUSTERPRO の使用方法

第 1 章 クラスタシステムとは？

本章では、クラスタシステムの概要について説明します。

本章で説明する項目は以下のとおりです。

• クラスタシステムの概要	16
• HA (High Availability) クラスタ.....	16
• 障害検出のメカニズム	20
• クラスタリソースの引き継ぎ	21
• Single Point of Failureの排除	24
• 可用性を支える運用	26

クラスタシステムの概要

現在のコンピュータ社会では、サービスを停止させることなく提供し続けることが成功への重要なカギとなります。例えば、1 台のマシンが故障や過負荷によりダウンしただけで、顧客へのサービスが全面的にストップしてしまうことがあります。そうすると、莫大な損害を引き起こすだけでなく、顧客からの信用を失いかねません。

このような事態に備えるのがクラスタシステムです。クラスタシステムを導入することにより、万一のときのシステム稼働停止時間(ダウンタイム)を最小限に食い止めたり、負荷を分散させたりすることでシステムダウンを回避することが可能になります。

クラスタとは、「群れ」「房」を意味し、その名の通り、クラスタシステムとは「複数のコンピュータを一群(または複数群)にまとめて、信頼性や処理性能の向上を狙うシステム」です。クラスタシステムには様々な種類があり、以下の 3 つに分類できます。この中で、CLUSTERPRO はハイアベイラビリティクラスタに分類されます。

◆ HA (ハイ アベイラビリティ) クラスタ

通常時は一方が現用系として業務を提供し、現用系障害発生時に待機系に業務を引き継ぐような形態のクラスタです。高可用性を目的としたクラスタで、データの引継ぎも可能です。共有ディスク型、データミラー型、遠隔クラスタがあります。

◆ 負荷分散クラスタ

クライアントからの要求を適切な負荷分散ルールに従って負荷分散ホストに要求を割り当てるクラスタです。高スケーラビリティを目的としたクラスタで、一般的にデータの引継ぎはできません。ロードバランスクラスタ、並列データベースクラスタがあります。

◆ HPC(High Performance Computing)クラスタ

全てのノードの CPU を利用し、単一の業務を実行するためのクラスタです。高性能化を目的としており、あまり汎用性はありません。

なお、HPC の 1 つであり、より広域な範囲のノードや計算機クラスタまでを束ねた、グリッドコンピューティングという技術も近年話題に上ることが多くなっています。

HA (High Availability) クラスタ

一般的にシステムの可用性を向上させるには、そのシステムを構成する部品を冗長化し、Single Point of Failure をなくすことが重要であると考えられます。Single Point of Failure とは、コンピュータの構成要素 (ハードウェアの部品) が 1 つしかないために、その個所で障害が起きると業務が止まってしまう弱点のことを指します。HA クラスタとは、サーバを複数台使用して冗長化することにより、システムの停止時間を最小限に抑え、業務の可用性 (availability) を向上させるクラスタシステムをいいます。

システムの停止が許されない基幹業務システムはもちろん、ダウンタイムがビジネスに大きな影響を与えてしまうそのほかのシステムにおいても、HA クラスタの導入が求められています。

HA クラスタは、共有ディスク型とデータミラー型に分けることができます。以下にそれぞれのタイプについて説明します。

共有ディスク型

クラスタシステムでは、サーバ間でデータを引き継がなければなりません。このデータを共有ディスク上に置き、ディスクを複数のサーバで利用する形態を共有ディスク型といいます。

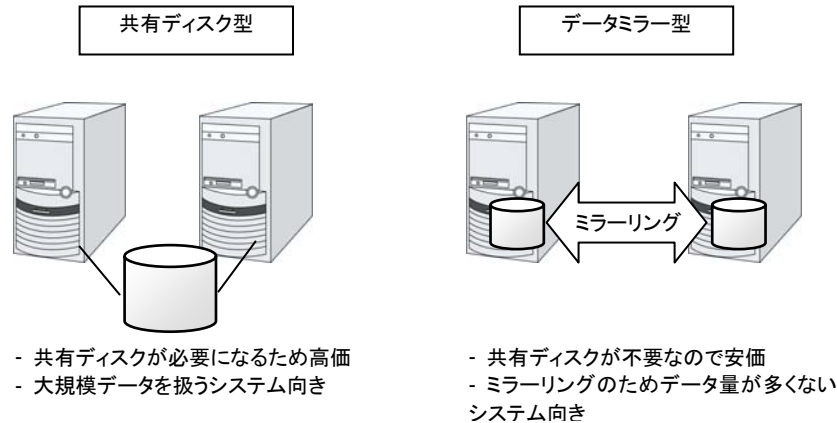


図 1-1 HAクラスタ構成図

業務アプリケーションを動かしているサーバ(現用系サーバ)で障害が発生した場合、クラスタシステムが障害を検出し、待機系サーバで業務アプリケーションを自動起動させ、業務を引き継がせます。これをフェイルオーバーといいます。クラスタシステムによって引き継がれる業務は、ディスク、IP アドレス、アプリケーションなどのリソースと呼ばれるもので構成されています。

クラスタ化されていないシステムでは、アプリケーションをほかのサーバで再起動させると、クライアントは異なる IP アドレスに再接続しなければなりません。しかし、多くのクラスタシステムでは、業務単位に仮想 IP アドレスを割り当てています。このため、クライアントは業務を行っているサーバが現用系か待機系かを意識する必要はなく、まるで同じサーバに接続しているように業務を継続できます。

データを引き継ぐためには、ファイルシステムの整合性をチェックしなければなりません。通常は、ファイルシステムの整合性をチェックするためにチェックコマンド（例えば、Linux の場合は `fsck` や `chkdsk`）を実行しますが、ファイルシステムが大きくなるほどチェックにかかる時間が長くなり、その間業務が止まってしまいます。この問題を解決するために、ジャーナリングファイルシステムなどでフェイルオーバー時間を短縮します。

業務アプリケーションは、引き継いだデータの論理チェックをする必要があります。例えば、データベースならばロールバックやロールフォワードの処理が必要になります。これらによって、クライアントは未コミットの SQL 文を再実行するだけで、業務を継続することができます。

障害からの復帰は、障害が検出されたサーバを物理的に切り離して修理後、クラスタシステムに接続すれば待機系として復帰できます。業務の継続性を重視する実際の運用の場合は、ここまでの復帰で十分な状態です。

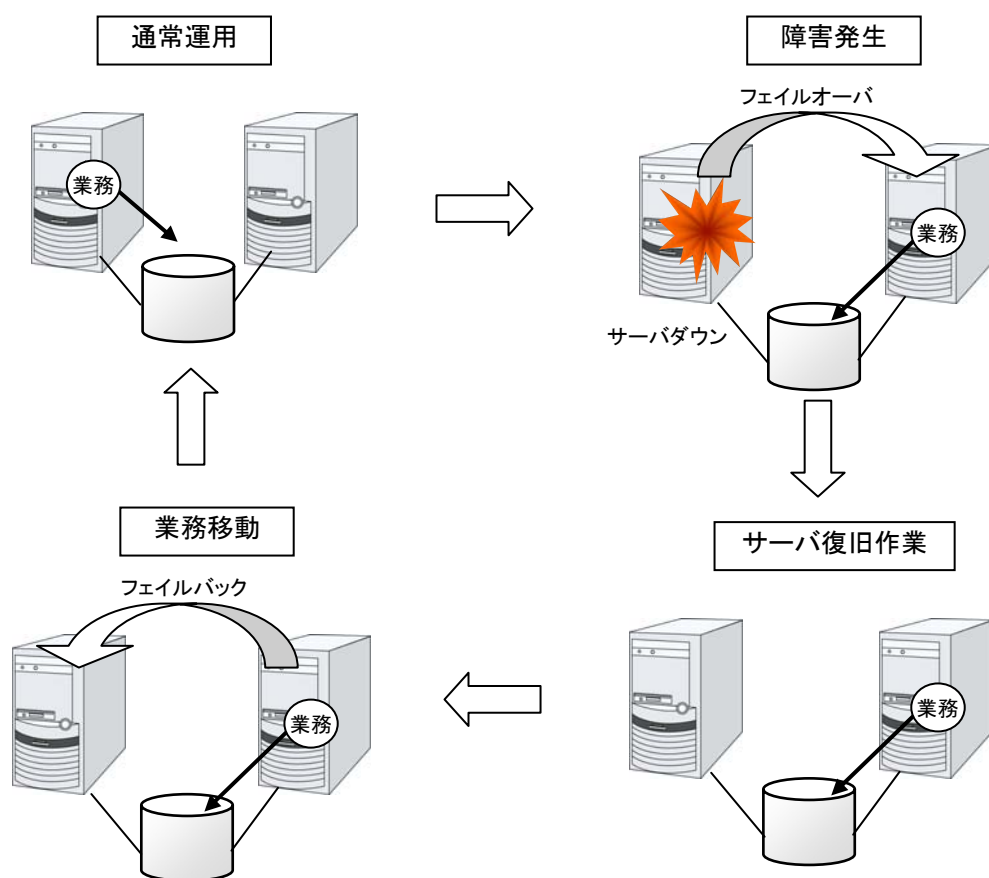


図 1-2 障害発生から復旧までの流れ

フェイルオーバー先のサーバのスペックが十分でなかったり、双方向スタンバイで過負荷になるなどの理由で元のサーバで業務を行うのが望ましい場合には、元のサーバで業務を再開するためにフェイルバックを行います。

図 1-3 のように、業務が 1 つであり、待機系では業務が動作しないスタンバイ形態を片方向スタンバイといいます。業務が 2 つ以上で、それぞれのサーバが現用系かつ待機系である形態を双方向スタンバイといいます。

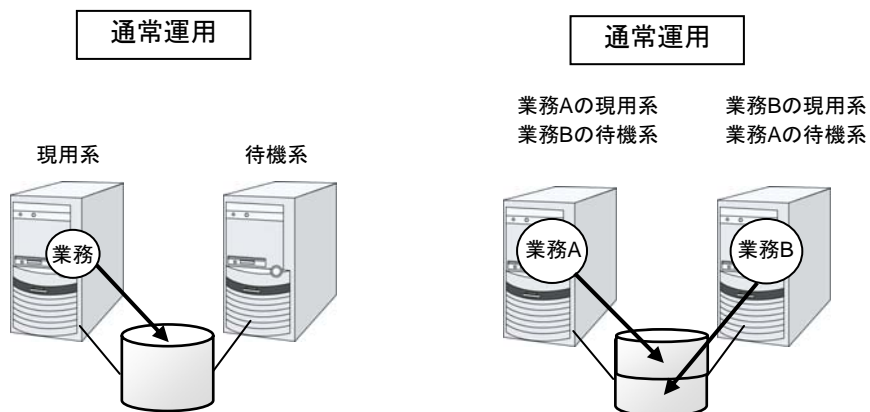


図 1-3 HA クラスタの運用形態

データミラー型

前述の共有ディスク型は大規模なシステムに適していますが、共有ディスクはおおむね高価なためシステム構築のコストが膨らんでしまいます。そこで共有ディスクを使用せず、各サーバのディスクをサーバ間でミラーリングすることにより、同じ機能をより低価格で実現したクラスタシステムをデータミラー型といいます。

しかし、サーバ間でデータをミラーリングする必要があるため、大量のデータを必要とする大規模システムには向きません。

アプリケーションからの Write 要求が発生すると、データミラーエンジンはローカルディスクにデータを書き込むと同時に、インタコネクトを通して待機系サーバにも Write 要求を振り分けます。インタコネクトとは、サーバ間をつなぐネットワークのことで、クラスタシステムではサーバの死活監視のために必要になります。データミラータイプでは死活監視に加えてデータの転送に使用することがあります。待機系のデータミラーエンジンは、受け取ったデータを待機系のローカルディスクに書き込むことで、現用系と待機系間のデータを同期します。

アプリケーションからの Read 要求に対しては、単に現用系のディスクから読み出すだけです。

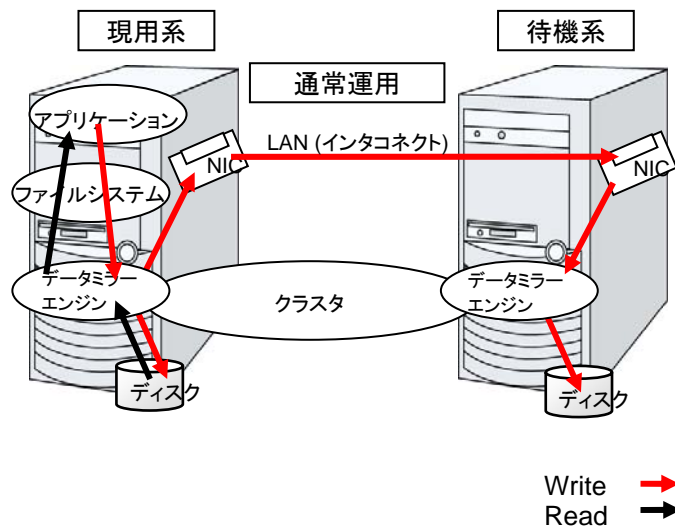


図 1-4 データミラーの仕組み

データミラーの応用例として、スナップショットバックアップの利用があります。データミラータイプのクラスタシステムは2カ所に共有のデータを持っているため、待機系のサーバをクラスタから切り離すだけで、バックアップ時間をかけることなくスナップショットバックアップとしてディスクを保存する運用が可能です。

フェイルオーバーの仕組みと問題点

ここまで、一口にクラスタシステムといってもフェイルオーバークラスタ、負荷分散クラスタ、HPC(High Performance Computing)クラスタなど、さまざまなクラスタシステムがあることを説明しました。そして、フェイルオーバークラスタは HA(High Availability)クラスタと呼ばれ、サーバそのものを多重化することで、障害発生時に実行していた業務をほかのサーバで引き継ぐことにより、業務の可用性(Availability)を向上することを目的としたクラスタシステムであることを見てきました。次に、クラスタの実装と問題点について説明します。

障害検出のメカニズム

クラスタソフトウェアは、業務継続に問題をきたす障害を検出すると業務の引き継ぎ(フェイルオーバー)を実行します。フェイルオーバー処理の具体的な内容に入る前に、簡単にクラスタソフトウェアがどのように障害を検出するか見ておきましょう。

ハートビートとサーバの障害検出

クラスタシステムにおいて、検出すべき最も基本的な障害はクラスタを構成するサーバ全てが停止してしまうものです。サーバの障害には、電源異常やメモリエラーなどのハードウェア障害や OS のパニックなどが含まれます。このような障害を検出するために、サーバの死活監視としてハートビートが使用されます。

ハートビートは、ping の応答を確認するような死活監視だけでもよいのですが、クラスタソフトウェアによっては、自サーバの状態情報などを相乗りさせて送るものもあります。クラスタソフトウェアはハートビートの送受信を行い、ハートビートの応答がない場合はそのサーバの障害とみなしてフェイルオーバー処理を開始します。ただし、サーバの高負荷などによりハートビートの送受信が遅延することも考慮し、サーバ障害と判断するまである程度の猶予時間が必要です。このため、実際に障害が発生した時間とクラスタソフトウェアが障害を検知する時間とにはタイムラグが生じます。

リソースの障害検出

業務の停止要因はクラスタを構成するサーバ全ての停止だけではありません。例えば、業務アプリケーションが使用するディスク装置や NIC の障害、もしくは業務アプリケーションそのものの障害などによっても業務は停止してしまいます。可用性を向上するためには、このようなリソースの障害も検出してフェイルオーバーを実行しなければなりません。

リソース異常を検出する手法として、監視対象リソースが物理的なデバイスの場合は、実際にアクセスしてみるという方法が取られます。アプリケーションの監視では、アプリケーションプロセスそのものの死活監視のほか、業務に影響のない範囲でサービスポートを試してみるような手段も考えられます。

共有ディスク型の諸問題

共有ディスク型のフェイルオーバークラスタでは、複数のサーバでディスク装置を物理的に共有します。一般的に、ファイルシステムはサーバ内にデータのキャッシュを保持することで、ディスク装置の物理的な I/O 性能の限界を超えるファイル I/O 性能を引き出しています。

あるファイルシステムを複数のサーバから同時にマウントしてアクセスするとどうなるでしょうか？

通常のファイルシステムは、自分以外のサーバがディスク上のデータを更新するとは考えていないので、キャッシュとディスク上のデータとに矛盾を抱えることとなり、最終的にはデータを破壊します。フェイルオーバークラスタシステムでは、次のネットワークパーティション症状などによる複数サーバからのファイルシステムの同時マウントを防ぐために、ディスク装置の排他制御を行っています。

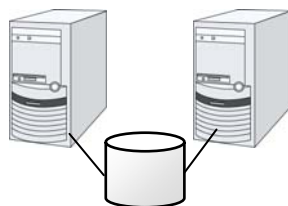


図 1-5 共有ディスクタイプのクラスタ構成

ネットワークパーティション症状(Split-brain-syndrome)

サーバ間をつなぐすべてのインタコネクトが切断されると、ハートビートによる死活監視で互いに相手サーバのダウンを検出し、フェイルオーバー処理を実行してしまいます。結果として、複数のサーバでファイルシステムを同時にマウントしてしまい、データ破壊を引き起こします。フェイルオーバークラスタシステムでは異常が発生したときに適切に動作しなければならないことが理解できると思います。

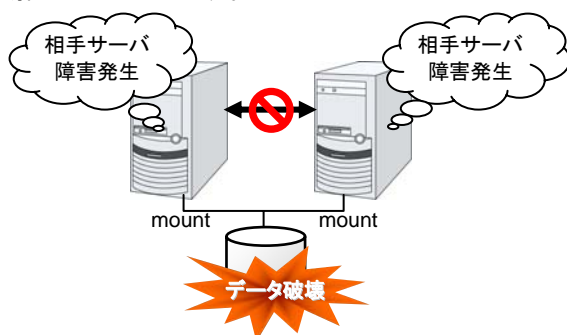


図 1-6 ネットワークパーティション症状

このような問題を「ネットワークパーティション症状」またはスプリットブレイン シンドローム (Split-brain-syndrome) と呼びます。フェイルオーバークラスタでは、すべてのインタコネクトが切断されたときに、確実に共有ディスク装置の排他制御を実現するためのさまざまな対応策が考えられています。

クラスタリソースの引き継ぎ

クラスタが管理するリソースにはディスク、IP アドレス、アプリケーションなどがあります。これらのクラスタリソースを引き継ぐための、フェイルオーバークラスタシステムの機能について説明します。

データの引き継ぎ

クラスタシステムでは、サーバ間で引き継ぐデータは共有ディスク装置上のパーティションに格納します。すなわち、データを引き継ぐとは、アプリケーションが使用するファイルが格納されているファイルシステムを健全なサーバ上でマウントしなおすことにほかなりません。共有ディスク装置は引き継ぐ先のサーバと物理的に接続されているので、クラスタソフトウェアが行うべきことはファイルシステムのマウントだけです。

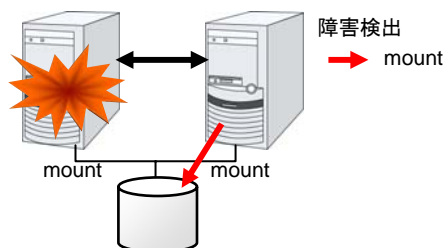


図 1-7 データの引き継ぎ

単純な話のようですが、クラスタシステムを設計・構築するうえで注意しなければならない点があります。

1 つは、ファイルシステムの復旧時間の問題です。引き継ごうとしているファイルシステムは、障害が発生する直前までほかのサーバで使用され、もしかしたらまさに更新中であつたかもしれません。このため、引き継ぐファイルシステムは通常ダーティであり、ファイルシステムの整合性チェックが必要な状態となっています。ファイルシステムのサイズが大きくなると、整合性チェックに必要な時間は莫大になり、場合によっては数時間もの時間がかかってしまいます。それがそのままフェイルオーバー時間(業務の引き継ぎ時間)に追加されてしまい、システムの可用性を低下させる要因になります。

もう 1 つは、書き込み保証の問題です。アプリケーションが大切なデータをファイルに書き込んだ場合、同期書き込みなどを利用してディスクへの書き込みを保証しようとします。ここでアプリケーションが書き込んだと思い込んだデータは、フェイルオーバー後にも引き継がれていることが期待されます。例えばメールサーバは、受信したメールをスプールに確実に書き込んだ時点で、クライアントまたはほかのメールサーバに受信完了を応答します。これによってサーバ障害発生後も、スプールされているメールをサーバの再起動後に再配信することができます。クラスタシステムでも同様に、一方のサーバがスプールへ書き込んだメールはフェイルオーバー後にもう一方のサーバが読み込めることを保証しなければなりません。

アプリケーションの引き継ぎ

クラスタソフトウェアが業務引き継ぎの最後に行う仕事は、アプリケーションの引き継ぎです。フォールトトレラントコンピュータ(FTC)とは異なり、一般的なフェイルオーバークラスタでは、アプリケーション実行中のメモリ内容を含むプロセス状態などを引き継ぎません。すなわち、障害が発生していたサーバで実行していたアプリケーションを健全なサーバで再実行することでアプリケーションの引き継ぎを行います。

