CLUSTERPRO® X 3.0 *for Solaris*

スタートアップガイド

2011.04.08
第3版

CLUSTERPRO

改版履歴

版数	改版日付	内 容
1	2010/10/01	新規作成
2	2011/01/25	内部バージョン3.0.2-1に対応しました。
3	2011/04/08	内部バージョン3.0.3-1に対応しました。

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、一切責任をおいません。
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。
Sun、Sun Microsystems、サンのロゴマーク、Solarisは、米国Sun Microsystems,Inc.の米国およびその他の国における商標または登録商標です。
Linuxは、Linus Torvalds氏の米国およびその他の国における、登録商標または商標です。
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。
Javaは、Sun Microsystems, Inc.の米国およびその他の国における商標または登録商標です。
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

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 4 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。

『CLUSTERPRO X 統合WebManager 管理者ガイド』(Integrated WebManager Administrator's Guide)
CLUSTERPRO を使用したクラスタシステムを CLUSTERPRO 統合WebManager で管理するシステム管理者、および統合WebManager の導入を行うシステムエンジニアを対象読者とし、統合WebManager を使用したクラスタシステム導入時に必須の事項について、実際の手順に則して詳細を説明します。

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]`
#	Solaris 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	`/Solaris/3.0/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	`pkgadd -d NECclusterpro-<`*バージョン番号*`>-<`*リリース番号*`>-x86.pkg`

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

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。そうなると、莫大な損害を引き起こすだけではなく、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。

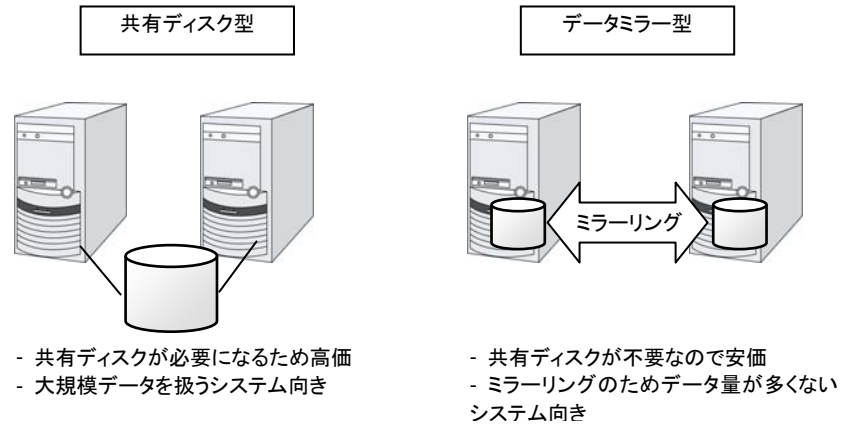


図 1-1 HAクラスタ構成図

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、Solaris の場合は fsck) 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。

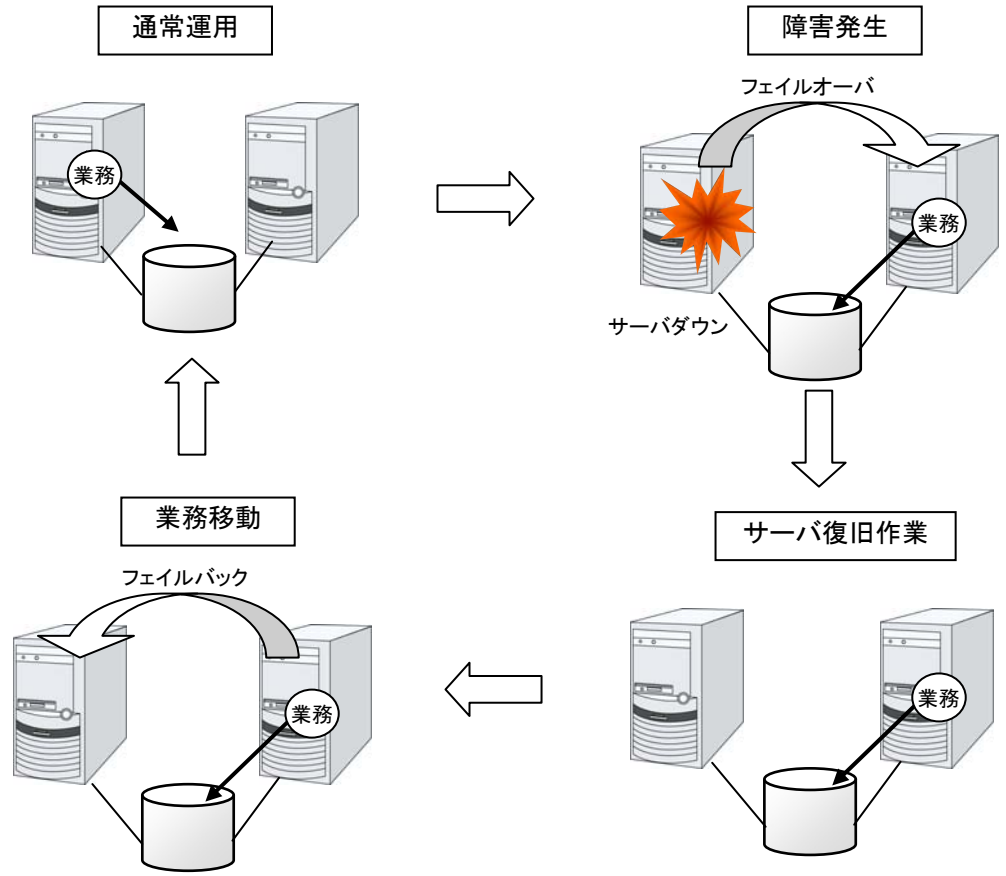


図 1-2 障害発生から復旧までの流れ

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。

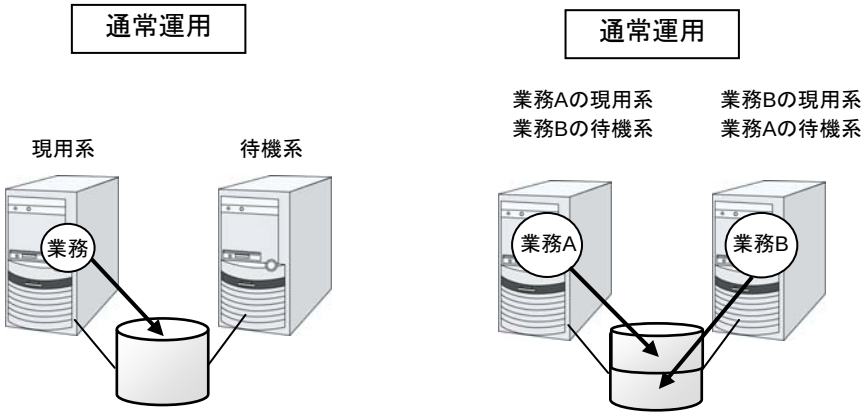


図 1-3 HA クラスタの運用形態

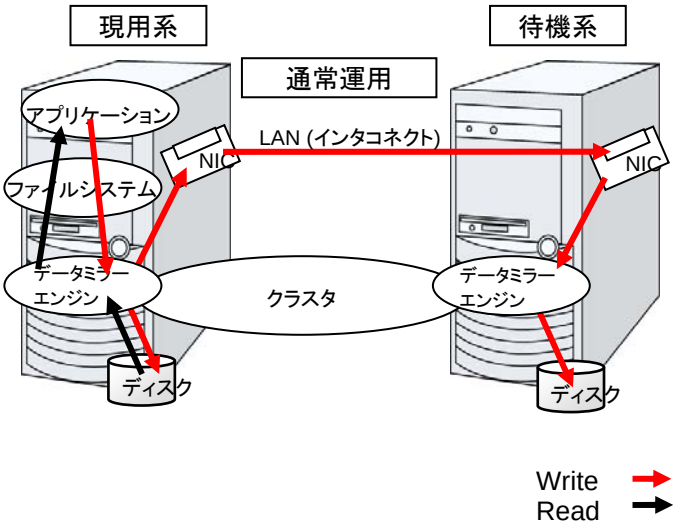
データミラー型

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。



注: CLUSTERPRO X 3.0 for Solaris ではデータミラー型のクラスタを構築することはできません。

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)を実行します。フェイルオーバ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。

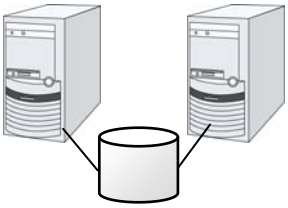


図 1-5 共有ディスクタイプのクラスタ構成

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。

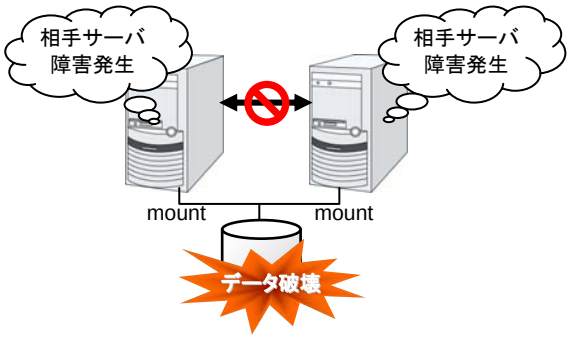


図 1-6 ネットワークパーティション症状

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。

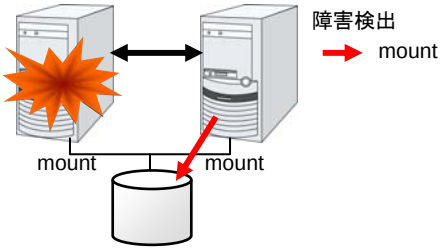


図 1-7 データの引き継ぎ

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ったかもしれません。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っ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ると思います。