例えば、データベース管理システム(DBMS)のインスタンスを引き継ぐ場合、インスタンスの起動時に自動的にデータベースの復旧(ロールフォワード/ロールバックなど)が行われます。このデータベース復旧に必要な時間は、DBMS のチェックポイントインターバルの設定などによってある程度の制御ができますが、一般的には数分程度必要となるようです。

多くのアプリケーションは再実行するだけで業務を再開できますが、障害発生後の業務復旧手順が必要なアプリケーションもあります。このようなアプリケーションのためにクラスタソフトウェアは業務復旧手順を記述できるよう、アプリケーションの起動の代わりにスクリプトを起動できるようにになっています。スクリプト内には、スクリプトの実行要因や実行サーバなどの情報をもとに、必要に応じて更新途中であつたファイルのクリーンアップなどの復旧手順を記述します。

フェイルオーバー総括

ここまでの内容から、次のようなクラスタソフトの動作が分かります。

- ◆ 障害検出(ハートビート/リソース監視)
- ◆ ネットワークパーティション症状解決(NP解決)
- ◆ クラスタ資源切り替え
 - データの引き継ぎ
 - IP アドレスの引き継ぎ
 - アプリケーションの引き継ぎ



図 1-8 フェイルオーバータイムチャート

クラスタソフトウェアは、フェイルオーバー実現のため、これらの様々な処置を 1 つ 1 つ確実に、短時間で実行することで、高可用性(High Availability)を実現しているのです。

Single Point of Failure の排除

高可用性システムを構築するうえで、求められるもしくは目標とする可用性のレベルを把握することは重要です。これはすなわち、システムの稼働を阻害し得るさまざまな障害に対して、冗長構成をとることで稼働を継続したり、短い時間で稼働状態に復旧したりするなどの施策を費用対効果の面で検討し、システムを設計するということです。

Single Point of Failure(SPOF)とは、システム停止につながる部位を指す言葉であると前述しました。クラスタシステムではサーバの多重化を実現し、システムの SPOF を排除することができますが、共有ディスクなど、サーバ間で共有する部分については SPOF となり得ます。この共有部分を多重化もしくは排除するようシステム設計することが、高可用性システム構築の重要なポイントとなります。

クラスタシステムは可用性を向上させますが、フェイルオーバーには数分程度のシステム切り替え時間が必要となります。従って、フェイルオーバー時間は可用性の低下要因の 1 つともいえます。このため、高可用性システムでは、まず単体サーバの可用性を高める ECC メモリや冗長電源などの技術が本来重要なのですが、ここでは単体サーバの可用性向上技術には触れず、クラスタシステムにおいて SPOF となりがちな下記の 3 つについて掘り下げて、どのような対策があるか見ていきたいと思います。

- ◆ 共有ディスク
- ◆ 共有ディスクへのアクセスパス
- ◆ LAN

共有ディスク

通常、共有ディスクはディスクアレイにより RAID を組むので、ディスクのベアドライブは SPOF となりません。しかし、RAID コントローラを内蔵するため、コントローラが問題となります。多くのクラスタシステムで採用されている共有ディスクではコントローラの二重化が可能になっています。

二重化された RAID コントローラの利点を生かすためには、通常は共有ディスクへのアクセスパスの二重化を行う必要があります。ただし、二重化された複数のコントローラから同時に同一の論理ディスクユニット(LUN)へアクセスできるような共有ディスクの場合、それぞれのコントローラにサーバを 1 台ずつ接続すればコントローラ異常発生時にノード間フェイルオーバーを発生させることで高可用性を実現できます。

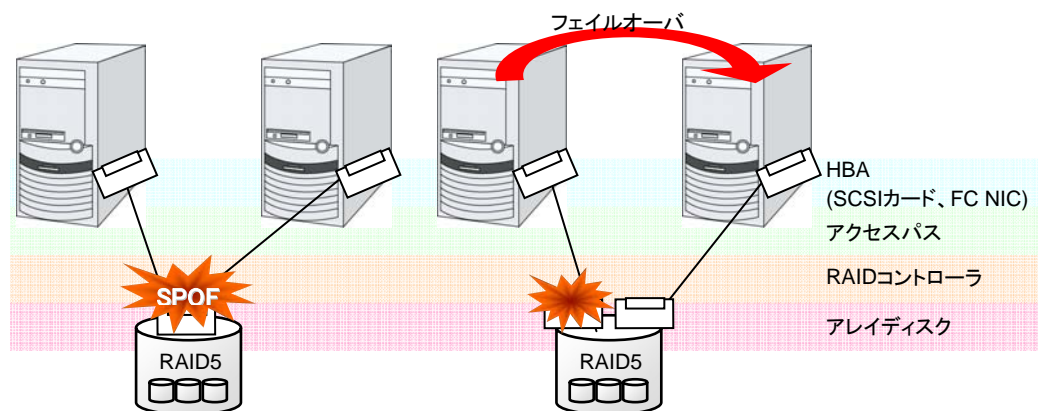


図 1-9 共有ディスクの RAID コントローラとアクセスパスが SPOF となっている例(左)と RAID コントローラとアクセスパスを分割した例

一方、共有ディスクを使用しないデータミラー型のフェイルオーバークラスタでは、すべてのデータをほかのサーバのディスクにミラーリングするため、SPOF が存在しない理想的なシステム構成を実現できます。ただし、欠点とはいえなくても、次のような点について考慮する必要があります。

- ◆ ネットワークを介してデータをミラーリングすることによるディスクI/O性能(特にwrite性能)
- ◆ サーバ障害後の復旧における、ミラー再同期中のシステム性能(ミラーコピーはバックグラウンドで実行される)
- ◆ ミラー再同期時間(ミラー再同期が完了するまでクラスタに組み込めない)

すなわち、データの参照が多く、データ容量が多くないシステムにおいては、データミラー型のフェイルオーバークラスタを採用するというのも可用性を向上させるポイントといえます。

共有ディスクへのアクセスパス

共有ディスク型クラスタの一般的な構成では、共有ディスクへのアクセスパスはクラスタを構成する各サーバで共有されます。SCSI を例に取れば、1 本の SCSI バス上に 2 台のサーバと共有ディスクを接続するということです。このため、共有ディスクへのアクセスパスの異常はシステム全体の停止要因となり得ます。

対策としては、共有ディスクへのアクセスパスを複数用意することで冗長構成とし、アプリケーションには共有ディスクへのアクセスパスが 1 本であるかのように見せることが考えられます。これを実現するデバイスドライバをパスフェイルオーバードライバなどと呼びます (パスフェイルオーバードライバは共有ディスクベンダーが開発してリリースするケースが多いのですが、Linux 版のパスフェイルオーバードライバは開発途上であったりしてリリースされていないようです。現時点では前述のとおり、共有ディスクのアレイコントローラごとにサーバを接続することで共有ディスクへのアクセスパスを分割する手法が Linux クラスタにおいては可用性確保のポイントとなります)。

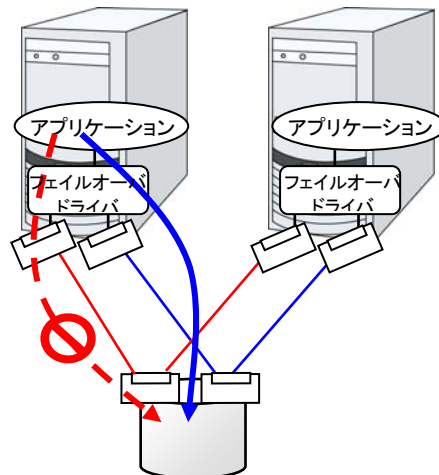


図 1-10 パスフェイルオーバードライバ

LAN

クラスタシステムに限らず、ネットワーク上で何らかのサービスを実行するシステムでは、LAN の障害はシステムの稼働を阻害する大きな要因です。クラスタシステムでは適切な設定を行えば NIC 障害時にノード間でフェイルオーバーを発生させて可用性を高めることは可能ですが、クラスタシステムの外側のネットワーク機器が故障した場合はやはりシステムの稼働を阻害します。

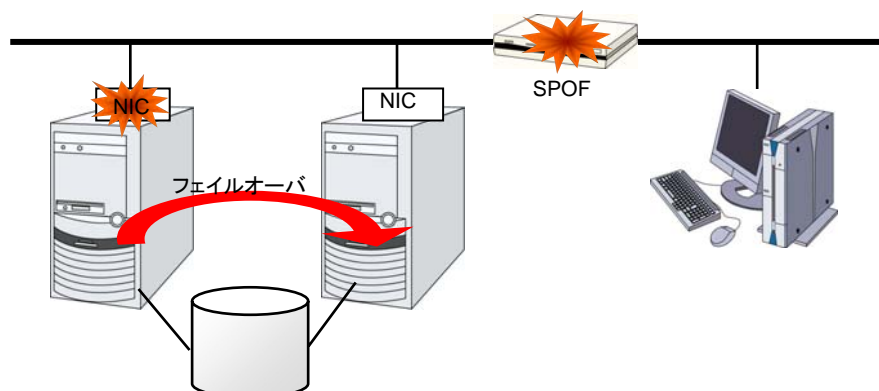


図 1-11 ルータが SPOF となる例

このようなケースでは、LAN を冗長化することでシステムの可用性を高めます。クラスタシステムにおいても、LAN の可用性向上には単体サーバでの技術がそのまま利用可能です。例えば、予備のネットワーク機器の電源を入れずに準備しておき、故障した場合に手動で入れ替えるといった原始的な手法や、高機能のネットワーク機器を冗長配置してネットワーク経路を多重化することで自動的に経路を切り替える方法が考えられます。また、インテル社の ANS ドライバのように NIC の冗長構成をサポートするドライバを利用することも考えられます。

ロードバランス装置 (Load Balance Appliance) やファイアウォールサーバ (Firewall Appliance) も SPOF となりやすいネットワーク機器です。これらもまた、標準もしくはオプションソフトウェアを利用することで、フェイルオーバー構成を組めるようになっているのが普通です。同時にこれらの機器は、システム全体の非常に重要な位置に存在するケースが多いため、冗長構成をとることはほぼ必須と考えるべきです。

可用性を支える運用

運用前評価

システムトラブルの発生要因の多くは、設定ミスや運用保守に起因するものであるといわれています。このことから考えても、高可用性システムを実現するうえで運用前の評価と障害復旧マニュアルの整備はシステムの安定稼働にとって重要です。評価の観点としては、実運用に合わせて、次のようなことを実践することが可用性向上のポイントとなります。

- ◆ 障害発生箇所を洗い出し、対策を検討し、擬似障害評価を行い実証する
- ◆ クラスタのライフサイクルを想定した評価を行い、縮退運転時のパフォーマンスなどの検証を行う
- ◆ これらの評価をもとに、システム運用、障害復旧マニュアルを整備する

クラスタシステムの設計をシンプルにすることは、上記のような検証やマニュアルが単純化でき、システムの可用性向上のポイントとなることが分かります。

障害の監視

上記のような努力にもかかわらず障害は発生するものです。ハードウェアには経年劣化があり、ソフトウェアにはメモリリークなどの理由や設計当初のキャパシティプランニングを超えた運用をしてしまうことによる障害など、長期間運用を続ければ必ず障害が発生してしまいます。このため、ハードウェア、ソフトウェアの可用性向上と同時に、さらに重要となるのは障害を監視して障害発生時に適切に対処することです。万が一サーバに障害が発生した場合を例にとると、クラスタシステムを組むことで数分の切り替え時間でシステムの稼働を継続できますが、そのまま放置しておけばシステムは冗長性を失い次の障害発生時にはクラスタシステムは何の意味もなさなくなってしまう。

このため、障害が発生した場合、すぐさまシステム管理者は次の障害発生に備え、新たに発生した SPOF を取り除くなどの対処をしなければなりません。このようなシステム管理業務をサポートするうえで、リモートメンテナンスや障害の通報といった機能が重要になります。Linux では、リモートメンテナンスの面ではいうまでもなく非常に優れていますし、障害を通報する仕組みも整いつつあります。

以上、クラスタシステムを利用して高可用性を実現するうえで必要とされる周辺技術やそのほかのポイントについて説明しました。簡単にまとめると次のような点に注意しましょうということになるかと思います。

- ◆ Single Point of Failure を排除または把握する
- ◆ 障害に強いシンプルな設計を行い、運用前評価に基づき運用・障害復旧手順のマニュアルを整備する
- ◆ 発生した障害を早期に検出し適切に対処する

第 2 章 CLUSTERPRO の使用方法

本章では、CLUSTERPRO を構成するコンポーネントの説明と、クラスタシステムの設計から運用手順までの流れについて説明します。

本章で説明する項目は以下のとおりです。

• CLUSTERPRO とは?.....	30
• CLUSTERPRO の製品構成.....	30
• CLUSTERPRO のソフトウェア構成.....	30
• ネットワークパーティション解決.....	34
• フェイルオーバーのしくみ.....	34
• リソースとは?.....	43
• CLUSTERPRO を始めよう!.....	47

CLUSTERPRO とは？

クラスタについて理解したところで、CLUSTERPRO の紹介を始めましょう。CLUSTERPRO とは、冗長化（クラスタ化）したシステム構成により、現用系のサーバでの障害が発生した場合に、自動的に待機系のサーバで業務を引き継がせることで、飛躍的にシステムの可用性と拡張性を高めることを可能にするソフトウェアです。

CLUSTERPRO の製品構成

CLUSTERPRO は大きく分けると 3 つのモジュールから構成されています。

- ◆ CLUSTERPRO Server

CLUSTERPRO の本体で、サーバの高可用性機能の全てが包含されています。また、WebManager のサーバ側機能も含まれます。

- ◆ CLUSTERPRO WebManager (WebManager)

CLUSTERPRO の運用管理を行うための管理ツールです。ユーザインターフェイスとして Web ブラウザを利用します。実体は CLUSTERPRO Server に組み込まれていますが、操作は管理端末上の Web ブラウザで行うため、CLUSTERPRO Server 本体とは区別されています。

- ◆ CLUSTERPRO Builder (Builder)

CLUSTERPRO の構成情報を作成するためのツールです。WebManager と同じく、ユーザインターフェイスとして Web ブラウザを利用します。Builder は Builder を利用する端末上で、CLUSTERPRO Server とは別にインストールする必要があります。

CLUSTERPRO のソフトウェア構成

CLUSTERPRO のソフトウェア構成は次の図のようになります。Linux サーバ上には「CLUSTERPRO Server (CLUSTERPRO 本体)」をインストールします。Builder は、管理 PC、あるいはサーバ上にインストールします。WebManager の本体 (CLUSTERPRO Server) は、CLUSTERPRO 本体と同時にインストールされているため、別途インストールする必要はありません。WebManager を利用する端末 (ブラウザ) は勿論管理 PC 上でも問題ありません。

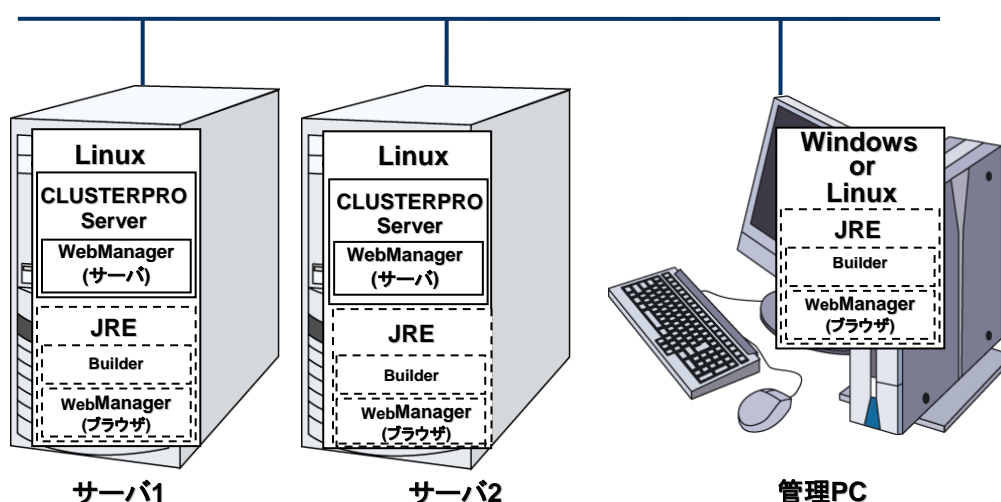


図 2-1 CLUSTERPRO のソフトウェア構成

CLUSTERPRO の障害監視のしくみ

CLUSTERPRO では、サーバ監視、業務監視、内部監視の 3 つの監視を行うことで、迅速かつ確実な障害検出を実現しています。以下にその監視の詳細を示します。

サーバ監視とは

サーバ監視とはフェイルオーバー型クラスタシステムの最も基本的な監視機能で、クラスタを構成するサーバが停止していないかを監視する機能です。

CLUSTERPRO はサーバ監視のために、定期的にサーバ同士で生存確認を行います。この生存確認をハートビートと呼びます。ハートビートは以下の通信パスを使用して行います。

◆ インタコネクト専用LAN

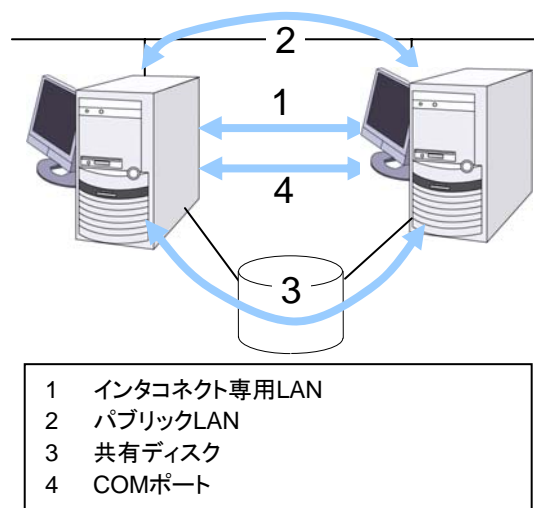
フェイルオーバー型クラスタ専用の通信パスで、一般の Ethernet NIC を使用します。ハートビートを行うと同時にサーバ間の情報交換に使用します。

◆ パブリックLAN

クライアントとの通信に使用している通信パスを予備のインタコネクトとして使用します。TCP/IP が使用できる NIC であればどのようなものでも構いません。ハートビートを行うと同時にサーバ間の情報交換に使用します。

◆ 共有ディスク

フェイルオーバー型クラスタを構成する全てのサーバに接続されたディスク上に、CLUSTERPRO 専用のパーティション(CLUSTER パーティション)を作成し、CLUSTER パーティション上でハートビートを行います。



◆ COM ポート

フェイルオーバー型クラスタを構成するサーバ間を、COM ポートを介してハートビート通信を行い、他サーバの生存を確認します。

これらの通信経路を使用することでサーバ間の通信の信頼性は飛躍的に向上し、ネットワークパーティション症状の発生を防ぎます。

注：ネットワークパーティション症状(Split-brain-syndrome)について：クラスタサーバ間の全ての通信路に障害が発生しネットワーク的に分断されてしまう状態のことです。ネットワークパーティション症状に対応できていないクラスタシステムでは、通信路の障害とサーバの障害を区別できず、同一資源を複数のサーバからアクセスしデータ破壊を引き起こす場合があります。

業務監視とは

業務監視とは、業務アプリケーションそのものや業務が実行できない状態に陥る障害要因を監視する機能です。

◆ アプリケーションの死活監視

アプリケーションを起動用のリソース (EXEC リソースと呼びます) により起動を行い、監視用のリソース (PID モニタリソースと呼びます) により定期的にプロセスの生存を確認することで実現します。業務停止要因が業務アプリケーションの異常終了である場合に有効です。

注：

- CLUSTERPRO が直接起動したアプリケーションが監視対象の常駐プロセスを起動し終了してしまうようなアプリケーションでは、常駐プロセスの異常を検出することはできません。
 - アプリケーションの内部状態の異常 (アプリケーションのストールや結果異常) を検出することはできません。
-

◆ リソースの監視

CLUSTERPRO のモニタリソースによりクラスタリソース(ディスクパーティション、IP アドレスなど)やパブリック LAN の状態を監視することで実現します。業務停止要因が業務に必要なリソースの異常である場合に有効です。

内部監視とは

内部監視とは、CLUSTERPRO 内部のモジュール間相互監視です。CLUSTERPRO の各監視機能が正常に動作していることを監視します。

次のような監視を CLUSTERPRO 内部で行っています。

◆ CLUSTERPROプロセスの死活監視

監視できる障害と監視できない障害

CLUSTERPRO には、監視できる障害とできない障害があります。クラスタシステム構築時、運用時に、どのような監視が検出可能なのか、または検出できないのかを把握しておくことが重要です。

サーバ監視で検出できる障害とできない障害

監視条件: 障害サーバからのハートビートが途絶

- ◆ 監視できる障害の例
 - ハードウェア障害(OS が継続動作できないもの)
 - panic
- ◆ 監視できない障害の例
 - OS の部分的な機能障害(マウス/キーボードのみが動作しない等)