- 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

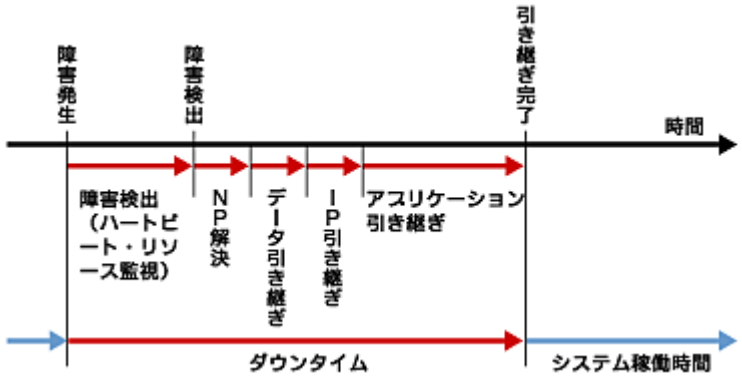


図 1-8 フェイルオーバタイムチャート

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。

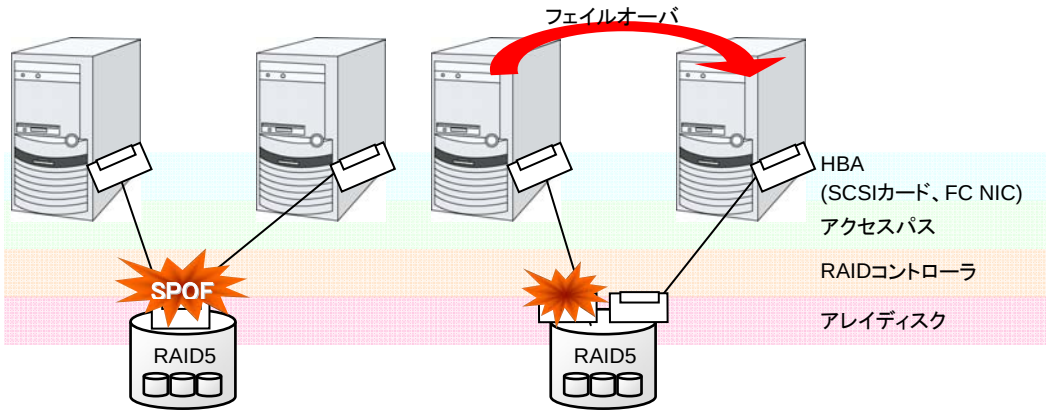


図 1-9 共有ディスクの RAID コントローラとアクセスパスが SPOF となっている例(左)と RAID コントローラとアクセスパスを分割した例

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、欠点とはいえないまでも、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ドライバなどと呼びます。このパスフェイルオーバドライバを利用し、共有ディスクへのアクセスパスを多重化することが可用性を向上させるポイントになります。

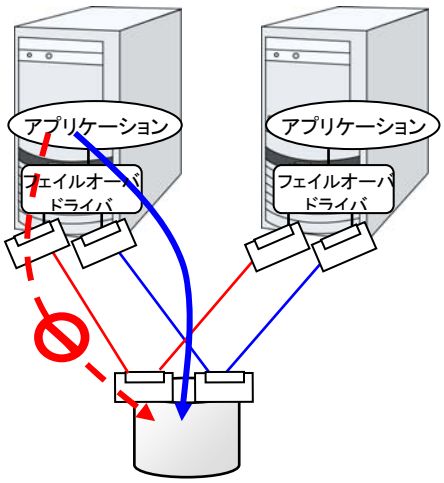


図 1-10 パスフェイルオーバドライバ

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。

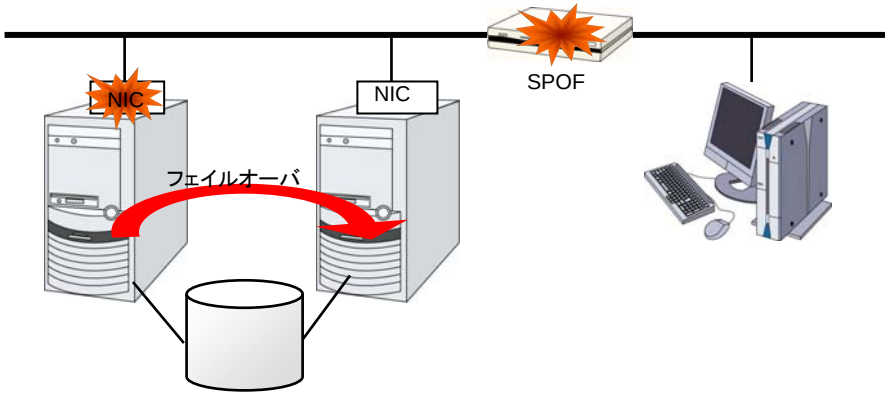


図 1-11 ルータが SPOF となる例

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という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もいわれています。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。

- 障害発生個所を洗い出し、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ると思います。

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。万がーサーバに障害が発生した場合を例に取ると、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います。

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。Solaris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?