業務監視で検出できる障害とできない障害

監視条件: 障害アプリケーションの消滅、継続的なリソース異常、あるネットワーク装置への通信路切断

- ◆ 監視できる障害の例
 - アプリケーションの異常終了
 - 共有ディスクへのアクセス障害(HBA¹の故障など)
 - パブリック LAN NIC の故障
- ◆ 監視できない障害の例
 - アプリケーションのストール/結果異常

アプリケーションのストール/結果異常を CLUSTERPRO で直接監視することはできませんが、アプリケーションを監視し異常検出時に自分自身を終了するプログラムを作成し、そのプログラムを EXEC リソースで起動、PID モニタリソースで監視することで、フェイルオーバーを発生させることは可能です。

¹ Host Bus Adapterの略で、共有ディスク側ではなく、サーバ本体側のアダプタのことです。
セクション I CLUSTERPRO の概要

ネットワークパーティション解決

CLUSTERPRO は、あるサーバからのハートビート途絶を検出すると、その原因が本当にサーバ障害なのか、あるいはネットワークパーティション症状によるものなのかの判別を行います。サーバ障害と判断した場合は、フェイルオーバ(健全なサーバ上で各種リソースを活性化し業務アプリケーションを起動)を実行しますが、ネットワークパーティション症状と判断した場合には、業務継続よりもデータ保護を優先させるため、緊急シャットダウンなどの処理を実施します。

ネットワークパーティション解決方式には下記の方法があります。

◆ ping 方式

関連情報: ネットワークパーティション解決方法の設定についての詳細は、『リファレンスガイド セクション II』の「第 8 章 ネットワークパーティション解決リソースの詳細」を参照してください。

フェイルオーバのしくみ

CLUSTERPRO は障害を検出すると、フェイルオーバ開始前に検出した障害がサーバの障害かネットワークパーティション症状かを判別します。この後、健全なサーバ上で各種リソースを活性化し業務アプリケーションを起動することでフェイルオーバを実行します。

このとき、同時に移動するリソースの集まりをフェイルオーバグループと呼びます。フェイルオーバグループは利用者から見た場合、仮想的なコンピュータとみなすことができます。

注: クラスタシステムでは、アプリケーションを健全なノードで起動しなおすことでフェイルオーバを実行します。このため、アプリケーションのメモリ上に格納されている実行状態をフェイルオーバすることはできません。

障害発生からフェイルオーバ完了までの時間は数分間必要です。以下にタイムチャートを示します。

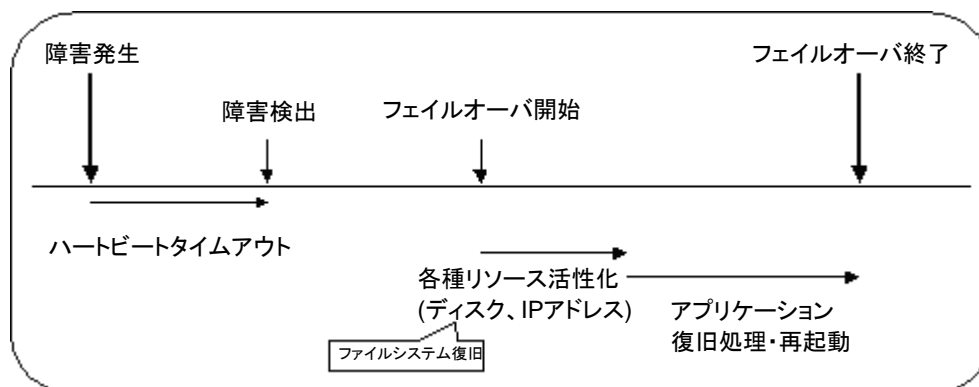


図 2-2 フェイルオーバのタイムチャート

◆ ハートビートタイムアウト

- 業務を実行しているサーバの障害発生後、待機系がその障害を検出するまでの時間です。

- 業務の負荷に応じてクラスタプロパティの設定値を調整します。
(出荷時設定では 90 秒に設定されています。)
- ◆ 各種リソース活性化
 - 業務で必要なリソースを活性化するための時間です。
 - 一般的な設定では数秒で活性化しますが、フェイルオーバーグループに登録されているリソースの種類や数によって必要時間は変化します。
(詳しくは、『CLUSTERPRO インストール & 設定ガイド』を参照してください。)
- ◆ 開始スクリプト実行時間
 - データベースのロールバック/ロールフォワードなどのデータ復旧時間と業務で使用するアプリケーションの起動時間です。
 - ロールバック/ロールフォワード時間など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、各ソフトウェア製品のドキュメントを参照してください。

フェイルオーバーリソース

CLUSTERPRO がフェイルオーバー対象とできる主なリソースは以下のとおりです。

- ◆ 切替パーティション (ディスクリソース、ミラーディスクリソース、ハイブリッドディスクリソースなど)
 - 業務アプリケーションが引き継ぐべきデータを格納するためのディスクパーティションです。
- ◆ フローティングIPアドレス (フローティングIPリソース)
 - フローティング IP アドレスを使用して業務へ接続することで、フェイルオーバーによる業務の実行位置(サーバ)の変化をクライアントは気にする必要がなくなります。
 - パブリック LAN アダプタへの IP アドレス動的割り当てと ARP パケットの送信により実現しています。ほとんどのネットワーク機器からフローティング IP アドレスによる接続が可能です。
- ◆ スクリプト (EXEC リソース)
 - CLUSTERPRO では、業務アプリケーションをスクリプトから起動します。
 - 共有ディスクにて引き継がれたファイルはファイルシステムとして正常であっても、データとして不完全な状態にある場合があります。スクリプトにはアプリケーションの起動のほか、フェイルオーバー時の業務固有の復旧処理も記述します。

注: クラスタシステムでは、アプリケーションを健全なノードで起動しなおすことでフェイルオーバーを実行します。このため、アプリケーションのメモリ上に格納されている実行状態をフェイルオーバーすることはできません。

フェイルオーバー型クラスタのシステム構成

フェイルオーバー型クラスタは、ディスクアレイ装置をクラスタサーバ間で共有します。サーバ障害時には待機系サーバが共有ディスク上のデータを使用し業務を引き継ぎます。

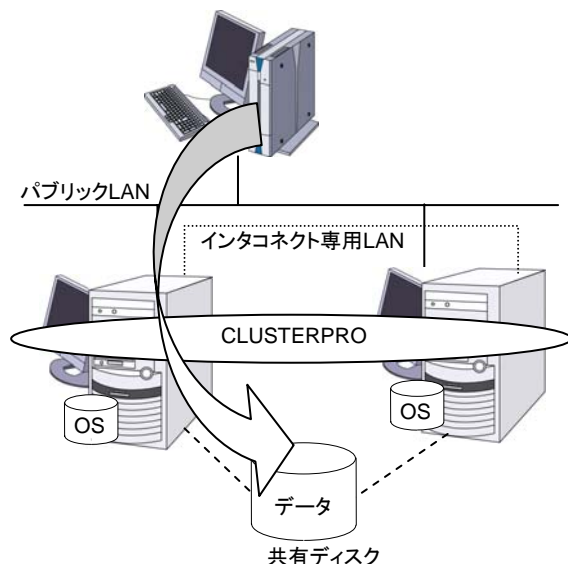


図 2-3 システム構成

フェイルオーバー型クラスタでは、運用形態により、次のように分類できます。

片方向スタンバイクラスタ

一方のサーバを現用系として業務を稼働させ、他方のサーバを待機系として業務を稼働させない運用形態です。最もシンプルな運用形態でフェイルオーバー後の性能劣化のない可用性の高いシステムを構築できます。

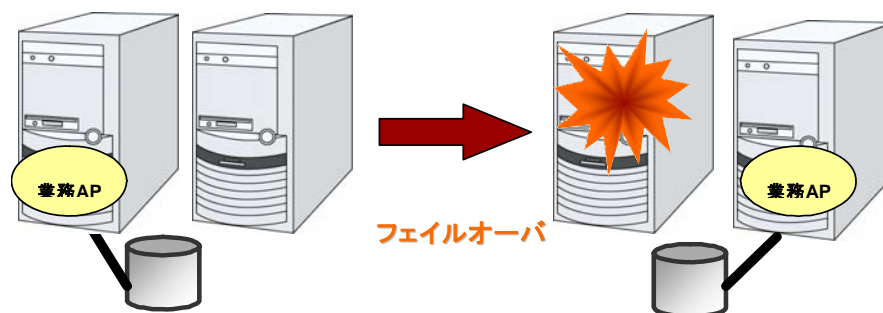


図 2-4 片方向スタンバイクラスタ

同一アプリケーション双方向スタンバイクラスタ

複数のサーバである業務アプリケーションを稼働させ相互に待機する運用形態です。アプリケーションは双方向スタンバイ運用をサポートしているものでなければなりません。ある業務データを複数に分割できる場合に、アクセスしようとしているデータによってクライアントからの接続先サーバを変更することで、データ分割単位での負荷分散システムを構築できます。

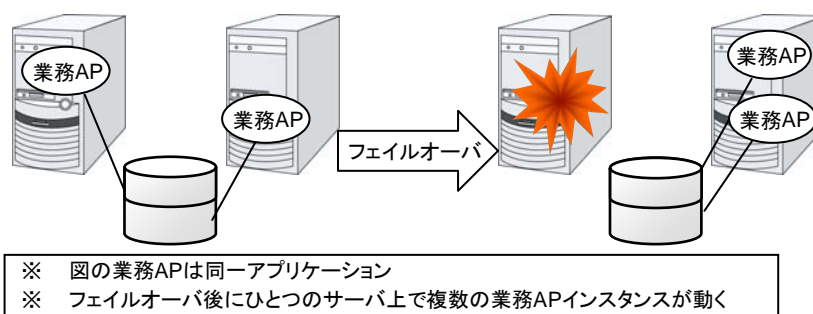


図 2-5 同一アプリケーション双方向スタンバイクラスタ

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をそれぞれ異なるサーバで稼動させ相互に待機する運用形態です。アプリケーションが双方向スタンバイ運用をサポートしている必要はありません。業務単位での負荷分散システムを構築できます。

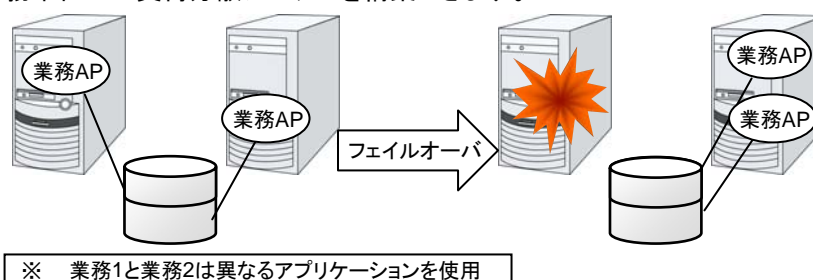


図 2-6 異種アプリケーション双方向スタンバイクラスタ

N + N 構成

ここまでの構成を応用し、より多くのノードを使用した構成に拡張することも可能です。下図は、3種の業務を3台のサーバで実行し、いざ問題が発生した時には1台の待機系にその業務を引き継ぐという構成です。片方向スタンバイでは、正常時のリソースの無駄は 1/2 ですが、この構成なら正常時の無駄を 1/4 まで削減でき、かつ、1台までの異常発生であればパフォーマンスの低下もありません。

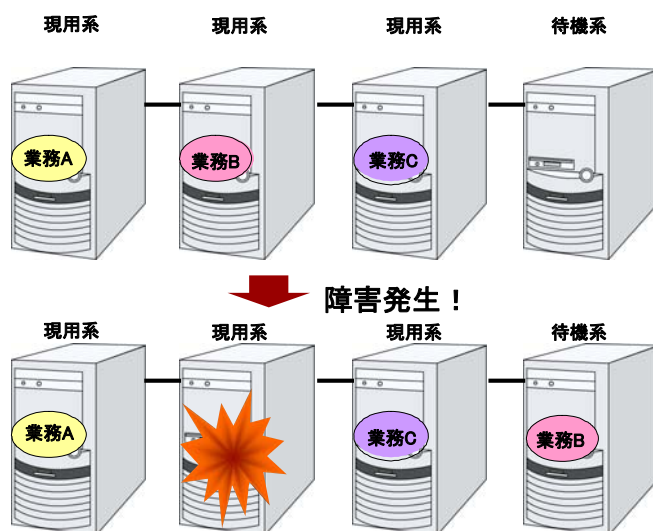


図 2-7 N + N 構成

共有ディスク型のハードウェア構成

共有ディスク構成の CLUSTERPRO の HW 構成は下図のようになります。

サーバ間の通信用に

- ◆ NICを2枚 (1枚は外部との通信と流用、1枚はCLUSTERPRO専用)
- ◆ RS232Cクロスケーブルで接続されたCOMポート
- ◆ 共有ディスクの特定領域

を利用する構成が一般的です。

共有ディスクとの接続インターフェイスは SCSI か FibreChannel ですが、最近では FibreChannel による接続が一般的です。

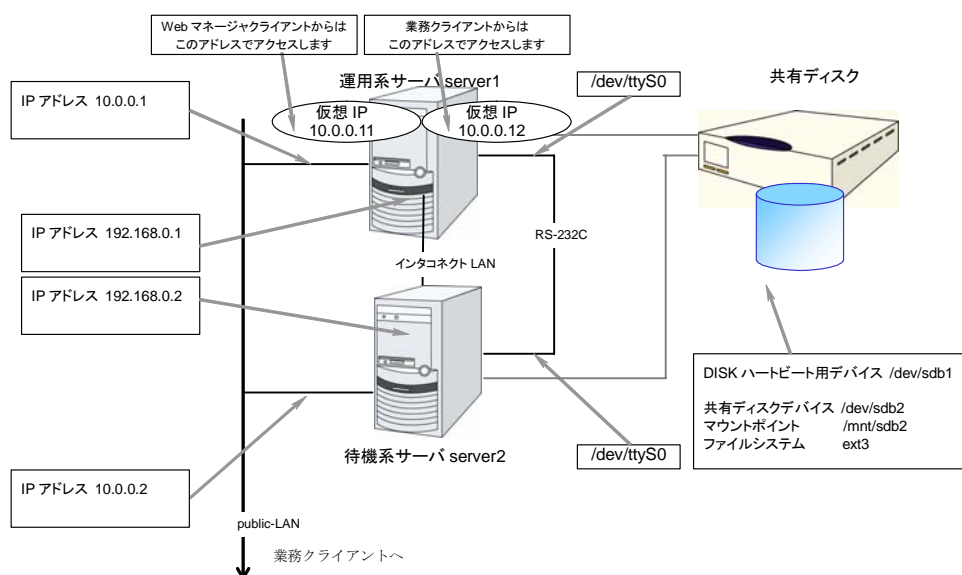


図 2-8 共有ディスク使用時のクラスタ環境のサンプル

ミラーディスク型のハードウェア構成

データミラー構成の CLUSTERPRO は、下図のような構成になります。

共有ディスク構成と比べ、ミラーディスクデータコピー用のネットワークが必要となりますが、通常、CLUSTERPRO の内部通信用 NIC と兼用します。

また、ミラーディスクは接続インターフェイス(IDE or SCSI)には依存しません。

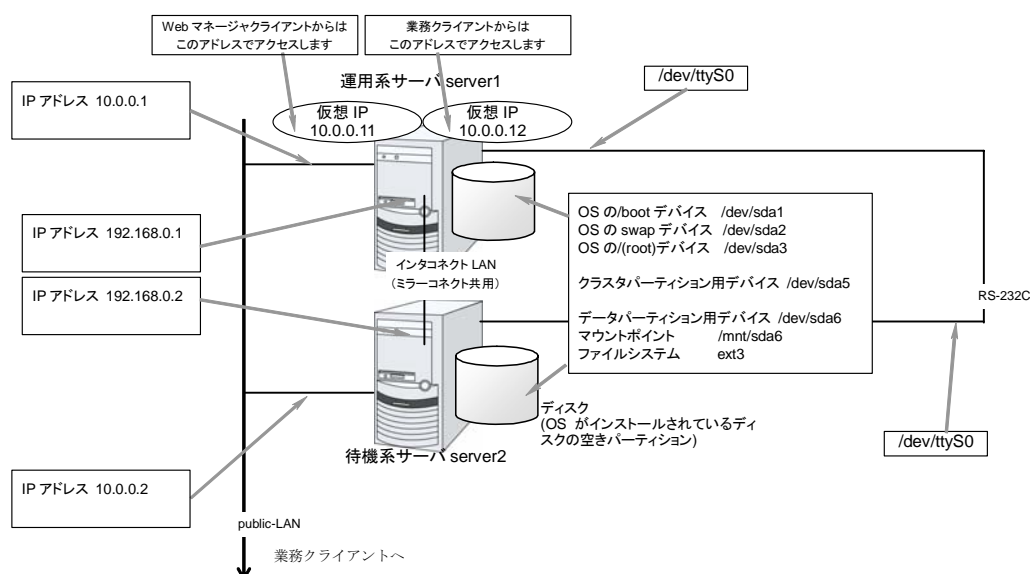


図 2-9 ミラーディスク使用時のクラスタ環境のサンプル(OS がインストールされているディスクにクラスタパーティション、データパーティションを確保する場合)

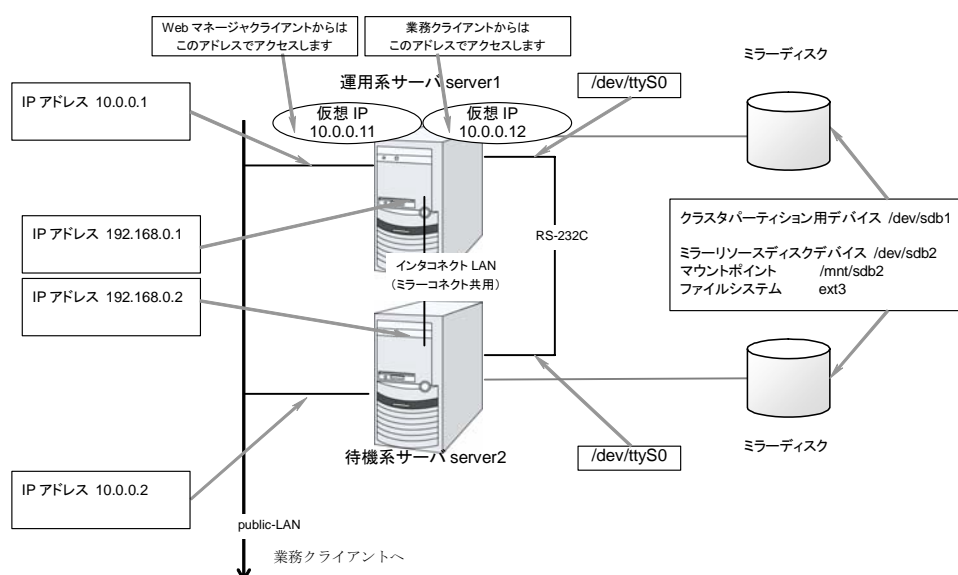


図 2-10 ミラーディスク使用時のクラスタ環境のサンプル(クラスタパーティション、データパーティション用のディスクを用意する場合)

ハイブリッドディスク型のハードウェア構成

ハイブリッド構成の CLUSTERPRO は、下図のような構成になります。

共有ディスク構成と比べ、データコピー用のネットワークが必要となりますが、通常、CLUSTERPRO の内部通信用 NIC と兼用します。

また、ディスクは接続インターフェイス(IDE or SCSI)には依存しません。

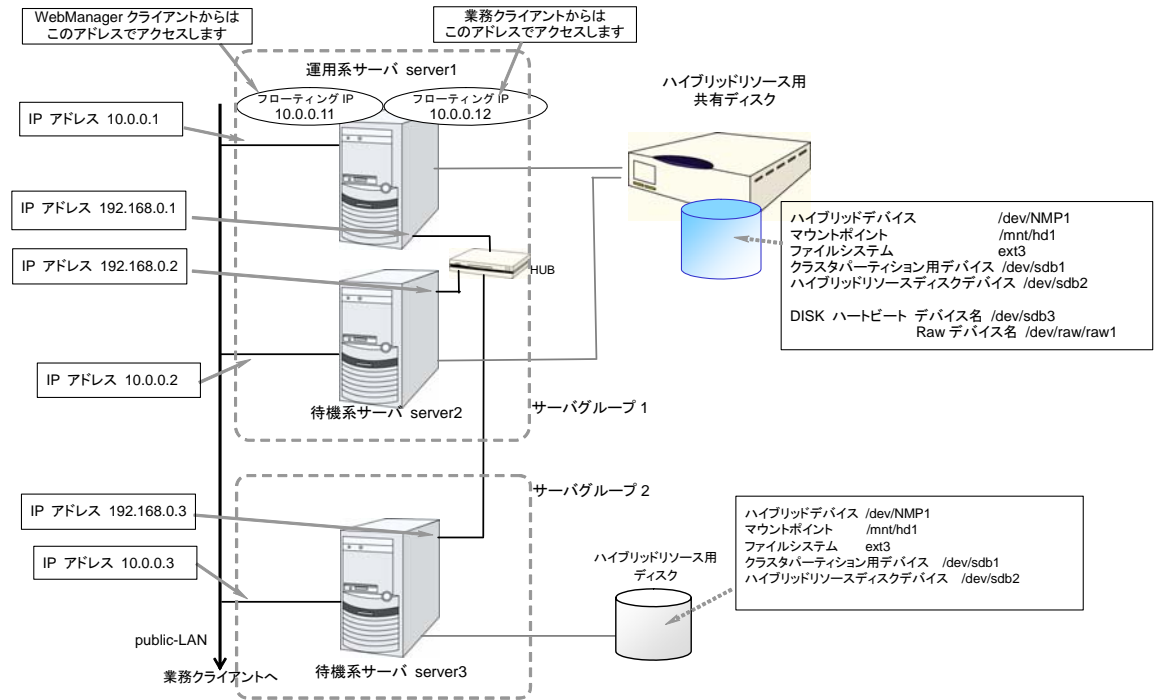


図 2-11 ハイブリッドディスク使用時のクラスタ環境のサンプル(2 台のサーバで共有ディスクを使用し、3 台目のサーバの通常のディスクへミラーリングする場合)

クラスタオブジェクトとは？

CLUSTERPRO では各種リソースを下のような構成で管理しています。

- ◆ クラスタオブジェクト
クラスタの構成単位となります。
- ◆ サーバオブジェクト
実体サーバを示すオブジェクトで、クラスタオブジェクトに属します。
- ◆ サーバグループオブジェクト
サーバを束ねるオブジェクトで、クラスタオブジェクトに属します。
- ◆ ハートビートリソースオブジェクト
実体サーバのNW部分を示すオブジェクトで、サーバオブジェクトに属します。
- ◆ ネットワークパーティション解決リソースオブジェクト
ネットワークパーティション解決機構を示すオブジェクトで、サーバオブジェクトに属します。
- ◆ グループオブジェクト
仮想サーバを示すオブジェクトで、クラスタオブジェクトに属します。
- ◆ グループリソースオブジェクト
仮想サーバの持つリソース(NW、ディスク)を示すオブジェクトでグループオブジェクトに属します。
- ◆ モニタリソースオブジェクト
監視機構を示すオブジェクトで、クラスタオブジェクトに属します。

リソースとは？

CLUSTERPRO では、監視する側とされる側の対象をすべてリソースと呼び、分類して管理します。このことにより、より明確に監視/被監視の対象を区別できるほか、クラスタ構築や障害検出時の対応が容易になります。リソースはハートビートリソース、ネットワークパーティション解決リソース、グループリソース、モニタリソースの 4 つに分類されます。以下にその概略を示します。

ハートビートリソース

サーバ間で、お互いの生存を確認するためのリソースです。

以下に現在サポートされているハートビートリソースを示します。

- ◆ LANハートビートリソース
Ethernetを利用した通信を示します。
- ◆ カーネルモードLANハートビートリソース
Ethernetを利用した通信を示します。
- ◆ COMハートビートリソース
RS232C(COM)を利用した通信を示します。
- ◆ ディスクハートビートリソース
共有ディスク上の特定パーティション(ディスクハートビート用パーティション)を利用した通信を示します。共有ディスク構成の場合のみ利用可能です。

ネットワークパーティション解決リソース

ネットワークパーティション症状を解決するためのリソースを示します。

- ◆ PING ネットワークパーティション解決リソース
PING 方式によるネットワークパーティション解決リソースです。

グループリソース

フェイルオーバーを行う際の単位となる、フェイルオーバーグループを構成するリソースです。

以下に現在サポートされているグループリソースを示します。

- ◆ フローティングIPリソース (fip)
仮想的なIPアドレスを提供します。クライアントからは一般のIPアドレスと同様にアクセス可能です。
- ◆ EXECリソース (exec)
業務(DB、httpd、etc..)を起動／停止するための仕組みを提供します。
- ◆ ディスクリソース (disk)
共有ディスク上の指定パーティションを提供します。(共有ディスク)構成の場合のみ利用可能です。
- ◆ ミラーディスクリソース (md)
ミラーディスク上の指定パーティションを提供します。(ミラーディスク)構成の場合のみ利用可能です。
- ◆ ハイブリッドディスクリソース (hd)

共有ディスク、またはディスク上の指定パーティションを提供します。(ハイブリッド)構成の場合のみ利用可能です。

- ◆ RAWリソース (raw)
共有ディスク上のRAWデバイスを提供します。共有ディスク構成の場合のみ利用可能です。
- ◆ VxVMディスクグループリソース (vxdg)
共有ディスク上のVxVMディスクグループを提供します。VxVMボリュームリソースと共に使用します。(共有ディスク)構成の場合のみ利用可能です
- ◆ VxVMボリュームリソース (vxvol)
共有ディスク上のVxVMボリュームを提供します。VxVMディスクグループリソースと共に使用します。(共有ディスク)構成の場合のみ利用可能です。
- ◆ NASリソース (nas)
NASサーバ上の共有リソースへ接続します。(クラスタサーバがNASのサーバ側として振る舞うリソースではありません。)
- ◆ 仮想 IP リソース (vip)
仮想的な IP アドレスを提供します。クライアントからは一般の IP アドレスと同様にアクセス可能です。ネットワークアドレスの異なるセグメント間で遠隔クラスタを構成する場合に使用します。

モニタリソース

クラスタシステム内で、監視を行う主体であるリソースです。

以下に現在サポートされているモニタリソースを示します。

- ◆ IPモニタリソース (ipw)
外部のIPアドレスの監視機構を提供します。
- ◆ ディスクモニタリソース (diskw)
ディスクの監視機構を提供します。共有ディスクの監視にも利用されます。
- ◆ ミラーディスクモニタリソース (mdw)
ミラーディスクの監視機構を提供します。
- ◆ ミラーディスクコネクタモニタリソース (mdnw)
ミラーディスクコネクタの監視機構を提供します。
- ◆ ハイブリッドディスクモニタリソース (hdw)
ハイブリッドディスクの監視機構を提供します。
- ◆ ハイブリッドディスクコネクタモニタリソース (hdnw)
ハイブリッドディスクコネクタの監視機構を提供します。
- ◆ PIDモニタリソース (pidw)
EXECリソースで起動したプロセスの死活監視機能を提供します。
- ◆ ユーザ空間モニタリソース (userw)
ユーザ空間のストール監視機構を提供します。
- ◆ RAWモニタリソース (raww)
ディスクの監視機構を提供します。RAWデバイスを使用するためreadサイズが小さいのでシステムへの負荷が軽減できます。共有ディスクの監視にも利用されます。
- ◆ NIC Link Up/Downモニタリソース (miiw)
LANケーブルのリンクステータスの監視機構を提供します。

- ◆ VxVMデーモンモニタリソース (vxdw)
VxVMのデーモンの監視機構を提供します。(共有ディスク)構成の場合のみ利用可能です。
- ◆ VxVMボリュームモニタリソース (vxvolw)
VxVMのボリュームの監視機構を提供します。(共有ディスク)構成の場合のみ利用可能です。
- ◆ マルチターゲットモニタリソース (mtw)
複数のモニタリソースを束ねたステータスを提供します。
- ◆ 仮想IPモニタリソース (vipw)
仮想IPリソースのRIPパケットを送出する機構を提供します。
- ◆ ARPモニタリソース (arpw)
フローティングIPリソースまたは仮想IPリソースのARPパケットを送出する機構を提供します。
- ◆ カスタムモニタリソース (genw)
監視処理を行うコマンドやスクリプトがある場合に、その動作結果によりシステムを監視する機構を提供します。
- ◆ DB2 モニタリソース (db2w)
IBM DB2 データベースへの監視機構を提供します。
- ◆ ftp モニタリソース (ftpw)
FTP サーバへの監視機構を提供します。
- ◆ http モニタリソース (httpw)
HTTP サーバへの監視機構を提供します。
- ◆ imap4 モニタリソース (imap4w)
IMAP4 サーバへの監視機構を提供します。
- ◆ MySQL モニタリソース (mysqlw)
MySQL データベースへの監視機構を提供します。
- ◆ nfs モニタリソース (nfsw)
nfs ファイルサーバへの監視機構を提供します。
- ◆ Oracle モニタリソース (oraclew)
Oracle データベースへの監視機構を提供します。
- ◆ OracleAS モニタリソース (oracleasw)
Oracle アプリケーションサーバへの監視機構を提供します。
- ◆ pop3 モニタリソース (pop3w)
POP3 サーバへの監視機構を提供します。
- ◆ PostgreSQL モニタリソース (psqlw)
PostgreSQL データベースへの監視機構を提供します。
- ◆ samba モニタリソース (sambaw)
samba ファイルサーバへの監視機構を提供します。
- ◆ smtp モニタリソース (smtpw)
SMTP サーバへの監視機構を提供します。
- ◆ Sybase モニタリソース (sybasew)
Sybase データベースへの監視機構を提供します。

- ◆ Tuxedo モニタリソース (tuxw)
Tuxedo アプリケーションサーバへの監視機構を提供します。
- ◆ Websphere モニタリソース (wasw)
Websphere アプリケーションサーバへの監視機構を提供します。
- ◆ Weblogic モニタリソース (wls w)
Weblogic アプリケーションサーバへの監視機構を提供します。
- ◆ WebOTX モニタリソース (otxw)
WebOTX アプリケーションサーバへの監視機構を提供します。

CLUSTERPRO を始めよう!

以上で CLUSTERPRO の簡単な説明が終了しました。

以降は、以下の流れに従い、対応するガイドを読み進めながら CLUSTERPRO を使用したクラスタシステムの構築を行ってください。

最新情報の確認

本ガイドのセクション II 『リリースノート (CLUSTERPRO 最新情報)』を参照してください。

クラスタシステムの設計

『インストール&設定ガイド』の「セクション I クラスタシステムの設計」および『リファレンスガイド』の「セクション II リソース詳細」を参照してください。

クラスタシステムの構築

『インストール&設定ガイド』の全編を参照してください。

オプションの監視コマンドを使用する場合は、監視対象アプリケーション別の『管理者ガイド』を参照してください。

クラスタシステムの運用開始後の障害対応

『リファレンスガイド』の「セクション III メンテナンス情報」を参照してください。

セクション II リリースノート (CLUSTERPRO 最新情報)

このセクションでは、CLUSTERPRO の最新情報を記載します。サポートするハードウェアやソフトウェアについての最新の詳細情報を記載します。また、制限事項や、既知の問題とその回避策についても説明します。

- 第 3 章 CLUSTERPRO の動作環境
- 第 4 章 最新バージョン情報
- 第 5 章 注意制限事項
- 第 6 章 アップデート手順

第 3 章 CLUSTERPRO の動作環境

本章では、CLUSTERPRO の動作環境について説明します。

本章で説明する項目は以下の通りです。

• ハードウェア	52
• ソフトウェア	54
• Builderの動作環境	65
• WebManagerの動作環境	67

ハードウェア

CLUSTERPRO は以下のアーキテクチャのサーバで動作します。

- ◆ IA32
- ◆ x86_64
- ◆ IA64 (Replicator, Replicator DR, Agent, Alert Serviceは未サポート)
- ◆ ppc64 (Replicator, Replicator DR, Agent, Alert Serviceは未サポート)

スペック

CLUSTERPRO Server で必要なスペックは下記の通りです。

- ◆ RS-232Cポート 1つ (3ノード以上のクラスタを構築する場合は不要)
- ◆ Ethernetポート 2つ以上
- ◆ 共有ディスク
- ◆ ミラー用ディスク または ミラー用空きパーティション
- ◆ CD-ROMドライブ

構築、構成変更時にオフライン版 Builderを使用する場合には、オフライン版 Builder とサーバとの間で構成情報のやりとりのため以下が必要です。

- ◆ FDドライブ、USBメモリなどのリムーバブルメディア または
- ◆ オフライン版Builderを動作させるマシンとファイルを共有する手段

動作確認済ディスクインターフェイス

Replicator のミラーディスク、Replicator DR のハイブリッドディスク(共有型でないディスク),として確認済みのディスクタイプは下記の通りです。

ディスクのタイプ	ホスト側ドライバ呼称	備考
IDE	ide	～120GBまで確認済
SCSI	aic7xxx	
SCSI	aic79xx	
SCSI	sym53c8xx	
SCSI	mptbase,mptscsih	
SCSI	mptsas	
RAID	megaraid(SCSIタイプ)	
RAID	megaraid (IDEタイプ)	～275GBまで確認済
S-ATA	sata-nv	～80GBまで確認済
S-ATA	ata-piix	～120GBまで確認済

動作確認済ネットワークインターフェイス

Replicator、Replicator DR のミラーディスク、ハイブリッドディスクのミラーディスクコネクタ(ミラー通信で使用する系)として確認済みのネットワークボードは下記の通りです。

チップ呼称	ドライバ呼称
Intel 82540EM Intel 82544EI Intel 82546EB Intel 82546GB Intel 82573L Intel 80003ES2LAN Intel 631xESB/632xESB	e1000
Broadcom BCM5701 Broadcom BCM5703 Broadcom BCM5721	bcm5700
Broadcom BCM5721	tg3

ここに掲載しているものは代表的な一例であり、これ以外の製品も利用可能です。

ソフトウェア

CLUSTERPRO Serverの動作環境

動作可能なディストリビューションとkernel

CLUSTERPRO 独自の kernel モジュールがあるため、CLUSTERPRO Serverの動作環境は kernel モジュールのバージョンに依存します。適合する kernel モジュール(ドライバ)を提供している kernel バージョンの情報を提示します。

下記以外のバージョンでは正常に動作しません。

IA32

ディストリビューション	kernel バージョン	Replicator Replicator DR サポート	clpka,clpkhb サポート	CLUSTERPRO Version	備考
Turbolinux 11 Server	2.6.23-5 smp64G-2.6.23-5	○	○	2.0.0-1~	
Red Hat Enterprise Linux AS/ES 4 (update5)	2.6.9-55.0.12.EL 2.6.9-55.0.12.ELsmp 2.6.9-55.0.12.ELhugemem	○	○	2.0.0-1~	
Red Hat Enterprise Linux AS/ES 4 (update6)	2.6.9-67.EL 2.6.9-67.ELsmp 2.6.9-67.ELhugemem	○	○	2.0.0-1~	
	2.6.9-67.0.4.EL 2.6.9-67.0.4.ELsmp 2.6.9-67.0.4.ELhugemem	○	○	2.0.0-1~	
	2.6.9-67.0.20.EL 2.6.9-67.0.20.ELsmp 2.6.9-67.0.20.ELhugemem	○	○	2.0.2-1~	
Red Hat Enterprise Linux AS/ES 4 (update7)	2.6.9-78.EL 2.6.9-78.ELsmp 2.6.9-78.ELhugemem	○	○	2.0.2-1~	
	2.6.9-78.0.1.EL 2.6.9-78.0.1.ELsmp 2.6.9-78.0.1.ELhugemem				
	2.6.9-78.0.8.EL 2.6.9-78.0.8.ELsmp 2.6.9-78.0.8.ELhugemem	○	○	2.0.3-1~	
	2.6.9-78.0.13.EL 2.6.9-78.0.13.ELsmp 2.6.9-78.0.13.ELhugemem				
	2.6.9-78.0.17.EL 2.6.9-78.0.17.ELsmp 2.6.9-78.0.17.ELhugemem				
Red Hat Enterprise Linux 5 (update1)	2.6.18-53.el5 PAE-2.6.18-53.el5 xen-2.6.18-53.el5	○	○	2.0.0-1~	

Red Hat Enterprise Linux 5 (update2)	2.6.18-92.el5 PAE-2.6.18-92.el5 xen-2.6.18-92.el5	○	○	2.0.2-1~	
	2.6.18-92.1.6.el5 PAE-2.6.18-92.1.6.el5 xen-2.6.18-92.1.6.el5				
	2.6.18-92.1.10.el5 PAE-2.6.18-92.1.10.el5 xen-2.6.18-92.1.10.el5				
	2.6.18-92.1.13.el5 PAE-2.6.18-92.1.13.el5 xen-2.6.18-92.1.13.el5	○	○	2.0.3-1~	
	2.6.18-92.1.22.el5 PAE-2.6.18-92.1.22.el5 xen-2.6.18-92.1.22.el5				
MIRACLE LINUX V4.0 SP2	2.6.9-42.18AX 2.6.9-42.18AXsmp 2.6.9-42.18AXhugemem	○	○	2.0.0-1~	
	2.6.9-42.23AX 2.6.9-42.23AXsmp 2.6.9-42.23AXhugemem	○	○	2.0.2-1~	
	2.6.9-42.26AX 2.6.9-42.26AXsmp 2.6.9-42.26AXhugemem				
MIRACLE LINUX V4.0 SP3	2.6.9-78.8AXS2 2.6.9-78.8AXS2smp 2.6.9-78.8AXS2hugemem	○	○	2.0.3-1~	
Asianux Server 3	2.6.18-8.10AX 2.6.18-8.10AXPAE 2.6.18-8.10AXxen	○	○	2.0.0-1~	
	2.6.18-8.15AX 2.6.18-8.15AXPAE 2.6.18-8.15AXxen	○	○	2.0.0-1~	
Asianux Server 3 (SP1)	2.6.18-53.11AXS3 2.6.18-53.11AXS3PAE 2.6.18-53.11AXS3xen	○	○	2.0.2-1~	
Novell SUSE LINUX Enterprise Server 10 (SP1)	2.6.16.46-0.12-default 2.6.16.46-0.12-smp 2.6.16.46-0.12-bigsmp 2.6.16.46-0.12-xen	○	○	2.0.0-1~	
Novell SUSE LINUX Enterprise Server 10 (SP2)	2.6.16.60-0.21-default 2.6.16.60-0.21-smp 2.6.16.60-0.21-bigsmp 2.6.16.60-0.21-xen	○	○	2.0.2-1~	
VMware ESX Server 3.5 (update2)	2.4.21-57.ELvmnix	×	○	2.0.2-1~	

x86_64

ディストリビューション	kernel バージョン	Replicator Replicator DR サポート	clpka,clpkhb サポート	CLUSTERPRO Version	備考
Turbolinux 11 Server	2.6.23-5	○	○	2.0.0-1~	
Red Hat Enterprise Linux AS/ES 4 (update5)	2.6.9-55.0.12.EL 2.6.9-55.0.12.ELsmp 2.6.9-55.0.12.ELlargesmp	○	○	2.0.0-1~	
Red Hat Enterprise Linux AS/ES 4 (update6)	2.6.9-67.EL 2.6.9-67.ELsmp 2.6.9-67.ELlargesmp	○	○	2.0.0-1~	
	2.6.9-67.0.4.EL 2.6.9-67.0.4.ELsmp 2.6.9-67.0.4.ELlargesmp				
	2.6.9-67.0.20.EL 2.6.9-67.0.20.ELsmp 2.6.9-67.0.20.ELlargesmp	○	○	2.0.2-1~	
Red Hat Enterprise Linux AS/ES 4 (update7)	2.6.9-78.EL 2.6.9-78.ELsmp 2.6.9-78.ELlargesmp	○	○	2.0.2-1~	
	2.6.9-78.0.1.EL 2.6.9-78.0.1.ELsmp 2.6.9-78.0.1. ELlargesmp				
	2.6.9-78.0.8.EL 2.6.9-78.0.8.ELsmp 2.6.9-78.0.8.ELlargesmp	○	○	2.0.3-1~	
	2.6.9-78.0.13.EL 2.6.9-78.0.13.ELsmp 2.6.9-78.0.13.ELlargesmp				
	2.6.9-78.0.17.EL 2.6.9-78.0.17.ELsmp 2.6.9-78.0.17.ELlargesmp				
Red Hat Enterprise Linux 5 (update1)	2.6.18-53.el5 xen-2.6.18-53.el5	○	○	2.0.0-1~	
Red Hat Enterprise Linux 5 (update2)	2.6.18-92.el5 xen-2.6.18-92.el5	○	○	2.0.2-1~	
	2.6.18-92.1.6.el5 xen-2.6.18-92.1.6.el5				
	2.6.18-92.1.10.el5 xen-2.6.18-92.1.10.el5				
	2.6.18-92.1.13.el5 xen-2.6.18-92.1.13.el5	○	○	2.0.3-1~	
	2.6.18-92.1.22.el5 xen-2.6.18-92.1.22.el5				
MIRACLE LINUX V4.0 SP2	2.6.9-42.18AX 2.6.9-42.18AXsmp 2.6.9-42.18AXlargesmp	○	○	2.0.0-1~	

ディストリビューション	kernel バージョン	Replicator Replicator DR サポート	clpka,clpkhb サポート	CLUSTERPRO Version	備考
	2.6.9-42.23AX 2.6.9-42.23AXsmp 2.6.9-42.23AXlargesmp	○	○	2.0.2-1~	
	2.6.9-42.26AX 2.6.9-42.26AXsmp 2.6.9-42.26AXlargesmp				
MIRACLE LINUX V4.0 SP3	2.6.9-78.8AXS2 2.6.9-78.8AXS2smp 2.6.9-78.8AXS2largesmp	○	○	2.0.3-1~	
Asianux Server 3	2.6.18-8.10AX 2.6.18-8.10AXxen	○	○	2.0.0-1~	
	2.6.18-8.15AX 2.6.18-8.15AXxen				
Asianux Server 3 (SP1)	2.6.18-53.11AXS3 2.6.18-53.11AXS3xen	○	○	2.0.2-1~	
Novell SUSE LINUX Enterprise Server 10 (SP1)	2.6.16.46-0.12-default 2.6.16.46-0.12-smp 2.6.16.46-0.12-xen	○	○	2.0.0-1~	
Novell SUSE LINUX Enterprise Server 10 (SP2)	2.6.16.60-0.21-default 2.6.16.60-0.21-smp 2.6.16.60-0.21-xen	○	○	2.0.2-1~	
Oracle Enterprise Linux 5 (5.1)	2.6.18-53.0.0.0.1.el5	○	○	2.0.0-1~	