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 を利用する端末上で、CLUSTERPRO Server とは別にインストールして利用するオフライン版と WebManager 画面のツールバーから設定モードアイコン、または[表示]メニューの[設定モード]をクリックして転換するオンライン版があります。通常インストール不要であり、オフラインで使用する場合のみ別途インストールします。

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。Solaris サーバ上には「CLUSTEPRRO Server(CLUSTERPRO 本体)」をインストールします。WebManager や Builder の本体機能は CLUSTERPRO Server に含まれるため、別途インストールする必要がありません。ただし、CLUSTERPRO Server にアクセスできない環境で Builder を使用する場合は、オフライン版の Builder を PC にインストールする必要があります。WebManager や Builder は管理 PC 上の Web ブラウザから利用するほか、クラスタを構成する各サーバ上の Web ブラウザでも利用できます。

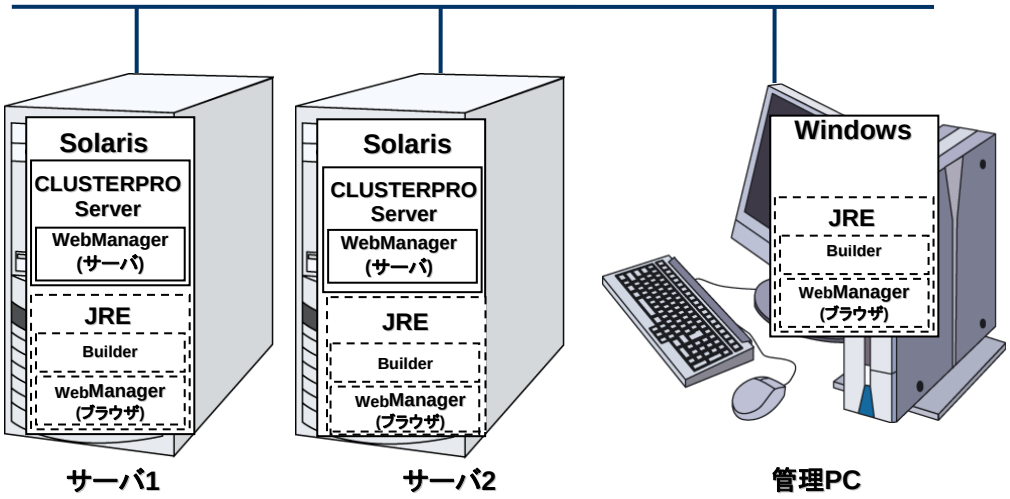


図 2-1 CLUSTERPRO のソフトウェア構成

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。

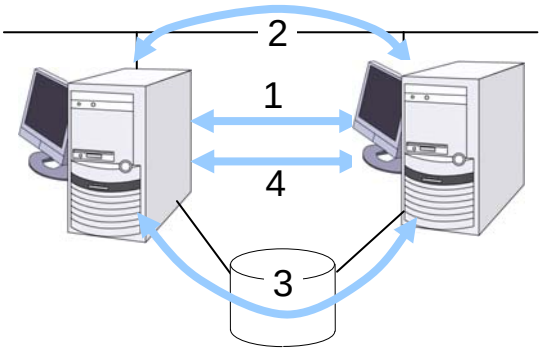
CLUSTERPRO はサーバ監視のために、定期的にサーバ同士で生存確認を行います。この生存確認をハートビートと呼びます。ハートビートは以下の通信パスを使用して行います。

- 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。



1	インタコネクト専用LAN
2	パブリックLAN
3	共有ディスク
4	COMポート

- 共有ディスク

  フェイルオーバ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(Sprit-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(HBA[1]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。

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』の「第 7 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。

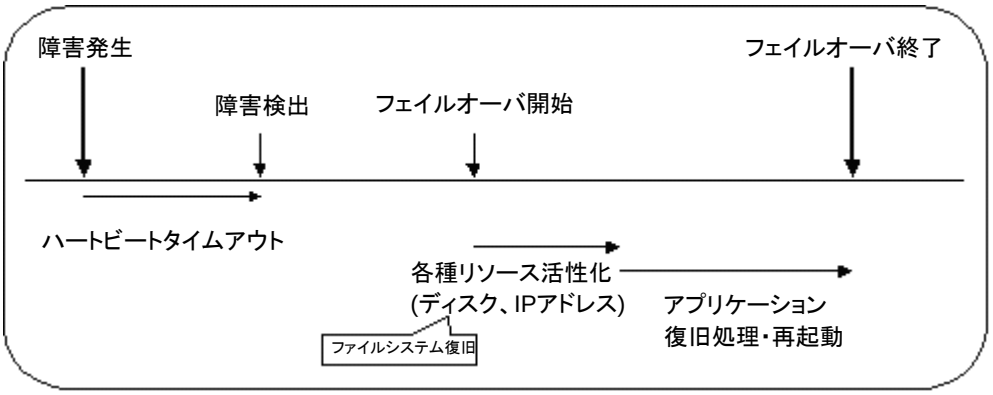


図 2-2 フェイルオーバのタイムチャート

- 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、『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。

フェイルオーバリソース

CLUSTERPRO がフェイルオーバ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。

フェイルオーバ型クラスタのシステム構成

フェイルオーバ型クラスタは、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。

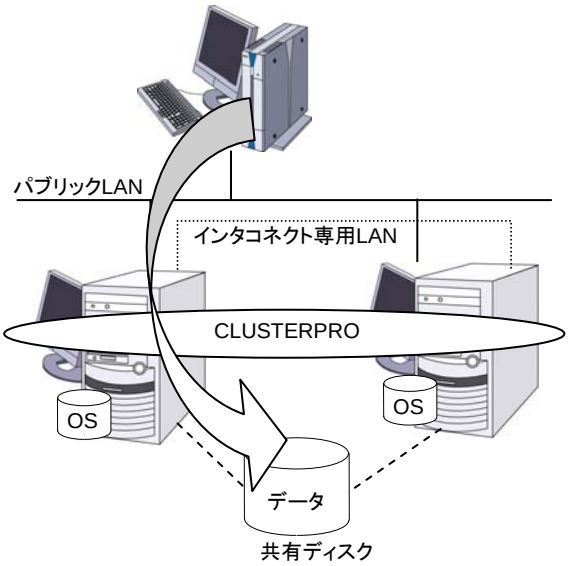


図 2-3 システム構成

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動させ、他方のサーバを待機系として業務を稼動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。