IA64

ディストリビューション	kernel バージョン	Replicator Replicator DR サポート	clpka,clpkhb サポート	CLUSTERPRO Version	備考
Red Hat Enterprise Linux AS/ES 4 (update6)	2.6.9-67.EL 2.6.9-67.ELlargesmp	×	○	2.0.0-1~	
Red Hat Enterprise Linux AS/ES 4 (update7)	2.6.9-78.EL 2.6.9-78.ELlargesmp	×	○	2.0.2-1~	
Red Hat Enterprise Linux 5 (update1)	2.6.18-53.el5 xen-2.6.18-53.el5	×	○	2.0.0-1~	
Red Hat Enterprise Linux 5 (update2)	2.6.18-92.el5 xen-2.6.18-92.el5	×	○	2.0.2-1~	
Asianux 2.0 SP1 準拠 ディストリビューション	2.6.9-34.21AX	×	○	2.0.0-1~	
Asianux Server 3	2.6.18-8.10AX	×	○	2.0.0-1~	
Novell SUSE LINUX Enterprise Server 10 (SP1)	2.6.16.46-0.12-default	×	○	2.0.0-1~	
Novell SUSE LINUX Enterprise Server 10 (SP2)	2.6.16.60-0.21-default	×	○	2.0.2-1~	

ppc64

ディストリビューション	kernel バージョン	Replicator Replicator DR サポート	clpka,clpkhb サポート	CLUSTERPRO Version	備考
Red Hat Enterprise Linux AS/ES 4 (update6)	2.6.9-67.EL 2.6.9-67.ELlargesmp	×	○	2.0.0-1~	
Red Hat Enterprise Linux AS/ES 4 (update7)	2.6.9-78.EL 2.6.9-78.ELlargesmp	×	○	2.0.2-1~	
Red Hat Enterprise Linux 5 (update1)	2.6.18-53.el5	×	○	2.0.0-1~	
Red Hat Enterprise Linux 5 (update2)	2.6.18-92.el5	×	○	2.0.2-1~	
Asianux 2.0 SP1 準拠 ディストリビューション	2.6.9-34.21AX	×	○	2.0.0-1~	
Asianux Server 3	2.6.18-8.10AX	×	○	2.0.0-1~	
Asianux Server 3 (SP1)	2.6.18-53.13AXS3	×	○	2.0.3-1~	
Novell SUSE LINUX Enterprise Server 10 (SP1)	2.6.16.46-0.12-default	×	○	2.0.0-1~	
Novell SUSE LINUX Enterprise Server 10 (SP2)	2.6.16.60-0.21-default	×	○	2.0.2-1~	

監視オプションの動作確認済アプリケーション情報

モニタリソースの監視対象のアプリケーションのバージョンの情報

IA32

モニタリソース	監視対象のアプリケーション	CLUSTERPRO Version	備考
Oracleモニタ	Oracle Database 10g Release 2 (10.2)	2.0.0-1~	
	Oracle Database 11g Release 1 (11.1)	2.0.0-1~	
DB2モニタ	DB2 V9.1	2.0.0-1~	
	DB2 V9.5	2.0.0-1~	
PostgreSQLモニタ	PostgreSQL 8.1	2.0.0-1~	
	PostgreSQL 8.2	2.0.0-1~	
	PostgreSQL 8.3	2.0.0-1~	
	PowerGres Plus V2.1	2.0.0-1~	
	PowerGres Plus V5.0	2.0.0-1~	
MySQLモニタ	MySQL 5.0	2.0.0-1~	
	MySQL 5.1	2.0.0-1~	
Sybaseモニタ	Sybase 12.5	2.0.0-1~	
sambaモニタ	Samba 3.0	2.0.0-1~	
nfsモニタ	NFS 1.0	2.0.0-1~	
	NFS 1.1	2.0.0-1~	
httpモニタ	Apache HTTP Server 2.0	2.0.0-1~	
	Apache HTTP Server 2.2	2.0.0-1~	
smtpモニタ	Sendmail 8.13	2.0.0-1~	

CLUSTERPRO X 2.0 for Linux スタートアップガイド

	Sendmail 8.14	2.0.0-1~	
	Postfix 2.2	2.0.0-1~	
pop3モニタ	Dovecot v0.99	2.0.0-1~	
	Dovecot v1.0	2.0.0-1~	
imap4モニタ	Dovecot v0.99	2.0.0-1~	
	Dovecot v1.0	2.0.0-1~	
ftpモニタ	vsftpd 2.0	2.0.0-1~	
Tuxedoモニタ	Tuxedo 9.1	2.0.0-1~	
OracleASモニタ	Oracle Application Server 10g Release 3 (10.3)	2.0.0-1~	
Weblogicモニタ	WebLogic Server 9.2	2.0.0-1~	
	WebLogic Server 10.0	2.0.0-1~	
Websphereモニタ	WebSphere 6.1	2.0.0-1~	
WebOTXモニタ	WebOTX V7.1	2.0.0-1~	

x86_64

モニタリソース	監視対象のアプリケーション	CLUSTERPRO Version	備考
Oracleモニタ	Oracle Database 10g Release 2 (10.2)	2.0.0-1~	
	Oracle Database 11g Release 1 (11.1)	2.0.0-1~	
DB2モニタ	DB2 V9.1	2.0.0-1~	
	DB2 V9.5	2.0.0-1~	
PostgreSQLモニタ	PostgreSQL 8.1	2.0.0-1~	
	PostgreSQL 8.2	2.0.0-1~	
	PostgreSQL 8.3	2.0.0-1~	
MySQLモニタ	MySQL 5.0	2.0.0-1~	
	MySQL 5.1	2.0.0-1~	
Sybaseモニタ	Sybase 12.5	2.0.0-1~	
sambaモニタ	Samba 3.0	2.0.0-1~	
nfsモニタ	NFS 1.0	2.0.0-1~	
	NFS 1.1	2.0.0-1~	
httpモニタ	Apache HTTP Server 2.0	2.0.0-1~	
	Apache HTTP Server 2.2	2.0.0-1~	
smtpモニタ	Sendmail 8.13	2.0.0-1~	
	Sendmail 8.14	2.0.0-1~	
	Postfix 2.2	2.0.0-1~	
pop3モニタ	Dovecot v0.99	2.0.0-1~	

	Dovecot v1.0	2.0.0-1~	
imap4モニタ	Dovecot v0.99	2.0.0-1~	
	Dovecot v1.0	2.0.0-1~	
ftpモニタ	vsftpd 2.0	2.0.0-1~	
Weblogicモニタ	WebLogic Server 9.2	2.0.0-1~	
	WebLogic Server 10.0	2.0.0-1~	
Websphereモニタ	WebSphere 6.1	2.0.0-1~	
WebOTXモニタ	WebOTX V7.1	2.0.0-1~	

IA64

監視オプションは未サポートです。

PPC64

監視オプションは未サポートです。

必要メモリ容量とディスクサイズ

	必要メモリサイズ		必要ディスクサイズ		備考
	ユーザモード	kernel モード	インストール直後	運用時最大	
IA32	64MB	同期モード、vxfs以外のファイルシステムを使用する場合 64MB+ (2MB×ミラーディスクリソース、ハイブリッドディスクリソース数)	140MB	640MB	
x86_64	64MB	同期モード、vxfsファイルシステムを使用する場合 256MB+ (2MB×ミラーディスクリソース、ハイブリッドディスクリソース数) 非同期モードの場合 (各ミラーディスクリソース、ハイブリッドディスクリソースの使用量の総和) ミラーディスクリソース、ハイブリッドディスクリソースの使用量 2MB + (I/Oサイズ×非同期キュー数) をリソース毎に加算したもの	140MB	640MB	
IA64	80MB	-	30MB	400MB	
ppc64	64MB	-	24MB	400MB	

注：I/O サイズは、vxfs ファイルシステムの場合 128KB、vxfs 以外のファイルシステムの場合 4KB になります。
非同期キュー数の設定値については「リファレンスガイド」の「ミラーディスクリソースを理解する」を参照してください。

Builder の動作環境

動作確認済OS、ブラウザ

最新情報は CLUSTERPRO のホームページで公開されている最新ドキュメントを参照してください。現在の対応状況は下記の通りです。

OS	ブラウザ	言語
Microsoft Windows® XP SP2 (IA32)	IE6 SP2	日本語/英語
	IE7	日本語/英語
Microsoft Windows Vista™ (IA32)	IE7	日本語/英語
Microsoft Windows Server 2003 SP1以降(IA32)	IE6 SP1	日本語/英語
Microsoft Windows Server 2008 (IA32)	IE7	日本語/英語
Novell SUSE LINUX Enterprise Server 9 SP3 (IA32)	FireFox 2.0.0.1	日本語/英語
Novell SUSE LINUX Enterprise Server 10 (IA32)	FireFox 2.0.0.2	日本語/英語
Red Hat Enterprise Linux AS/ES 4 update6 (IA32)	FireFox 1.5.0.12	日本語/英語
Red Hat Enterprise Linux 5 update1 (IA32)	FireFox 1.5.0.12	日本語/英語
MIRACLE LINUX V4.0 SP2 (IA32)	FireFox 1.5.0.9	日本語/英語
Asianux Server 3 (IA32)	FireFox 1.5.0.12	日本語/英語
	Konqueror3.5.5	日本語/英語
Turbolinux 11 Server (IA32)	FireFox 2.0.0.8	日本語/英語

注: 64bit、x86_64、ppc64 上では「Builder」は動作しません。構築時、構成変更時には 32bit マシンを用意してください。

Java実行環境

Builder を使用する場合には、Java 実行環境が必要です。

Sun Microsystems

Java™ Runtime Environment

Version 5.0 Update6 (1.5.0_06)以降

必要メモリ容量/ディスク容量

必要メモリ容量 32MB 以上

必要ディスク容量 5MB(Java 実行環境に必要な容量を除く)

セクション II リリースノート (CLUSTERPRO 最新情報)

オフライン版Builderが対応するCLUSTERPROのバージョン

オフライン版Builderバージョン	CLUSTERPRO X rpmバージョン
2.0.0-1	2.0.0-1
2.0.2-1	2.0.2-1
2.0.2-1	2.0.3-1

注：オフライン版 Builder のバージョンと CLUSTERPRO rpm バージョンは上記の対応表の組み合わせで使用してください。それ以外の組み合わせで使用すると正常に動作しない可能性があります。

WebManager の動作環境

動作確認済OS、ブラウザ

現在の対応状況は下記の通りです。

OS	ブラウザ	言語
Microsoft Windows® XP(IA32)	IE6 SP2	日本語/英語
	IE7	日本語/英語
Microsoft Windows Vista™ (IA32)	IE7	日本語/英語
Microsoft Windows Server 2003 SP1以降(IA32, x86_64)	IE6 SP1	日本語/英語
Microsoft Windows Server 2008 (IA32)	IE7	日本語/英語
Novell SUSE LINUX Enterprise Server 9 SP3 (IA32)	Firefox 2.0.0.1	日本語/英語
Novell SUSE LINUX Enterprise Server 10 (IA32)	Firefox 2.0.0.2	日本語/英語
Red Hat Enterprise Linux AS/ES 4 update6 (IA32)	Firefox 1.5.0.12	日本語/英語
Red Hat Enterprise Linux 5 update1 (IA32)	Firefox 1.5.0.12	日本語/英語
MIRACLE LINUX V4.0 SP2 (IA32)	Firefox 1.5.0.9	日本語/英語
Asianux Server 3 (IA32)	Firefox 1.5.0.12	日本語/英語
	Konqueror3.5.5	日本語/英語
Turbolinux 11 Server (IA32)	Firefox 2.0.0.8	日本語/英語

注: WebManager は 64bit、x86_64、ppc64 の Linux 上では 動作しません。Linux マシンでクラスタの管理をするには 32bit OS を用意してください。

Java実行環境

WebManager を使用する場合には、Java 実行環境が必要です。

Sun Microsystems

Java(TM) Runtime Environment

Version 5.0 Update6 (1.5.0_06)以降

必要メモリ容量/ディスク容量

必要メモリ容量 40MB 以上

必要ディスク容量 600KB(Java 実行環境に必要な容量を除く)

第 4 章 最新バージョン情報

本章では、CLUSTERPRO の最新情報について説明します。新しいリリースで強化された点、改善された点などをご紹介します。

• CLUSTERPRO とマニュアルの対応一覧.....	70
• 機能強化	71
• 修正情報	72

CLUSTERPRO とマニュアルの対応一覧

本書では下記のバージョンの CLUSTERPRO を前提に説明してあります。CLUSTERPRO のバージョンとマニュアルの版数に注意してください。

CLUSTERPROのバージョン	マニュアル	版数	備考
2.0.3-1	インストール & 設定ガイド	第3版	
	スタートアップガイド	第3版	
	リファレンスガイド	第3版	

機能強化

各バージョンにおいて以下の機能強化を実施しています。

項番	内部バージョン	機能強化項目
1	2.0.2-1	新しくリリースされた kernel に対応しました。
2	2.0.2-1	待機系サーバの CPU クロック数を制御して省電力モードにする機能を追加しました。
3	2.0.2-1	グループリソースの活性/非活性異常時と、監視リソースの異常検出時に、最終動作を実行する前に任意のスクリプトを実行する機能を追加しました。
4	2.0.2-1	指定した処理実行要求を他クラスタのサーバに発行する機能 (clptmreq コマンド) を追加しました。
5	2.0.2-1	カスタムモニタリソースを追加しました。
6	2.0.2-1	筐体 ID 連携機能を制御する clpledctrl コマンドを追加しました。
7	2.0.2-1	ミラーディスクリソースとハイブリッドディスクリソースが GPT 形式のハードディスクに対応しました。
8	2.0.2-1	WebManager のアラートビューに表示項目のカスタマイズ機能とポップアップ画面によるメッセージ確認機能を追加しました。
9	2.0.2-1	Oracle 監視リソースにリスナー監視機能を追加しました。
10	2.0.2-1	OracleAS 監視リソースに opmn プロセス監視機能を追加しました。
11	2.0.3-1	新しくリリースされた kernel に対応しました。
12	2.0.3-1	Oracle 監視処理の負荷を低減しました。

修正情報

各バージョンにおいて以下の修正を実施しています。

項番	内部バージョン	修正項目
1	2.0.2-1	時刻変更時、まれにディスク HB が停止してしまうことを修正しました。
2	2.0.2-1	モニタリソース一時停止状態でサスペンドを行ったとき、モニタリソースの再開ができなくなることを修正しました。
3	2.0.2-1	サーバ名が 240 文字までしか利用できない問題を修正しました。
4	2.0.2-1	lankhb のポート番号が重複していると、異常終了してしまう問題を修正しました。
5	2.0.2-1	exec リソースの設定で、パス名にダブルクォートを入力すると設定が反映されない問題を修正しました。
6	2.0.2-1	クラスタ生成ウィザードで一度サーバを登録した後に削除すると、構成情報が不正になることがある問題を修正しました。
7	2.0.3-1	ミラーディスクを手動で切り離した後、ミラー復帰を行うと失敗する可能性がある問題を修正しました。

第 5 章 注意制限事項

本章では、注意事項や既知の問題とその回避策について説明します。

本章で説明する項目は以下の通りです。

• システム構成検討時	74
• OSインストール前、OSインストール時.....	81
• OSインストール後、CLUSTERPROインストール前	86
• CLUSTERPROの情報作成時	92
• CLUSTERPRO運用後	98

システム構成検討時

HW の手配、システム構成、共有ディスクの構成時に留意すべき事項について説明します。

Builder、WebManagerの動作OSについて

- ◆ 64bit、x86_64、ppc64上では「Builder」は動作しません。構築時、構成変更時には32bit マシンを用意してください。
- ◆ WebManager は64bit、x86_64、ppc64 のLinux上では 動作しません。Linuxマシンでクラスタの管理をするには32bit OSを用意してください。

ミラーディスクの要件について

- ◆ ミラーリソースとして使用するディスクは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。
- ◆ ミラーディスクリソースを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対象とすることはできません。
- ◆ ミラーリソースを使用するにはミラー用のパーティション(データパーティションとクラスタパーティション)が必要です。
- ◆ ミラー用のパーティションの確保の方法は以下の2つがあります。
 - OS(root パーティションやswap パーティション)と同じディスク上にミラー用のパーティション(クラスタパーティションとデータパーティション)を確保する
 - OS とは別のディスク(または LUN)を用意(追加)してミラー用のパーティションを確保する
- ◆ 以下を参考に上記を選定してください。
 - 障害時の保守性、性能を重視する場合
 - OS とは別にミラー用のディスクを用意することを推奨します。
 - H/W Raid の仕様の制限で LUN の追加ができない場合
H/W Raid のプリインストールモデルで LUN 構成変更が困難な場合
 - OS と同じディスクにミラー用のパーティションを確保します。
- ◆ ミラーリソースを複数使用する場合には、さらにミラーリソース毎に個別のディスクを用意(追加)することを推奨します。
同一のディスク上に複数のミラーリソースを確保すると性能の低下やミラー復帰に時間がかかることがあります。これらの現象はLinux OSのディスクアクセスの性能に起因するものです。
- ◆ ミラー用のディスクとして使用するにはディスクをサーバ間で同じにする必要があります。
 - ディスクのタイプ
両サーバのミラーディスクまたは、ミラー用のパーティションを確保するディスクは、ディスクのタイプを同じにしてください。
動作確認済みのディスクのタイプについては 52 ページの「動作確認済ディスクインターフェイス」を参照してください。

例)

組み合わせ	サーバ1	サーバ2
OK	SCSI	SCSI
OK	IDE	IDE
NG	IDE	SCSI

- ◆ ミラー用のディスクとして使用するディスクのジオメトリがサーバ間で異なる場合の注意

fdisk コマンドなどで確保したパーティションサイズはシリンダあたりのブロック(ユニット)数でアラインされます。

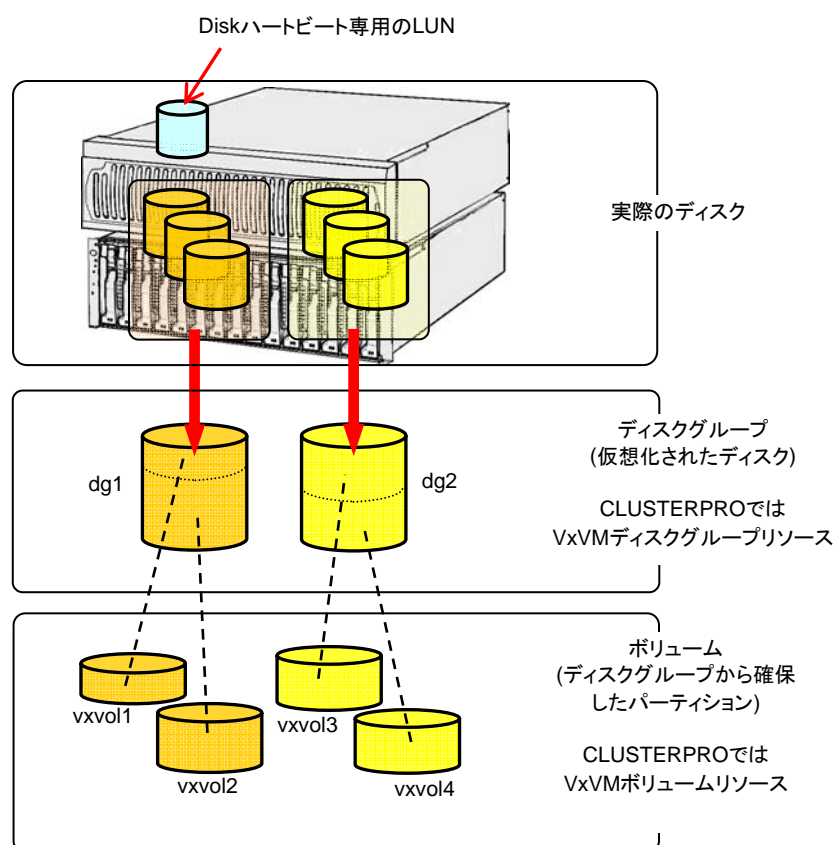
データパーティションのサイズと初期ミラー構築の方向の関係が以下になるようにデータパーティションを確保してください。

コピー元のサーバ ≤ コピー先のサーバ

コピー元のサーバとは、ミラーリソースが所属するフェイルオーバーグループのフェイルオーバーポリシーが高いサーバを指します。コピー先のサーバとは、ミラーリソースが所属するフェイルオーバーグループのフェイルオーバーポリシーが低いサーバを指します。

共有ディスクの要件について

- ◆ 共有ディスクはLinuxのmd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。
- ◆ 共有ディスクでLinuxの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を使用する場合、以下の制限があります。
 - ディスクリソースに設定されたパーティションの ReadOnly,ReadWrite の制御を CLUSTERPRO が行うことができません。
ディスクリソースが活性化されていないサーバで誤って同じファイルシステムをマウントしないように運用回避をしてください。
 - LVM の論理ボリュームをディスクモニタリソースで監視できません。
ディスクモニタリソースの監視先として LVM のボリュームを構成する実デバイスを設定し、マルチターゲットモニタを組み合わせで監視をしてください。
- ◆ VxVMを使用する場合、CLUSTERPROのディスクハートビート用に共有ディスク上に、VxVMで制御対象としないLUNが必要です。共有ディスクのLUNの設計時に留意してください。



ハイブリッドディスクとして使用するディスクの要件について

- ◆ ハイブリッドディスクリソースとして使用するディスクは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。
- ◆ ハイブリッドディスクリソースを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対象とすることはできません。
- ◆ ハイブリッドディスクリソースを使用するにはハイブリッドディスク用のパーティション(データパーティションとクラスタパーティション)が必要です。
- ◆ さらにハイブリッドディスク用のディスクを共有ディスク装置で確保する場合には、共有ディスク装置を共有するサーバ間のディスクハートビートリソース用のパーティションが必要です。
- ◆ ハイブリッドディスク用のディスクを共有ディスク装置でないディスクから確保する場合、パーティションの確保の方法は以下の2つがあります。
 - ・ OS(root パーティションや swap パーティション)と同じディスク上にハイブリッドディスク用のパーティション(クラスタパーティションとデータパーティション)を確保する
 - ・ OS とは別のディスク(または LUN)を用意(追加)してハイブリッドディスク用のパーティションを確保する
- ◆ 以下を参考に上記を選定してください。
 - ・ 障害時の保守性、性能を重視する場合
 - OS とは別にハイブリッドディスク用のディスクを用意することを推奨します。
 - ・ H/W Raid の仕様の制限で LUN の追加ができない場合
 - H/W Raid のプラインストールモデルで LUN 構成変更が困難な場合
 - OS と同じディスクにハイブリッドディスク用のパーティションを確保します。
- ◆ ハイブリッドディスクリソースを複数使用する場合には、さらにハイブリッドディスクリソース毎に個別のLUNを用意(追加)することを推奨します。
 同一のディスク上に複数のハイブリッドディスクリソースを確保すると性能の低下やミラー復帰に時間がかかることがあります。これらの現象はLinux OSのディスクアクセスの性能に起因するものです。

必要なパーティションの種類	ハイブリッドディスクリソースを確保する装置	
	共有ディスク装置	共有型でないディスク装置
データパーティション	必要	必要
クラスタパーティション	必要	必要
ディスクハートビート用パーティション	必要	不要
OSと同じディスク(LUN)上での確保	—	可能

- ◆ ハイブリッドディスク用のディスクとして使用するディスクのタイプやジオメトリがサーバ間で異なる場合の注意

データパーティションのサイズと初期ミラー構築の方向の関係が以下になるようにデータパーティションを確保してください。

コピー元のサーバ ≤ コピー先のサーバ

コピー元のサーバとは、ハイブリッドディスクリソースが所属するフェイルオーバーグループのフェイルオーバーポリシーが高いサーバを指します。コピー先のサーバとは、ハイブリッドディスクリソースが所属するフェイルオーバーグループのフェイルオーバーポリシーが低いサーバを指します。

NIC Link Up/Downモニタリソース

NIC のボード、ドライバによっては、必要な ioctl() がサポートされていない場合があります。NIC Link Up/Down モニタリソースの動作可否は、各ディストリビュータが提供する ethtool コマンドで確認することができます。

```

ethtool eth0
Settings for eth0:
    Supported ports: [ TP ]
    Supported link modes:   10baseT/Half 10baseT/Full
                           100baseT/Half 100baseT/Full
                           1000baseT/Full
    Supports auto-negotiation: Yes
    Advertised link modes:  10baseT/Half 10baseT/Full
                           100baseT/Half 100baseT/Full
                           1000baseT/Full
    Advertised auto-negotiation: Yes
    Speed: 1000Mb/s
    Duplex: Full
    Port: Twisted Pair
    PHYAD: 0
    Transceiver: internal
    Auto-negotiation: on
    Supports Wake-on: umbg
    Wake-on: g
    Current message level: 0x00000007 (7)
    Link detected: yes

```

- ◆ ethtool コマンドの結果で LAN ケーブルのリンク状況 ("Link detected: yes") が表示されない場合
 - CLUSTERPRO の NIC Link Up/Down モニタリソースが動作不可能な可能性が高いです。IP モニタリソースで代替してください。
- ◆ ethtool コマンドの結果で LAN ケーブルのリンク状況 ("Link detected: yes") が表示される場合
 - 多くの場合 CLUSTERPRO の NIC Link Up/Down モニタリソースが動作可能ですが、希に動作不可能な場合があります。
 - 特に以下のようなハードウェアでは動作不可能な場合があります。IP モニタリソースで代替してください。
 - ブレードサーバのように実際の LAN のコネクタと NIC のチップとの間にハードウェアが実装されている場合

実機で CLUSTERPRO を使用して NIC Link Up/Down モニタリソースの使用可否を確認する場合には以下の手順で動作確認を行ってください。

1. NIC Link Up/Down モニタリソースを構成情報に登録してください。
NIC Link Up/Down モニタリソースの異常検出時回復動作の設定は「何もしない」を選択してください。

2. クラスタを起動してください。
3. NIC Link Up/Down モニタリソースのステータスを確認してください。
LAN ケーブルのリンク状態が正常状態時に NIC Link Up/Down モニタリソースのステータスが異常となった場合、NIC Link Up/Down モニタリソースは動作不可です。
4. LAN ケーブルのリンク状態を異常状態(リンクダウン状態)にしたときに NIC Link Up/Down モニタリソースのステータスが異常となった場合、NIC Link Up/Down モニタリソースは動作不可です。
ステータスが正常のまま変化しない場合、NIC Link Up/Down モニタリソースは動作不可です。

ミラーリソース、ハイブリッドディスクリソースのwrite性能について

- ◆ ミラーディスク、ハイブリッドディスクリソースのwrite処理はネットワークを経由して相手サーバのディスクへwrite、自サーバのディスクへwriteを行います。readは自サーバ側のディスクからのみreadします。
- ◆ 上記の理由により、クラスタ化していない単体サーバと比べてwrite性能が劣化します。writeに単体サーバ並みに高スループットが要求されるシステム(更新系が多いデータベースシステムなど)には共有ディスク使用をご提案ください。

OS インストール前、OS インストール時

OS をインストールするときに決定するパラメータ、リソースの確保、ネーミングルールなどで留意して頂きたいことです。

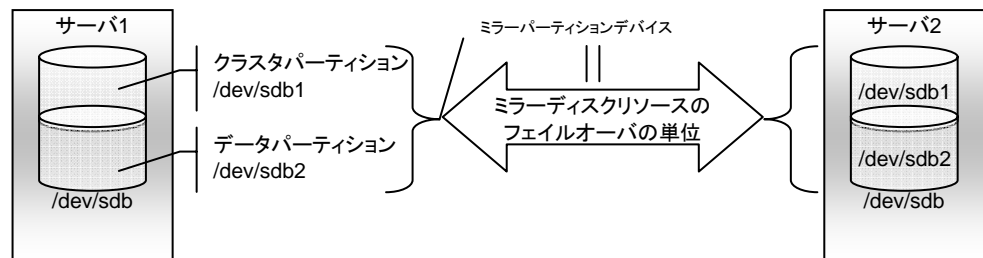
/opt/nec/clusterproのファイルシステムについて

システムの対障害性の向上のために、ジャーナル機能を持つファイルシステムを使用することを推奨します。

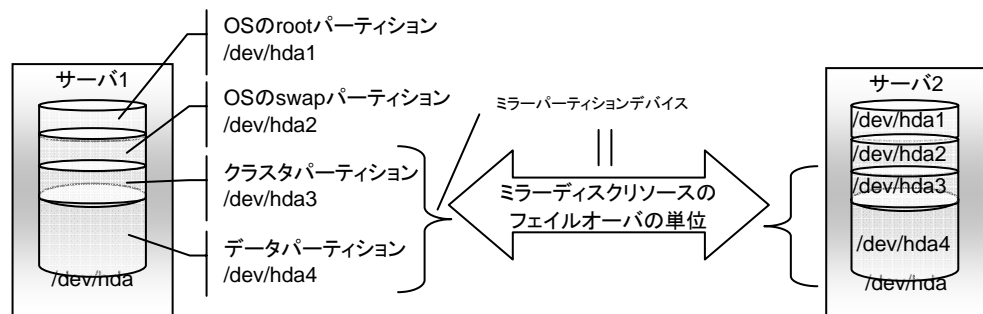
ミラー用のディスクについて

◆ ディスクのパーティション

例)両サーバに 1 つの SCSI ディスクを増設して 1 つのミラーディスクのペアにする場合



例)両サーバの OS が格納されている IDE ディスクの空き領域を使用してミラーディスクのペアにする場合



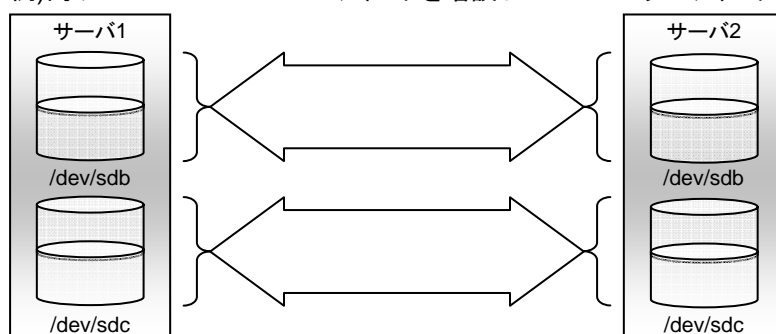
- ミラーパーティションデバイスは CLUSTERPRO のミラーリングドライバが上位に提供するデバイスです。
- クラスタパーティションとデータパーティションの 2 つのパーティションをペアで確保してください。
- OS(root パーティションや swap パーティション)と同じディスク上にミラーパーティション(クラスタパーティション、データパーティション)を確保することも可能です。
 - 障害時の保守性、性能を重視する場合
OS(root パーティションや swap パーティション)と別にミラー用のディスクを用意することを推奨します。
 - H/W Raid の仕様の制限で LUN の追加ができない場合
H/W Raid のブリーインストールモデルで LUN 構成変更が困難な場合
OS(root パーティションや swap パーティション)と同じディスクにミラーパーティション(クラスタパーティション、データパーティション)を確保することも可能です。

◆ ディスクの配置

ミラーディスクとして複数のディスクを使用することができます。

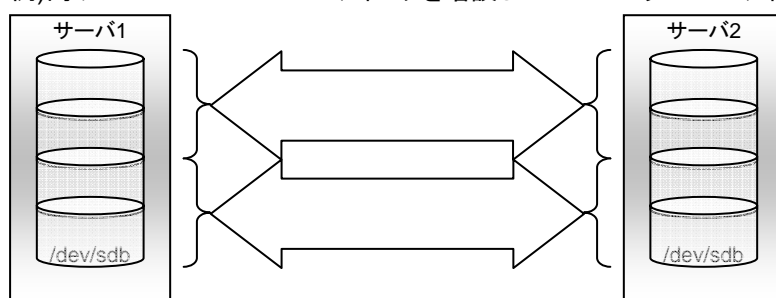
また 1 つのディスクに複数のミラーパーティションデバイスを割り当てて使用することができます。

例)両サーバに 2 つの SCSI ディスクを増設して 2 つのミラーディスクのペアにする場合。



- 1 つのディスク上にクラスタパーティションとデータパーティションをペアで確保してください。
- データパーティションを 1 つ目のディスク、クラスタパーティションを 2 つ目のディスクとするような使い方はできません。

例)両サーバに 1 つの SCSI ディスクを増設して 2 つのミラーパーティションにする場合



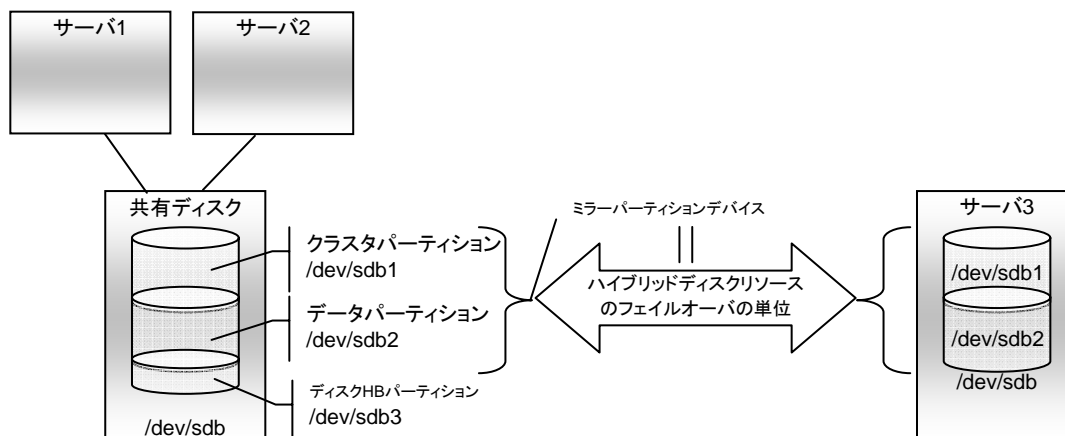
- ◆ ディスクに対してLinuxのmdやLVMによるストライプセット、ボリュームセット、ミラーリング、パリティ付きストライプセットの機能はサポートしていません。

ハイブリッドディスクリソース用のディスクについて

◆ ディスクのパーティション

共有ディスクまたは共有型でないディスク(サーバ内蔵、サーバ間で共有していない外付型ディスク筐体など)を使用することができます。

例) 2 台のサーバで共有ディスクを使用し 3 台目のサーバでサーバに内蔵したディスクを使用する場合



- ミラーパーティションデバイスは CLUSTERPRO のミラーリングドライバが上位に提供するデバイスです。
- クラスタパーティションとデータパーティションの 2 つのパーティションをペアで確保してください。
- 共有型でないディスク(サーバ内蔵、サーバ間で共有していない外付型ディスク筐体など)を使用する場合には OS(root パーティションや swap パーティション)と同じディスク上にミラーパーティション(クラスタパーティション、データパーティション)を確保することも可能です。
 - 障害時の保守性、性能を重視する場合
OS(root パーティションや swap パーティション)と別にミラー用のディスクを用意することを推奨します。
 - H/W Raid の仕様の制限で LUN の追加ができない場合
H/W Raid のプリインストールモデルで LUN 構成変更が困難な場合
OS(root パーティションや swap パーティション)と同じディスクにミラーパーティション(クラスタパーティション、データパーティション)を確保することも可能です。
- さらにハイブリッドディスク用のディスクを共有ディスク装置で確保する場合には、共有ディスク装置を共有するサーバ間のディスクハートビートリソース用のパーティションを確保してください。
- 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、パリティ付きストライプセットの機能はサポートしていません。

依存するライブラリ

libxml2

OS インストール時に、libxml2 をインストールしてください。

依存するドライバ

softdog

- ◆ ユーザ空間モニタリソースの監視方法がsoftdogの場合、このドライバが必要です。
- ◆ ローダブルモジュール構成にしてください。スタティックドライバでは動作しません。

ミラードライバ

- ◆ ミラードライバはメジャー番号 218を使用します。

他のデバイスドライバでは、メジャー番号の 218 を使用しないでください。

カーネルモードLANハートビートドライバ、キープアライブドライバ

- ◆ カーネルモードLANハートビートドライバは、メジャー番号 10、マイナ番号 240を使用します。
- ◆ キープアライブドライバは、メジャー番号 10、マイナ番号 241を使用します。

他のドライバが上記のメジャー及びマイナ番号を使用していないことを確認してください。

RAWモニタリソース用のパーティション確保

- ◆ RAWモニタリソースを設定する場合、監視専用のパーティションを用意してください。パーティションサイズは10MB確保してください。

SELinuxの設定

- ◆ SELinuxの設定は permissive または disabled にしてください。
- ◆ enforcing に設定するとCLUSTERPROで必要な通信が行えない場合があります。

OSインストール後、CLUSTERPROインストール前

OS のインストールが完了した後、OS やディスクの設定を行うときに留意頂いて頂きたいことです。

通信ポート番号

CLUSTERPRO では、以下のポート番号を使用します。このポート番号については「ミラードライバ間キープアライブ」以外は、Builder での変更が可能です。

下記ポート番号には、CLUSTERPRO 以外のプログラムからアクセスしないようにしてください。

サーバにファイアウォールの設定を行う場合には、下記のポート番号にアクセスできるようにしてください。

[サーバ・サーバ間][サーバ内ループバック]

					備考
サーバ	自動割り当て ¹	—	サーバ	29001/TCP	内部通信
サーバ	自動割り当て	—	サーバ	29002/TCP	データ転送
サーバ	自動割り当て	—	サーバ	29002/UDP	ハートビート
サーバ	自動割り当て	—	サーバ	29003/UDP	アラート同期
サーバ	自動割り当て	—	サーバ	29004/TCP	ミラーエージェント間通信
サーバ	自動割り当て	—	サーバ	29006/UDP	ハートビート(カーネルモード)
サーバ	自動割り当て	—	サーバ	XXXX ² /TCP	ミラーディスクリソースデータ同期
サーバ	自動割り当て	—	サーバ	XXXX ³ /TCP	ミラードライバ間通信
サーバ	自動割り当て	—	サーバ	XXXX ⁴ /TCP	ミラードライバ間通信
サーバ	自動割り当て	—	サーバ	icmp	ミラードライバ間キープアライブ、 FIP/VIP リソースの重複確認、 ミラーエージェント
サーバ	自動割り当て	—	サーバ	XXXX ⁵ /UDP	内部ログ用通信

[サーバ・WebManager 間]

					備考
Web Manager	自動割り当て	—	サーバ	29003/TCP	http 通信

[統合 WebManager を接続しているサーバ・管理対象のサーバ間]

					備考
統合 WebManager を接続したサーバ	自動割り当て	—	サーバ	29003/TCP	http 通信

[その他]

					備考
サーバ	自動割り当て	—	ネットワーク警告灯	514/TCP	ネットワーク警告灯制御
サーバ	自動割り当て	—	サーバの BMC のマ ネージメント LAN	623/UDP	BMC 制御 (強制停止/筐体ランプ 連携)
サーバ	自動割り当て	—	監視先	icmp	IP モニタ
サーバ	自動割り当て	—	NFS サーバ	icmp	NAS リソースの NFS サーバ死活 確認
サーバ	自動割り当て	—	監視先	icmp	Ping 方式ネットワークパーティ ション解決リソースの監視先

1. 自動割り当てでは、その時点で使用されていないポート番号が割り当てられます。
2. ミラーディスク、ハイブリッドディスクリソースごとに使用するポート番号です。ミラーディスクリソース、ハイブリッドディスク作成時に設定します。初期値として 29051 が設定されます。また、ミラーディスクリソース、ハイブリッドディスクの追加ごとに 1 を加えた値が自動的に設定されます。変更する場合は、Builder の[ミラーディスクリソースプロパティ]-[詳細]タブ、[ハイブリッドディスクリソースプロパティ]-[詳細]タブで設定します。詳細については『リファレンスガイド』の第 5 章、「グループリソースの詳細」を参照してください。
3. ミラーディスクリソース、ハイブリッドディスクごとに使用するポート番号です。ミラーディスクリソース、ハイブリッドディスク作成時に設定します。初期値として 29031 が設定されます。また、ミラーディスクリソース、ハイブリッドディスクの追加ごとに 1 を加えた値が自動的に設定されます。変更する場合は、Builder の[ミラーディスクリソースプロパティ]-[詳細]タブ、[ハイブリッドディスクリソースプロパティ]-[詳細]タブで設定します。詳細については『リファレンスガイド』の第 5 章、「グループリソースの詳細」を参照してください。
4. ミラーディスクリソース、ハイブリッドディスクごとに使用するポート番号です。ミラーディスクリソース、ハイブリッドディスク作成時に設定します。初期値として 29071 が設定されます。また、ミラーディスクリソース、ハイブリッドディスクの追加ごとに 1 を加えた値が自動的に設定されます。変更する場合は、Builder の[ミラーディスクリソースプロパティ]-[詳細]タブ、[ハイブリッドディスクリソースプロパティ]-[詳細]タブで設定します。詳細については『リファレンスガイド』の第 5 章、「グループリソースの詳細」を参照してください。
5. クラスタプロパティ、ポート番号(ログ)タブでログの通信方法に[UDP]を選択し、ポート番号で設定したポート番号を使用します。デフォルトのログの通信方法 [UNIX ドメイン]では通信ポートは使用しません。

通信ポート番号の自動割り当て範囲の変更

- ◆ OSが管理している通信ポート番号の自動割り当ての範囲とCLUSTERPROが使用する通信ポート番号と重複する場合があります。
- ◆ 通信ポート番号の自動割り当ての範囲とCLUSTERPROが使用する通信ポート番号が重複する場合には、重複しないようにOSの設定を変更してください。