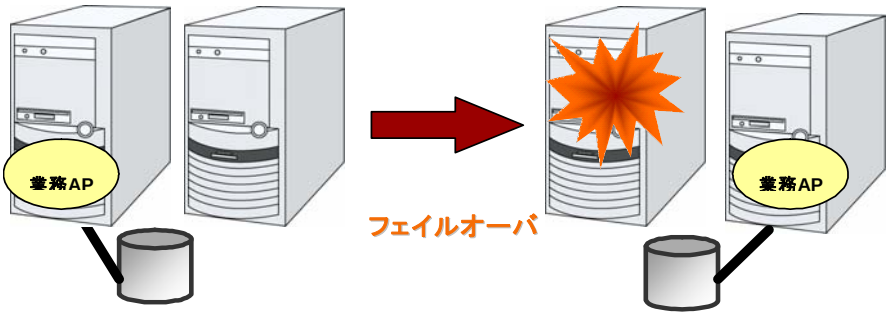
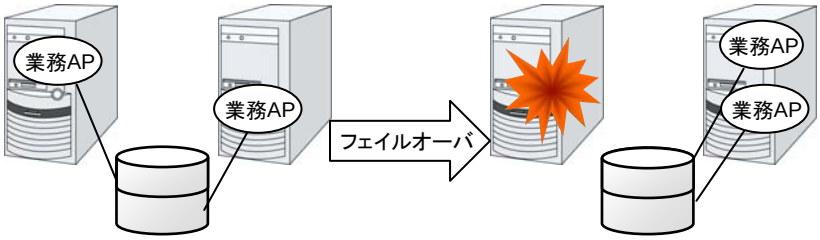


図 2-4 片方向スタンバイクラスタ

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動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。

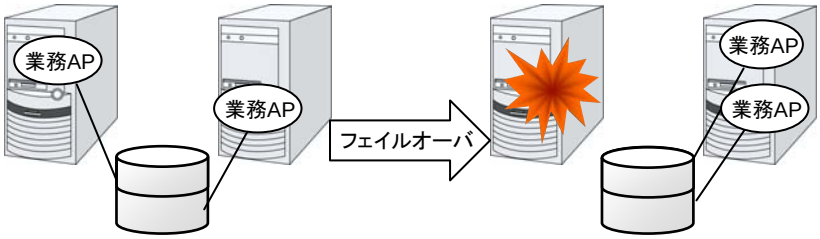


※　図の業務APは同一アプリケーション
※　フェイルオーバ後にひとつのサーバ上で複数の業務APインスタンスが動く

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。



※　業務1と業務2は異なるアプリケーションを使用

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 でしたが、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。