OS の設定状態の確認例/表示例

通信ポート番号の自動割り当ての範囲はディストリビューションに依存します。

```
#cat /proc/sys/net/ipv4/ip_local_port_range
1024 65000
```

これは、アプリケーションが OS へ通信ポート番号の自動割り当てを要求した場合、1024～65000 の範囲でアサインされる状態です。

```
#cat /proc/sys/net/ipv4/ip_local_port_range
32768 61000
```

これは、アプリケーションが OS へ通信ポート番号の自動割り当てを要求した場合、32768～61000 の範囲でアサインされる状態です。

OS の設定の変更例

/etc/sysctl.conf に以下の行を追加します。(30000～65000 に変更する場合)

```
net.ipv4.ip_local_port_range = 30000 65000
```

時刻同期の設定

クラスタシステムでは、複数のサーバの時刻を定期的に同期する運用を推奨します。ntp などを使用してサーバの時刻を同期させてください。

NICデバイス名について

ifconfig コマンドの仕様により、CLUSTERPRO で動作可能な NIC デバイス名の文字列の長さに制限があります。また、フローティング IP リソースの数によって異なります。

NIC のデバイス名をデフォルト(eth0,eth1 など)から変更する場合には下記の範囲内の長さで設定してください。

bonding のデバイス名についても同様の制限があります。下記の "NIC デバイス名の文字列の長さ" の範囲内で bonding デバイスの名称を設定してください。

フローティングIPリソース、 仮想IPリソースの個数	NICデバイス名の文字列の長さ
0～10	7文字まで
11～100	6文字まで
101～	5文字まで

共有ディスクについて

- ◆ サーバの再インストール時等で共有ディスク上のデータを引き続き使用する場合は、パーティションの確保やファイルシステムの作成はしないでください。
- ◆ パーティションの確保やファイルシステムの作成をおこなうと共有ディスク上のデータは削除されます。
- ◆ 共有ディスク上のファイルシステムはCLUSTERPROが制御します。共有ディスクのファイルシステムをOSの/etc/fstabにエントリしないでください。
- ◆ 共有ディスクの設定手順は「インストールガイド&設定ガイド」を参照してください。

ミラー用のディスクについて

- ◆ ミラーディスクリソース管理用パーティション(クラスタパーティション)とミラーディスクリソースで使用するパーティション(データパーティション)を設定します。
- ◆ ミラーディスク上のファイルシステムはCLUSTERPROが制御します。ミラーディスクのファイルシステムをOSの/etc/fstabにエントリしないでください。
- ◆ ミラー用ディスクの設定手順は「インストールガイド&設定ガイド」を参照してください。

ハイブリッドディスクリソース用のディスクについて

- ◆ ハイブリッドディスクリソースの管理用パーティション(クラスタパーティション)とハイブリッドディスクリソースで使用するパーティション(データパーティション)を設定します。
- ◆ さらにハイブリッドディスク用のディスクを共有ディスク装置で確保する場合には、共有ディスク装置を共有するサーバ間のディスクハートビートリソース用のパーティションを確保します。
- ◆ ハイブリッドディスク上のファイルシステムはCLUSTERPROが制御します。ハイブリッドディスクのファイルシステムをOSの/etc/fstabにエントリしないでください。
- ◆ ハイブリッドディスク用ディスクの設定手順は「インストールガイド&設定ガイド」を参照してください。

OS起動時間の調整

電源が投入されてから、OS が起動するまでの時間が、下記の 2 つの時間より長くなるように調整してください。

- ◆ 共有ディスクを使用する場合に、ディスクの電源が投入されてから使用可能になるまでの時間
- ◆ ハートビートタイムアウト時間

設定手順は「インストールガイド&設定ガイド」を参照してください。

ネットワークの確認

- ◆ インタコネクトやミラーディスクコネクトで使用するネットワークの確認をします。クラスタ内のすべてのサーバで確認します。
- ◆ 設定手順は「インストールガイド&設定ガイド」を参照してください。

ipmiutil, OpenIPMIについて

- ◆ 以下の機能でipmiutilまたはOpenIPMIを使用します。
 - グループリソースの活性異常時/非活性異常時の最終アクション
 - モニタリソースの異常時アクション
 - ユーザ空間モニタリソース
 - シャットダウンストール監視
 - 強制停止機能
 - 筐体 ID ランプ連携
- ◆ CLUSTERPROにipmiutil、OpenIPMIは添付しておりません。ユーザ様ご自身で別途ipmiutilまたはOpenIPMI の rpm ファイルをインストールしてください。
- ◆ ipmiutil、OpenIPMIに関する以下の事項について、弊社は対応いたしません。ユーザ様の判断、責任にてご使用ください。
 - ipmiutil、OpenIPMI 自体に関するお問い合わせ

- ipmiutil、OpenIPMI の動作保証
 - ipmiutil、OpenIPMI の不具合対応、不具合が原因の障害
 - 各サーバの ipmiutil、OpenIPMI の対応状況のお問い合わせ
-
- ◆ ご使用予定のサーバ(ハードウェア)のipmiutil、OpenIPMI対応可否についてはユーザー様にて事前に確認ください。
 - ◆ ハードウェアとしてIPMI規格に準拠している場合でも実際にはipmiutil、OpenIPMIが動作しない場合がありますので、ご注意ください。
 - ◆ サーバベンダが提供するサーバ監視ソフトウェアを使用する場合には ユーザ空間モニタリソースとシャットダウンストール監視の監視方法にIPMIを選択しないでください。
これらのサーバ監視ソフトウェアとipmiutilは共にサーバ上のBMC(Baseboard Management Controller)を使用するため競合が発生して正しく監視が行うことができません。

ユーザ空間モニタリソース(監視方法softdog)について

- ◆ 監視方法にsoftdogを設定する場合、OS標準添付のheartbeatを動作しない設定にしてください。
- ◆ SUSE LINUX 10では監視方法にsoftdogを設定する場合、i8xx_tcoドライバと同時に使用することができません。i8xx_tcoドライバを使用しない場合は、i8xx_tcoをロードしない設定にしてください。

ログ収集について

- ◆ SUSE LINUX 10ではCLUSTERPROのログ収集機能でOSのsyslogを採取する場合、ロテートされたsyslog(message)ファイルのサフィックスが異なるためsyslogの世代の指定機能が動作しません。
ログ収集機能のsyslogの世代の指定を行うためにはsyslogのロテートの設定を下記のように変更して運用する必要があります。
- /etc/logrotate.d/syslog ファイルの compress と dateext をコメントアウトする

CLUSTERPRO の情報作成時

CLUSTERPRO の構成情報の設計、作成前にシステムの構成に依存して確認、留意が必要な事項です。

サーバのリセット、パニック、パワーオフ

CLUSTERPRO が「サーバのリセット」または「サーバのパニック」、または「サーバのパワーオフ」を行う場合、サーバが正常にシャットダウンされません。そのため下記のリスクがあります。

- ◆ マウント中のファイルシステムへのダメージ
- ◆ 保存していないデータの消失
- ◆ OSのダンプ採取の中断

「サーバのリセット」または「サーバのパニック」が発生する設定は下記です。

- ◆ グループリソース活性時/非活性時異常時の動作
 - sysrq パニック
 - keepalive リセット
 - keepalive パニック
 - BMC リセット
 - BMC パワーオフ
 - BMC サイクル
 - BMC NMI
- ◆ モニタリソース異常検出時の最終動作
 - sysrq パニック
 - keepalive リセット
 - keepalive パニック
 - BMC リセット
 - BMC パワーオフ
 - BMC サイクル
 - BMC NMI
- ◆ ユーザ空間監視のタイムアウト検出時動作
 - 監視方法 softdog
 - 監視方法 ipmi
 - 監視方法 keepalive

注: 「サーバのパニック」は監視方法が keepalive の場合のみ設定可能です。

- ◆ シャットダウンストール監視
 - 監視方法 softdog
 - 監視方法 ipmi
 - 監視方法 keepalive

注: 「サーバのパニック」は監視方法が keepalive の場合のみ設定可能です。

- ◆ 強制停止機能の動作
 - BMC リセット
 - BMC パワーオフ
 - BMC サイクル
 - BMC NMI

グループリソースの非活性異常時の最終アクション

非活性異常検出時の最終動作に「何もしない」を選択すると、グループが非活性失敗のまま停止しません。
本番環境では「何もしない」は設定しないように注意してください。

execリソースから起動されるアプリケーションのスタックサイズについて

- ◆ スタックサイズが 2MB に設定された状態で exec リソースが実行されます。このため、exec リソースから起動されるアプリケーションで2MB 以上のスタックサイズが必要な場合には、スタックオーバーフローが発生します。
スタックオーバーフローが発生する場合には、アプリケーションを起動する前にスタックサイズを設定してください。

1. 「この製品で作成したスクリプト」を使用している場合
アプリケーションを起動する前に、ulimit コマンドでスタックサイズを設定してください。
2. 「ユーザアプリケーション」を使用している場合
「この製品で作成したスクリプト」に変更し、スクリプト内からアプリケーションを起動するように編集してください。
アプリケーションを起動する前に、ulimit コマンドでスタックサイズを設定してください。

- 開始スクリプトの編集例

```
-----
#!/bin/sh
#*****
#*          start.sh          *
#*****

ulimit -s unlimited  # スタックサイズ変更(無制限)

"実行するアプリケーション"
-----
```

- ◆ execリソースのスクリプトを変更する場合は、『リファレンスガイド セクションⅡ』の「第5章 グループリソースの詳細 execリソースを理解する」を参照してください。

VxVMが使用するRAWデバイスの確認

ボリューム RAW デバイスの実 RAW デバイスについて事前に調べておいてください。

1. CLUSTERPRO をインストールする前に、片サーバで活性化する全てのディスクグループをインポートし、全てのボリュームを起動した状態にします。
2. 以下のコマンドを実行します。

```
# raw -qa
/dev/raw/raw2: bound to major 199, minor 2
/dev/raw/raw3: bound to major 199, minor 3
```

①②

例)ディスクグループ名、ボリューム名がそれぞれ以下の場合

- ディスクグループ名 dg1
- dg1 配下のボリューム名 vol1、vol2

3. 以下のコマンドを実行します。

```
# ls -l /dev/vx/dsk/dg1/
brw----- 1 root root 199, 2 5月 15 22:13 vol1
brw----- 1 root root 199, 3 5月 15 22:13 vol2
```

③

4. ②と③のメジャー/マイナ番号が等しいことを確認します。

これにより確認された RAW デバイス①は CLUSTERPRO のディスクハートビートリソース、RAW リソース、RAW モニタリソースでは絶対に設定しないでください。

ミラーディスクリソース、ハイブリッドディスクリソースの復帰回数制限について

- ◆ ミラーディスクを切り離れた状態からミラー復帰を行う際、復帰中にミラーディスクの更新があった場合にはリアルタイムに同期されず差分情報として蓄積されます。ミラー復帰が完了し、なお差分情報がある場合に再度ミラー復帰を行います。復帰回数制限の回数分のミラー復帰の繰り返しを行います。ミラー復帰の繰り返し回数が復帰回数制限の値を超えた場合、ミラー復帰が正常に終了せず、ミラーディスクの状態が正常になりません。ミラーディスクを切り離れた状態からミラー復帰を行う場合、復帰中のミラーディスクの更新を反映するには、復帰回数制限を しない に設定してください。
- ◆ ミラーディスクリソースの復帰回数制限を変更する場合は、『リファレンスガイド セクション II』の「第3章 Builderの機能 クラスタ - ミラーエージェントとタブ」を参照してください。

ミラーディスクのファイルシステムの選択について

現在動作確認を完了しているファイルシステムは下記の通りです。

- ◆ ext3
- ◆ xfs
- ◆ reiserfs
- ◆ jfs
- ◆ vxfs

ハイブリッドディスクのファイルシステムの選択について

現在動作確認を完了しているファイルシステムは下記の通りです。

- ◆ ext3
- ◆ reiserfs

RAWモニタリソースについて

- ◆ RAWモニタリソースを設定する場合、既にmountしているパーティションまたはmountする可能性のあるパーティションの監視はできません。また、既にmountしているパーティションまたはmountする可能性のあるパーティションのwhole device(ディスク全体を示すデバイス)をデバイス名に設定して監視することもできません。
- ◆ 監視専用のパーティションを用意してRAWモニタリソースに設定してください。

遅延警告割合

遅延警告割合を 0 または、100 に設定すれば以下のようなことを行うことが可能です。

- ◆ 遅延警告割合に0を設定した場合
監視毎に遅延警告がアラート通報されます。
この機能を利用し、サーバが高負荷状態での監視リソースへのポーリング時間を算出し、監視リソースの監視タイムアウト時間を決定することができます。
- ◆ 遅延警告割合に100を設定した場合
遅延警告の通報を行いません。

テスト運用以外で、0%等の低い値を設定しないように注意してください。

ディスクモニタリソースの監視方法TURについて

- ◆ SCSIのTest Unit ReadyコマンドやSG_IOコマンドをサポートしていないディスク、ディスクインターフェイス(HBA)では使用できません。
ハードウェアがサポートしている場合でもドライバがサポートしていない場合があるのでドライバの仕様も合わせて確認してください。
- ◆ S-ATAインターフェイスのディスクの場合には、ディスクコントローラのタイプや使用するディストリビューションにより、OSにIDEインターフェイスのディスク(hd)として認識される場合とSCSIインターフェイスのディスク(sd)として認識される場合があります。
IDEインターフェイスとして認識される場合には、すべてのTUR方式は使用できません。

SCSI インターフェイスとして認識される場合には、TUR(legacy) が使用できます。
TUR(generic)は使用できません。

- ◆ Read方式に比べてOSやディスクへの負荷は小さくなります。
- ◆ Test Unit Readyでは、実際のメディアへのI/Oエラーは検出できない場合があります。

WebManagerの画面更新間隔について

- ◆ WebManagerタブの「画面データ更新インターバル」には、基本的に30秒より小さい値を設定しないでください。

LANハートビートの設定について

- ◆ LANハートビートリソースは最低一つ設定する必要があります。
- ◆ インタコネクト専用のLANをLANハートビートリソースとして登録し、さらにパブリックLANもLANハートビートリソースとして登録することを推奨します(LANハートビートリソースを二つ以上設定することを推奨します)。

カーネルモードLANハートビートの設定について

- ◆ カーネルモードLANハートビートが使用できるディストリビューション、カーネルの場合にはLANハートビートとカーネルモードLANハートビートの併用を推奨します。
- ◆ インタコネクト専用のLANをLANハートビートリソース、カーネルモードLANハートビートリソースとして登録し、さらにパブリックLANもLANハートビートリソース、カーネルモードLANハートビートリソースとして登録することを推奨します(LANハートビートリソースとカーネルモードLANハートビートリソースを二つ以上設定することを推奨します)。

COMハートビートの設定について

- ◆ ネットワークが断線した場合に両系で活性することを防ぐため、COMが使用できる環境であればCOMハートビートリソースを使用することを推奨します。

CLUSTERPRO運用後

クラスタとして運用を開始した後に発生する事象で留意して頂きたい事項です。

hotplugサービスについて

hotplug サービスがデバイスをサーチするときに以下のログが messages ファイルにエントリされます。

```
kernel: <liscal liscal_make_request> NMP0 I/O port is close,
mount(0), io(0).
kernel: Buffer I/O error on device NMP1, logical block 0
```

hotplug サービスが起動する時点でミラーリソースが起動していないため、この現象が発生します。この現象は異常ではありません。

hotplug を使用しない設定に変更して、coldplug で運用する場合には本現象は発生しません。

udev環境でのミラードライバロード時のエラーメッセージについて

udev 環境でミラードライバのロード時に、以下のログが messages ファイルにエントリされることがあります。

```
kernel: <liscal liscal_make_request> NMP1 device does not exist.
kernel: Buffer I/O error on device NMP1, logical block 112
```

この現象は異常ではありません。

このエラーメッセージの出力を回避したい場合は/etc/udev/rules.d/配下に下記の設定ファイルを追加してください。

ファイル名: 50-liscal-udev.rules

```
ACTION=="add", DEVPATH==" /block/NMP*", OPTIONS+="ignore_device"
```

X-Window上のファイル操作ユーティリティについて

X-Window 上で動作する一部のファイル操作ユーティリティ(GUI でファイルやディレクトリのコピーや移動などの操作を行うもの)に以下の挙動をするものがあります。

- ◆ ブロックデバイスが使用可能であるかサーチする
- ◆ サーチの結果、マウントが可能なファイルシステムがあればマウントする

上記のような仕様のファイル操作ユーティリティは使用しないでください。

上記のような動作は CLUSTERPRO の動作に支障が発生する可能性があります。

ドライバロード時のメッセージについて

ミラードライバを load した際に、以下のメッセージがコンソール、syslog に表示されることがあります。この現象は異常ではありません。

```
kernel: liscal: no version for "xxxxx" found: kernel tainted.  
kernel: liscal: module license 'unspecified' taints kernel.
```

(xxxxx には任意の文字列が入ります)

clpka ドライバ, clpkhb ドライバを load した際に、以下のメッセージがコンソール、syslog に表示されることがあります。この現象は異常ではありません。

```
kernel: clpkhb: no version for "xxxxx" found: kernel tainted.  
kernel: clpkhb: module license 'unspecified' taints kernel.
```

```
kernel: clpka: no version for "xxxxx" found: kernel tainted.  
kernel: clpka: module license 'unspecified' taints kernel.
```

(xxxxx には任意の文字列が入ります)

ipmiのメッセージについて

ユーザ空間モニタリソースに IPMI を使用する場合、syslog に下記の kernel モジュール警告ログが多数出力されます。

```
modprobe: modprobe: Can't locate module char-major-10-173
```

このログ出力を回避したい場合は、/dev/ipmikcs を rename してください。

複数のミラーディスクリソース、ハイブリッドディスクリソースを使用している場合の活性時のsyslogのメッセージについて

2 つ以上のミラーディスクリソース、ハイブリッドディスクリソースを設定している場合、ミラーリソース、ハイブリッドディスクリソースの活性時に OS の messages ファイルに以下のメッセージがエントリされることがあります。

この現象は一部のディストリビューションの fsck コマンドの挙動(本来、fsck の対象でないブロックデバイスへアクセスをする挙動)によるものです。

```
kernel: <liscal liscal_make_request> NMPx I/O port is close,
mount(0), io(0).
kernel: Buffer I/O error on device /dev/NMPx, logical block xxxx
```

CLUSTERPRO としては問題はありません。messages ファイルを圧迫するなどの問題がある場合にはミラーディスクリソース、ハイブリッドディスクリソースの以下の設定を変更してください。

Mount 実行前の fsck アクションを「実行しない」

Mount 失敗時の fsck アクションを「実行する」

回復動作中の操作制限

モニタリソースの異常検出時の設定で回復対象にグループリソース(ディスクリソース、EXEC リソース、...)を指定し、モニタリソースが異常を検出した場合の回復動作遷移中(再活性化 → フェイルオーバー → 最終動作)には、以下のコマンドまたは、WebManager からのクラスタ及びグループへの制御は行わないでください。

- ◆ クラスタの停止 / サスペンド
- ◆ グループの開始 / 停止 / 移動

モニタリソース異常による回復動作遷移中に上記の制御を行うと、そのグループの他のグループリソースが停止しないことがあります。

また、モニタリソース異常状態であっても最終動作実行後であれば上記制御を行うことが可能です。

コマンド編に記載されていない実行形式ファイルやスクリプトファイルについて

インストールディレクトリ配下にコマンド編に記載されていない実行形式ファイルやスクリプトファイルがありますが、CLUSTERPRO 以外からは実行しないでください。

実行した場合の影響については、サポート対象外とします。

kernel ページアロケートエラーのメッセージについて

TurboLinux 10 Server で Replicator を使用する場合に、syslog に以下のメッセージが出力されることがあります。ただし、物理メモリサイズや I/O 負荷に依存するため出力されない場合もあります。

```
kernel: [kernel モジュール名]: page allocation failure. order:X,
mode:0xXX
```

このメッセージが出力される場合には、下記のカーネルパラメータを変更する必要があります。sysctl コマンド等を使用して OS 起動時にパラメータが変更されるように設定してください。

```
/proc/sys/vm/min_free_kbytes
```

min_free_kbytes に設定可能な最大値は、サーバに搭載されている物理メモリサイズによって異なります。下記の表を参照して設定してください。

物理メモリサイズ(Mbyte)	最大値
1024	1024
2048	1448
4096	2048
8192	2896
16384	4096

ログ収集時のメッセージ

ログ収集を実行した場合、コンソールに以下のメッセージが表示されることがありますが、異常ではありません。ログは正常に収集されています。

```
hd#: bad special flag: 0x03
ip_tables: (C) 2000-2002 Netfilter core team
```

(hd#にはサーバ上に存在する IDE のデバイス名が入ります)

クラスタシャットダウン・クラスタシャットダウンリブート

ミラーディスクリソース、ハイブリッドディスクリソース使用時は、グループ活性処理中に `clpstdn` コマンドまたは WebManager からクラスタシャットダウン、クラスタシャットダウンリブートを実行しないでください。