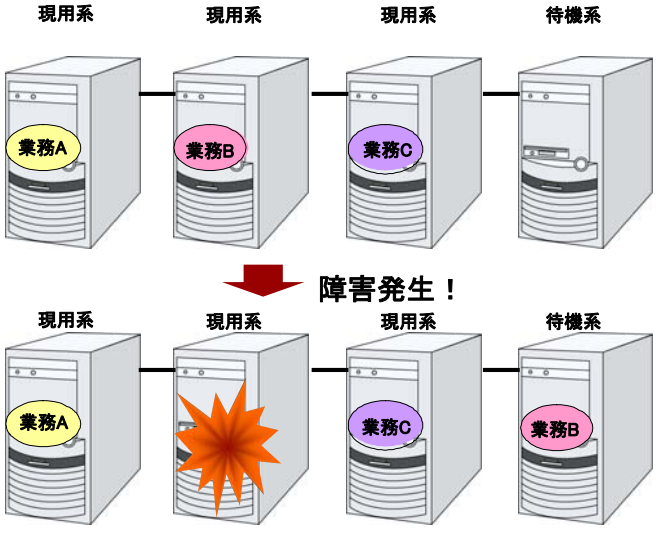


図 2-7 N + N 構成

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。

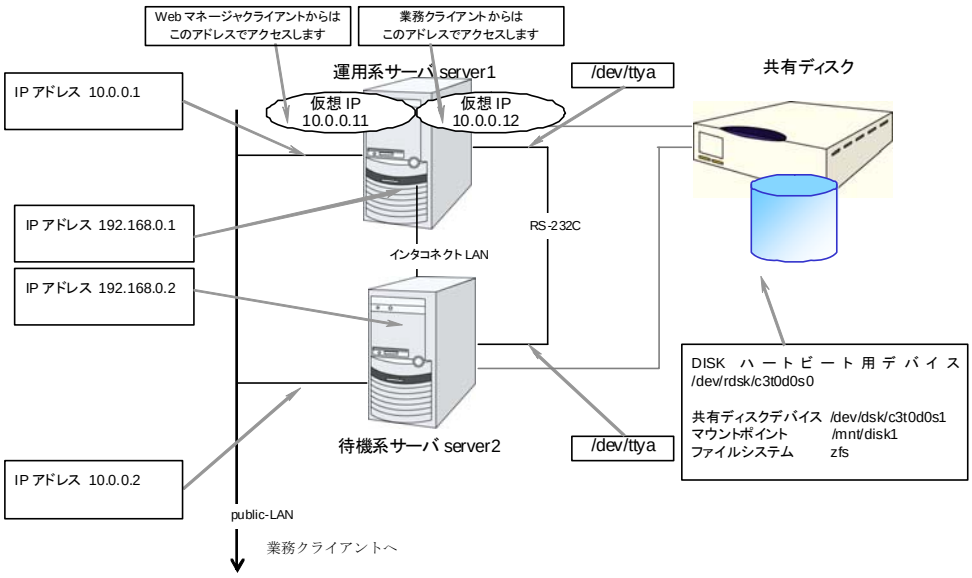


図 2-8 共有ディスク使用時のクラスタ環境のサンプル

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。
- 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。
- 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

フェイルオーバを行う際の単位となる、フェイルオーバ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。
- 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。

- ボリュームマネージャリソース (volmgr)
  ボリュームマネージャリソースは、ボリュームマネージャによって管理される論理ディスクを制御します。
- 仮想マシンリソース (vm)
  仮想マシンの起動、停止、マイグレーションを行います。
- ダイナミック DNS リソース (ddns)
  Dynamic DNS サーバに仮想ホスト名と活性サーバの IP アドレスを登録します。

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。
- 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。
- 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。
- 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。
- 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。
- 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。
- 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。
- ボリュームマネージャモニタリソース (volmgrw)
  ボリュームマネージャにより管理されている論理ディスクの監視機構を提供します。

- 仮想マシンモニタリソース (vmw)
  仮想マシンの生存確認を行います。
- 外部連携モニタリソース(mrw)
  ”異常発生通知受信時に実行する異常時動作の設定”と”異常発生通知の WebManager 表示” を実現するためのモニタリソースです。
- ダイナミックDNSモニタリソース (ddnsw)
  定期的にDynamic DNSサーバに仮想ホスト名と活性サーバのIPアドレスを登録します。

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

CLUSTERPRO は以下のアーキテクチャのサーバで動作します。

- i86pc(x86)
- i86pc(x86_64)

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
- 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。

- 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

ソフトウェア

CLUSTERPRO Serverの動作環境

動作可能なバージョン

CLUSTERPRO 独自のドライバモジュールがあります。動作確認を行ったバージョン情報を下記に提示します。

i86pc(x86)

バージョン	clpka 動作可否	CLUSTERPROVersion	備考
Solaris10 10/08	○	3.0.0-1~	
Solaris10 10/09	○	3.0.0-1~	

i86pc(x86_64)

バージョン	clpka 動作可否	CLUSTERPROVersion	備考
Solaris10 10/08	○	3.0.0-1~	
Solaris10 10/09	○	3.0.0-1~	

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

i86pc(x86)

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)	3.0.0-1~	
PostgresSQLモニタ	PostgresSQL 8.3	3.0.0-1~	
	PostgresSQL 8.4	3.0.0-1~	
MySQLモニタ	MySQl 4.0	3.0.0-1~	
	MySQL 5.1	3.0.0-1~	
sambaモニタ	Samba 3.2	3.0.0-1~	
nfsモニタ	バージョン指定無し	3.0.0-1~	

i86pc(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)	3.0.0-1~	
sambaモニタ	Samba 3.2	3.0.0-1~	
nfsモニタ	バージョン指定無し	3.0.0-1~	

仮想マシンリソースの動作環境

仮想マシンリソースの動作確認を行った仮想化基盤のバージョン情報を下記に提示します。

仮想化基盤	バージョン	CLUSTERPRO Version	備考
Solaris Zones	Solaris 10 10/08	3.0.0-1~	x86/x86_64

必要メモリ容量とディスクサイズ

	必要メモリサイズ	インストール直後の必要ディスクサイズ	備考
	ユーザモード	インストール直後	
i86pc(x86)	64MB	20MB	
i86pc(x86_64)	64MB	20MB	

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/中国語
	IE7	日本語/英語/中国語
Microsoft Windows Vista® (IA32)	IE7	日本語/英語/中国語
	IE8	日本語/英語/中国語
Microsoft Windows® 7(IA32)	IE8	日本語/英語/中国語
Microsoft Windows Server 2003 SP1 以降(IA32)	IE6 SP1	日本語/英語/中国語
Microsoft Windows Server 2008 (IA32)	IE7	日本語/英語/中国語

注: 64bit マシン上では「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 6.0 Update21 (1.6.0_21)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)

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 パッケージバージョン
3.0.0-1	3.0.0-1
3.0.2-1	3.0.2-1
3.0.3-1	3.0.3-1

注: オフライン版 Builder のバージョンとCLUSTERPRO パッケージの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/中国語
	IE7	日本語/英語/中国語
Microsoft Windows Vista™ (IA32)	IE7	日本語/英語/中国語
	IE8	日本語/英語/中国語
Microsoft Windows® 7(IA32)	IE8	日本語/英語/中国語
Microsoft Windows Server 2003 SP1 以降(IA32, x86_64)	IE6 SP1	日本語/英語/中国語
Microsoft Windows Server 2008 (IA32)	IE7	日本語/英語/中国語

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™ Runtime Environment

Version 6.0 Update21 (1.6.0_21)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

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
3.0.3-1	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2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	3.0.0-1	WebManager と builder が同一ブラウザ画面から操作可能になりました。
2	3.0.0-1	クラスタ構成ウィザードを刷新しました。
3	3.0.0-1	クラスタ構成ウィザードで一部設定項目の自動取得が可能になりました。
4	3.0.0-1	統合 WebManager をブラウザ上から操作可能に変更しました。
5	3.0.0-1	設定情報のアップロード時、設定内容をチェックする機能を実装しました。
6	3.0.0-1	障害発生時に自律的にフェイルオーバ先を選択することが可能になりました。
7	3.0.0-1	サーバグループを跨ぐフェイルオーバを抑制する機能が実装されました。
8	3.0.0-1	障害検出時のフェイルオーバ対象として「全グループ」が選択可能になりました。
9	3.0.0-1	起動同期待ちをスキップ可能になりました。
10	3.0.0-1	CLUSTERPRO の外部で発生した障害を CLUSTERPRO で管理可能になりました。
11	3.0.0-1	監視対象アプリケーションのタイムアウト発生時、ダンプ情報を取得することが可能になりました。
12	3.0.0-1	オラクル監視で異常を検出した際、オラクルの詳細情報を取得することが可能になりました。
13	3.0.0-1	仮想的なホスト名を DynamicDNS サーバに登録する機能が実装されました。
14	3.0.0-1	Solaris コンテナの大域ゾーンをクラスタ化した場合、非大域ゾーンをリソースとして扱えるようにしました。
15	3.0.0-1	対応 OS を拡充しました。
16	3.0.0-1	対応アプリケーションを拡充しました。
17	3.0.0-1	対応ネットワーク警告灯を拡充しました。
18	3.0.2-1	モニタリソースの回復対象に全グループを指定した場合、WebManager 上での表示を改善しました。

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	修正バージョン / 発生バージョン	修正項目	原因
1	3.0.2-1 / 3.0.0-1	VMライセンスが利用できなかった問題を修正しました。	ライセンス管理テーブルに不足があったため。
2	3.0.2-1 / 3.0.0-1	グループリソース、モニタリソースの異常時最終動作が、Builderでは「クラスタサービス～」、WebManagerでは「クラスタデーモン～」と表示される。	機能間で統一されていない用語があったため。
3	3.0.2-1 / 3.0.0-1	Builderで仮想マシングループのプロパティから排他属性が設定できてしまう。	ウィザードでは設定できないように制限したが、プロパティでは制限処理が漏れていたため。
4	3.0.2-1 / 3.0.0-1	XenServerが利用不可な環境でXenServerのVMモニタの設定を行うと、VMモニタが異常終了(core dump)することがある。	VMモニタの初期化処理でNULLポインタアクセスが発生するため。
5	3.0.2-1 / 3.0.0-1	WebManagerをFIPで接続し、「設定の反映」を実行した場合に「FIP接続に関する注意」が表示されないことがある。	FIPの接続を判断する処理で考慮が漏れていたため。
6	3.0.2-1 / 3.0.0-1	[clprexec]コマンドを使用した場合、syslog、アラートに「Unknown request」が出力されることがある。	syslog、アラートへの出力文字列を作成する処理で「スクリプト実行」、「グループフェイルオーバ」の考慮が漏れていたため。
7	3.0.2-1 / 3.0.0-1	WebManagerで、停止しているサーバのpingnpのステータスが正常と表示される。	NPの状態を初期化していないため、情報が取得できない場合に不定値になっていたため。
8	3.0.2-1 / 3.0.0-1	モニタリソースのプロパティ画面で設定を変更しても「適用」がクリック出来なくなることがある。	判定処理で考慮が漏れていたため。
9	3.0.2-1 / 3.0.0-1	Builderのインタコネクト設定画面で、インタコネクトを複数選択した状態で削除を行うと一部しか削除されない。	複数のインタコネクトが選択されることの考慮が漏れていたため。
10	3.0.2-1 / 3.0.0-1	WebManagerサービス停止時に異常終了することがある。	リアルタイム更新用スレッドが使用するMutexリソースを解放するタイミングに誤りがあったため。
11	3.0.2-1 / 3.0.0-1	サーバ名を変更して再起動する場合にアラート同期サービスが異常終了することがある。	サーバー覧取得処理に問題があったため。