グループ活性処理中はグループ非活性ができません。このため、ミラーディスクリソース、ハイブリッドディスクリソースが正常に非活性されていない状態で OS がシャットダウンされ、ミラーブレイクが発生することがあります

特定サーバのシャットダウン、リブート

ミラーディスクリソース、ハイブリッドディスクリソース使用時は、グループ活性処理中に `clpdown` コマンドまたは WebManager からサーバのシャットダウン、シャットダウンリブートコマンドを実行しないでください。

グループ活性処理中はグループ非活性ができません。このため、ミラーディスクリソース、ハイブリッドディスクリソースが正常に非活性されていない状態で OS がシャットダウンされ、ミラーブレイクが発生することがあります。

サービス起動／停止用スクリプトについて

以下の場合に、サービスの起動／停止スクリプトでエラーが出力されます。

- ◆ CLUSTERPRO Server インストール直後 (SUSE Linux の場合)
OSシャットダウン時に下記のサービス停止スクリプトでエラーが出力されます。各サービスが起動されていないことが原因で出力されるエラーのため問題はありません。
 - `clusterpro_alertsync`
 - `clusterpro_webmgr`
 - `clusterpro`
 - `clusterpro_md`
 - `clusterpro_trn`
 - `clusterpro_evt`
- ◆ クラスタ構築前
OS起動時に下記のサービス起動スクリプトでエラーが出力されます。クラスタ未構築が原因で出力されるエラーのため問題はありません。
 - `clusterpro_md`
- ◆ クラスタ構築後 (SUSE Linux の場合)
ミラーディスクリソース、ハイブリッドディスクリソースを使用していない場合、OSシャットダウン時に下記のサービス停止スクリプトでエラーが出力されます。ミラーディスクリソース、ハイブリッドディスクリソースを使用していない場合、ミラーエージェントが起動されないことが原因で出力されるエラーのため問題はありません。
 - `clusterpro_md`

- ◆ サービスの手動停止後のOSシャットダウン(SUSE Linux の場合)
clpclコマンドやWebManagerからサービスを停止後、OSシャットダウン時に停止したサービスの停止スクリプトでエラーが出力されます。サービスが停止していることが原因で出力されるエラーのため問題はありません。
 - clusterpro
 - clusterpro_md

以下の場合に、サービスの停止スクリプトが不正な順序で実行されます。

- ◆ `chkconfig --del name` を実行し全サービスを無効化した後のOSシャットダウン
CLUSTERPROのサービスを無効化した後、OSシャットダウン時にCLUSTERPROのサービスが不正な順序で停止されます。OSシャットダウン時に無効化したCLUSTERPROのサービスが停止されないことが原因で発生します。
WebManager から実行するクラスタシャットダウンや、`clpstdn` コマンドなどCLUSTERPROのコマンドを使用してのクラスタシャットダウンの場合は不正な順序で停止されても問題ありません。

EXECリソースで使用するスクリプトファイルについて

EXEC リソースで使用するスクリプトファイルは各サーバ上の下記のディレクトリに配置されます。

`/インストールパス/scripts/グループ名/EXEC リソース名/`

クラスタ構成変更時に下記の変更を行った場合、変更前のスクリプトファイルはサーバ上からは削除されません。

- EXEC リソースを削除した場合や EXEC リソース名を変更した場合
- EXEC リソースが所属するグループを削除した場合やグループ名を変更した場合

変更前のスクリプトファイルが必要ない場合は、削除しても問題ありません。

活性時監視設定のモニタリソースについて

活性時監視設定のモニタリソースの一時停止／再開には下記の制限事項があります。

- ◆ モニタリソースを一時停止後、監視対象リソースを停止させた場合
モニタリソースごとに監視再開時の動作が異なります。
 - PID モニタリソース
監視を再開しても監視対象リソースが停止しているため監視できません。
 - PID モニタリソース以外
監視を再開すると監視を継続します。
- ◆ モニタリソースを一時停止後、監視対象リソースを他サーバへ移動させた場合
モニタリソースごとに、監視再開時の動作が異なります。
 - PID モニタリソース(移動元サーバ)
監視を再開しても監視対象リソースが停止しているため監視できません。
 - PID モニタリソース以外(移動元サーバ)
モニタリソースを再開すると監視を継続します。
 - PID モニタリソース(移動先サーバ)
監視対象リソースが活性化されたタイミングで監視を開始します。移動元のサーバでの監視再開操作は影響しません。
 - PID モニタリソース以外(移動先サーバ)
監視対象リソースが活性化されたタイミングで監視を開始します。移動元のサーバでの監視再開操作は影響しません。

活性時監視設定のモニタリソースの回復対象にクラスタが設定されている場合には下記の制限事項があります。

- ◆ モニタリソースを一時停止後、監視対象リソースが停止した場合
モニタリソースごとに監視再開後の異常検出時の動作が異なります。
 - PID モニタリソース
異常を検出しません。
 - PID モニタリソース以外
異常を検出した場合、設定されたアクションを実行します。
- ◆ モニタリソースを一時停止後、監視対象リソースが他サーバへ移動した場合
モニタリソースごとに監視再開後の異常検出時の動作が異なります。
 - PID モニタリソース(移動元サーバ)
異常を検出しません。
 - PID モニタリソース以外(移動元サーバ)
異常を検出した場合、設定されたアクションを実行します。

WebManagerについて

- ◆ WebManagerで表示される内容は必ずしも最新の状態を示しているわけではありません。最新の情報を取得したい場合、[リロード]ボタンを選択して最新の情報を取得してください。
- ◆ WebManagerが情報を取得中にサーバダウン等発生すると、情報の取得に失敗し、一部オブジェクトが正しく表示できない場合があります。次の自動更新まで待つか、[リロード]ボタンを選択して最新の情報を再取得してください。
- ◆ Linux上のブラウザを利用する場合、ウィンドウマネージャの組み合わせによっては、ダイアログが背後に回ってしまう場合があります。[ALT]+[TAB]キーなどでウィンドウを切り替えてください。
- ◆ CLUSTERPROのログ収集は複数のWebManagerから同時に実行することはできません。
- ◆ 接続先と通信できない状態で操作を行うと、制御が戻ってくるまでしばらく時間が必要な場合があります。
- ◆ マウスポインタが処理中を表す、腕時計や砂時計になっている状態で、ブラウザ外にカーソルを移動すると、処理中であってもカーソルが矢印の状態にもどってしまうことがあります。
- ◆ Proxyサーバを経由する場合は、WebManagerのポート番号を中継できるように、Proxyサーバの設定をしてください。
- ◆ CLUSTERPROのアップデートを行なった場合、ブラウザを終了してください。Javaのキャッシュをクリアしてブラウザを再起動してください。

Builder について

- ◆ 以下の製品とはクラスタ構成情報の互換性がありません。
 - CLUSTERPRO X 2.0 for Linux 以外の Linux 版のトレッキングツール
- ◆ Webブラウザを終了すると(メニューの[終了]やウィンドウフレームの[X]ボタン等)、現在の編集内容が破棄されます。構成を変更した場合でも保存の確認ダイアログが表示されません。
編集内容の保存が必要な場合は、終了する前に、Builder のメニューバー[ファイル]-[情報ファイルの保存]を行ってください。
- ◆ Webブラウザをリロードすると(メニューの[最新の情報に更新]やツールバーの[現在のページを再読み込み]ボタン等)、現在の編集内容が破棄されます。構成を変更した場合でも保存の確認ダイアログが表示されません。
編集内容の保存が必要な場合は、リロードする前に、Builder のメニューバー[ファイル]-[情報ファイルの保存]を行ってください。
- ◆ Builderでのクラスタ構成情報作成時には下記の点に注意してください。
 - 数値を入力するテキストボックス
0 で始まる数値は入力しないでください。
例えば、タイムアウトに 10 秒を設定する場合には「010」ではなく、「10」を入力してください。

ミラーディスク、ハイブリッドディスクリソースについて

- ◆ 運用を開始した後で、ミラーパーティションのサイズを変更したい場合は『リファレンスガイド』の「第 10 章 保守情報 ミラーリソースのパーティションのオフセットやサイズを変更する」を参照してください。

第 6 章 アップデート手順

本章では、CLUSTERPRO のアップデート手順について説明します。

本章で説明する項目は以下の通りです。

- CLUSTERPRO X 1.0 からのアップデート手順..... 108

CLUSTERPRO X 1.0 からのアップデート手順

サーバ rpm のバージョン 1.0.1-1～1.0.2-1 から 2.0.0-1 以降へのアップデートには下記の手順を実行してください。

CLUSTERPRO Server RPMのアップデート

CLUSTERPRO Server RPM は root ユーザでインストールしてください。次の手順に従って、サーバ RPM をすべてのサーバでインストールしてください。

1. 全サーバで `chkconfig --del name` を実行して以下の順序でサービスを無効にします。 *name* には以下のサービスを指定します。
 - clusterpro_alertsync
 - clusterpro_webmgr
 - clusterpro
 - clusterpro_md
 - clusterpro_trn
 - clusterpro_evt
2. Web マネージャまたは `clpstdn` コマンドを使用してクラスタをシャットダウン、リブートしてください。
3. インストール CD-ROM の媒体を mount します。
4. CLUSTERPRO のサービスが起動していないことを確認してから、rpm コマンドを実行してパッケージファイルをインストールします。
アーキテクチャによりインストール用 RPM が異なります。

CD-ROM 内の `/Linux/2.0/jp/server` に移動して、

```
rpm -U clusterpro-<バージョン>.<アーキテクチャ>.rpm
```

を実行します。

アーキテクチャには i686、x86_64、ia64、ppc64 があります。インストール先の環境に応じて選択してください。アーキテクチャは、`arch` コマンドなどで確認できます。

CLUSTERPRO は以下の場所にインストールされます。このディレクトリを変更するとアンインストールできなくなりますので注意してください。

インストールディレクトリ: `/opt/nec/clusterpro`

5. インストール終了後、インストール CD-ROM 媒体を `umount` し、インストール CD-ROM 媒体を取り除きます。
6. `chkconfig --add name` を実行して以下の順序でサービスを有効にします。 *name* には以下のサービスを指定します。SUSE Linux では `--force` オプションをつけて実行してください。
 - clusterpro_evt
 - clusterpro_trn

- clusterpro_webmgr
 - clusterpro_alertsync
7. 3～6 の手順を全てのサーバで実行します。
 8. クラスタを構成している全てのサーバを再起動します。
 9. ライセンス登録を行います。ライセンス登録の詳細は『インストール&設定ガイド』の「第 4 章 ライセンスを登録する」を参照してください。
 10. クラスタを構成している 1 台のサーバに WebManager を接続します。
 11. 接続した WebManager からオンライン Builder を起動します。
オンライン Builder の起動方法は『インストール&設定ガイド』を参照してください。
 12. クラスタプロパティの[情報]タブを開いて、クラスタで使用する言語を設定します。言語設定の詳細は『リファレンスガイド』の「第 3 章 Builder の機能」を参照してください。
 13. クラスタを構成している全てのサーバが起動していることを確認して、オンライン Builder から構成情報のアップロードを実行します。オンライン Builder の操作方法は「リファレンスガイド」を参照してください。
 14. `chkconfig --add name` を実行して以下の順序でサービスを有効にします。 *name* には以下のサービスを指定します。
 - clusterpro_md
 - clusterpro
 15. 14 の手順を全てのサーバで実行します。
 16. WebManager から マネージャ再起動 を実行します。
 17. WebManager から ミラーエージェント開始 を実行します。
 18. WebManager を接続しているブラウザを再起動します。
 19. WebManager から クラスタ開始 を実行します。

サーバ rpm のバージョン 1.1.0-1 以降から 2.0.0-1 以降へのアップデートには下記の手順を実行してください。

CLUSTERPRO Server RPMのアップデート

CLUSTERPRO Server RPM は root ユーザでインストールしてください。次の手順に従って、サーバ RPM をすべてのサーバでインストールしてください。

1. 全サーバで `chkconfig --del name` を実行して以下の順序でサービスを無効にします。name には以下のサービスを指定します。
 - clusterpro_alertsync
 - clusterpro_webmgr
 - clusterpro
 - clusterpro_md
 - clusterpro_trn
 - clusterpro_evt
2. Web マネージャまたは `clpstdn` コマンドを使用してクラスタをシャットダウン、リブートしてください。
3. インストール CD-ROM の媒体を mount します。
4. CLUSTERPRO のサービスが起動していないことを確認してから、rpm コマンドを実行してパッケージファイルをインストールします。
アーキテクチャによりインストール用 RPM が異なります。

CD-ROM 内の `/Linux/2.0/jp/server` に移動して、

```
rpm -U clusterpro-<バージョン>.<アーキテクチャ>.rpm
```

を実行します。

アーキテクチャには i686、x86_64、ia64、ppc64 があります。インストール先の環境に応じて選択してください。アーキテクチャは、`arch` コマンドなどで確認できます。

CLUSTERPRO は以下の場所にインストールされます。このディレクトリを変更するとアンインストールできなくなりますので注意してください。

インストールディレクトリ: `/opt/nec/clusterpro`

5. インストール終了後、インストール CD-ROM 媒体を `umount` し、インストール CD-ROM 媒体を取り除きます。
6. ライセンス登録を行います。ライセンス登録の詳細は『インストール&設定ガイド』の「第 4 章 ライセンスを登録する」を参照してください。
ライセンス登録を行った場合に以下のメッセージがコンソールに出力されます。この現象は異常ではありません。

```
Command succeeded.
```

```
But the license was not applied to all the servers in the cluster  
because there are one or more servers that are not started up.
```

7. `chkconfig --add name` を実行して以下の順序でサービスを有効にします。 *name* には以下のサービスを指定します。
 - clusterpro_evt
 - clusterpro_trn
 - clusterpro_md
 - clusterpro
 - clusterpro_webmgr
 - clusterpro_alertsync
8. 3～7 の手順を全てのサーバで実行します。
9. クラスタを構成している全てのサーバを再起動します。

付録

- 付録 A 用語集
- 付録 B 索引

付録 A 用語集

あ

インタコネクト クラスタ サーバ間の通信パス
(関連) プライベート LAN、パブリック LAN

か

仮想IPアドレス 遠隔地クラスタを構築する場合に使用するリソース (IPアドレス)

管理クライアント WebManager が起動されているマシン

起動属性 クラスタ起動時、自動的にフェイルオーバーグループを起動するか、手動で起動するかを決定するフェイルオーバーグループの属性
管理クライアントより設定が可能

共有ディスク 複数サーバよりアクセス可能なディスク

共有ディスク型クラスタ 共有ディスクを使用するクラスタシステム

切替パーティション 複数のコンピュータに接続され、切り替えながら使用可能なディスクパーティション
(関連) ディスクハートビート用パーティション

クラスタ システム 複数のコンピュータを LAN などをつないで、1 つのシステムのように振る舞わせるシステム形態

クラスタ シャットダウン クラスタシステム全体 (クラスタを構成する全サーバ) をシャットダウンさせること

クラスタパーティション ミラーディスク、ハイブリッドディスクに設定するパーティション。ミラーディスク、ハイブリッドディスクの管理に使用する。
関連(ディスクハートビート用パーティション)

現用系 ある 1 つの業務セットについて、業務が動作しているサーバ
(関連) 待機系

さ

セカンダリ (サーバ) 通常運用時、フェイルオーバーグループがフェイルオーバーする先のサーバ
(関連) プライマリ サーバ

た

待機系 現用系ではない方のサーバ
(関連) 現用系

ディスクハートビート用パーティション 共有ディスク型クラスタで、ハートビート通信に使用するためのパーティション

データパーティション 共有ディスクの切替パーティションのように使用することが可能なローカルディスク
ミラーディスク、ハイブリッドディスクに設定するデータ用のパーティション
(関連) クラスタパーティション

な

ネットワークパーティション 全てのハートビートが途切れてしまうこと
(関連) インタコネクト、ハートビート

ノード クラスタシステムでは、クラスタを構成するサーバを指す。ネットワーク用語では、データを他の機器に経由することのできる、コンピュータやルータなどの機器を指す。

は

ハートビート サーバの監視のために、サーバ間で定期的にお互いに通信を行うこと
(関連) インタコネクト、ネットワークパーティション

パブリック LAN サーバ / クライアント間通信パスのこと
(関連) インタコネクト、プライベート LAN

フェイルオーバー 障害検出により待機系が、現用系上の業務アプリケーションを引き継ぐこと

フェイルバック あるサーバで起動していた業務アプリケーションがフェイルオーバーにより他のサーバに引き継がれた後、業務アプリケーションを起動していたサーバに再び業務を戻すこと

フェイルオーバー グループ 業務を実行するのに必要なクラスタリソース、属性の集合

フェイルオーバー グループの移動	ユーザが意図的に業務アプリケーションを現用系から待機系に移動させること
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の中でのフェイルオーバー優先順位を持つ属性
プライベート LAN	クラスタを構成するサーバのみが接続された LAN (関連) インタコネクト、パブリック LAN
プライマリ (サーバ)	フェイルオーバーグループでの基準で主となるサーバ (関連) セカンダリ (サーバ)
フローティング IP アドレス	フェイルオーバーが発生したとき、クライアントのアプリケーションが接続先サーバの切り替えを意識することなく使用できる IP アドレス クラスタサーバが所属する LAN と同一のネットワーク アドレス内で、他に使用されていないホスト アドレスを割り当てる

ま

マスタサーバ	Builder の [クラスタのプロパティ]-[マスタサーバ] で先頭に表示されているサーバ
ミラーディスクコネクト	ミラーディスク、ハイブリッドディスクでデータのミラーリングを行うために使用する LAN。プライマリインタコネクトと兼用で設定することが可能。
ミラーディスクシステム	共有ディスクを使用しないクラスタシステム サーバのローカルディスクをサーバ間でミラーリングする

付録 B 索引

B

Builder, 51, 65, 74, 106

C

CLUSTERPRO, 29, 30
COMハートビート, 93, 94, 97

H

HA クラスタ, 16
hotplugサービス, 98

I

ipmiのメッセージ, 100

J

Java実行環境, 65, 67

K

kernel, 54
kernelページアロケートエラーのメッセージ, 101

L

LANハートビート, 97

N

NIC Link Up/Downモニタリソース, 79
NICデバイス名, 89

O

OS, 65, 67
OS起動時間, 90

R

RAWデバイス, 94
RAWモニタリソース, 85, 95

S

Single Point of Failure (SPOF), 15, 24

T

TUR, 95

U

udev環境でのliscalドライバロード時のエラーメッセージ, 98

W

WebManager, 51, 67, 74, 105
write性能, 80

あ

アップデート, 108, 110
アップデート手順, 108
アプリケーションの引き継ぎ, 22

い

依存するドライバ, 85
依存するライブラリ, 85

か

カーネルモードLANハートビート, 97
カーネルモードLANハートビート、キープアライブドライバ, 85
活性時監視設定のモニタリソース, 104
画面更新間隔, 97
監視できる障害とできない障害, 33

き

機能強化, 71
業務監視, 32
共有ディスク, 89
共有ディスク要件, 75

く

クラスタオブジェクト, 42
クラスタシステム, 15, 16
クラスタシャットダウン, 102
クラスタシャットダウンリポート, 102
クラスタリソースの引き継ぎ, 21
グループリソース, 43, 93

け

検出できる障害とできない障害, 33

さ

サーバ監視, 31

サーバのリセット、パニック, 92
サービス起動／停止用スクリプト, 102
最終アクション, 93

し

時刻同期, 88
システム構成, 36
実行形式ファイル, 101
修正情報, 72
障害監視, 27, 31
障害検出, 15, 20

す

スクリプトファイル, 101, 104
スペック, 52

せ

製品構成, 30

そ

ソフトウェア, 54
ソフトウェア構成, 30

ち

遅延警告割合, 95

つ

通信ポート番号, 86

て

ディスクインターフェイス, 52
ディスクサイズ, 64
ディスク容量, 65, 67
ディストリビューション, 54
データの引き継ぎ, 21

と

動作OS, 74
動作確認済アプリケーション情報, 60
特定サーバのシャットダウン, 102
特定サーバのシャットダウンレポート, 102
ドライバロード時のメッセージ, 98

な

内部監視, 32

ね

ネットワーク, 90
ネットワークインターフェイス, 53
ネットワークパーティション解決リソース, 43
ネットワークパーティション問題, 21

は

ハードウェア, 52
ハードウェア構成, 39, 40, 41
ハートビートリソース, 43
ハイブリッドディスク, 106
ハイブリッドディスク要件, 77
ハイブリッド用ディスク, 84, 90

ふ

ファイルシステム, 81, 95
ファイル操作ユーティリティ, 98
フェイルオーバー, 23, 34
フェイルオーバーリソース, 35
ブラウザ, 65, 67

み

ミラーディスク要件, 74
ミラードライバ, 85
ミラー用ディスク, 81, 89

め

メモリ容量, 64, 65, 67

も

モニタリソース, 44

ゆ

ユーザ空間モニタリソース, 91

り

リソース, 29, 43

ろ

ログ収集, 91
ログ収集時のメッセージ, 101