項番	修正バージョン / 発生バージョン	修正項目	原因
12	3.0.2-1 / 3.0.0-1	mdwがタイムアウト、或いは強制killされた場合、OS資源をリークしてしまう。	獲得したsemaphoreを開放するタイミングが無くなるため。
13	3.0.2-1 / 3.0.0-1	初期ミラー構築を行わない設定にした場合、その後一度全面同期をするまで差分同期が有効にならない。	ユーザが意図して初期ミラー構築を行っていない場合でも、ディスク内容の完全一致を保証するフラグが成立しないため。
14	3.0.2-1 / 3.0.0-1	クラスタ生成ウィザードでクラスタ名を変更しても既定値に戻ることがある。	クラスタ生成ウィザードでクラスタ名を変更して次へ進んだ後で、クラスタ名変更画面に戻ると発生する。
15	3.0.2-1 / 3.0.0-1	volmgrwモニタで異常を検出しても回復動作が実行されない。	回復動作を行うかどうかの判定処理が間違っていたため。
16	3.0.2-1 / 3.0.0-1	volmgrリソースのタイムアウトが正しく設定されない。	タイムアウトを計算するための式が間違っているため。
17	3.0.2-1 / 3.0.0-1	キーワードを256文字以上設定すると、mrwモニタを設定していても、外部監視連携が動作しないことがある。	キーワードを保存するためのバッファサイズが不足していたため。
18	3.0.2-1 / 3.0.0-1	シャットダウンストール監視を無効にすると、user空間監視モニタが起動できない。	user空間監視モニタの初期化処理でシャットダウンストール監視の確認処理を行っていたため。
19	3.0.2-1 / 3.0.0-1	シャットダウンストール監視のタイムアウト時間が変更できない。	常にハートビートのタイムアウト時間が使用されるようになっていたため。
20	3.0.2-1 / 3.0.0-1	本体サービスを手動起動にすると、WebManagerサービスとアラートサービスが起動できなくなる。	WebManagerサービスとアラートサービスの設定として本体サービスへの依存関係が設定されているため。
21	3.0.2-1 / 3.0.0-1	miiwモニタで正常な状態にもかかわらず異常と判断することがある。	警告が出力されることを考慮していないことが原因で、警告部分をステータスとして扱い判定するため。
22	3.0.2-1 / 3.0.0-1	FIP、VIPリソースで、IPv6アドレスを設定しalias番号を指定した場合に、活性、非活性に失敗することがある。	IPv6でalias番号指定時に、活性、非活性処理で使用している[ifconfig]コマンドの引数を間違っているため。
23	3.0.3-1 / 3.0.0-1~3.0.2-1	設定モードでVMモニタリソースの「外部マイグレーション発生時の待ち時間」に数値以外(文字や記号)が設定できてしまう。	Builderによる入力ガードに考慮漏れがあったため。
24	3.0.3-1 / 3.0.0-1~3.0.2-1	サーバプライオリティ変更時の反映方法がクラスタサスペンド、リジュームとWebManager再起動になっているが、実際にはクラスタを停止、開始とWebManager再起動が必要となる。	グループリソースの起動サーバがサーバIDとして共有メモリ上に保存されているため、サーバIDが変わると起動サーバの情報が一致しなくなっていたため。
25	3.0.3-1 / 3.0.0-1~3.0.2-1	EXECリソースのタイムアウトとして0を指定すると、EXECリソースの活性が失敗し、緊急シャットダウンしてしまう。	Builderによる入力ガードに考慮漏れがあったため。

項番	修正バージョン / 発生バージョン	修正項目	原因
26	3.0.3-1 / 3.0.0-1~3.0.2-1	特定の環境にて、Builderのクラスタ生成ウィザードでサーバ追加ボタンを押すとアプリケーションエラーが発生する。	JRE側の不具合のため。
27	3.0.3-1 / 3.0.0-1~3.0.2-1	ハイブリッド構成の場合にミラーエージェントが起動しないことがある。	サーバグループを検索するロジックに問題があったため。
28	3.0.3-1 / 3.0.0-1~3.0.2-1	起動待ち合わせ時間に0を指定するとクラスタ本体プロセスが起動しないことがある。	起動待ち合わせ時間に0分が設定された場合は、起動待ちタイムアウトとHB送信開始タイムアウトが同値になってしまい、タイミングによって起動待ち合わせが上手く行えないため。
29	3.0.3-1 / 3.0.0-1~3.0.2-1	複数のモニタ異常が同時に発生し、同じ完全排他グループをフェイルオーバしようとした場合に、両系活性が発生することがある。	グループステータスの返却値に考慮漏れがあったため。
30	3.0.3-1 / 3.0.0-1~3.0.2-1	FIP強制活性の設定が無視される。	別設定値で該当設定が上書かれる実装になってしまっていたため。
31	3.0.3-1 / 3.0.0-1~3.0.2-1	ユーザ空間モニタリソースの遅延警告のアラート(syslog)に表示される時刻の単位が誤っており、tickcountで表示されるべき数値が秒で表示される。	出力時の変換方法を誤っていたため。
32	3.0.3-1 / 3.0.0-1~3.0.2-1	アラートメッセージの内容は512Byteを超えた場合に、アラートデーモンが異常終了する。	アラートメッセージ用のバッファサイズに不足があったため。

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

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。

機能一覧と必要なライセンス

下記オプション製品はサーバ台数分必要となります。

使用したい機能	必要なライセンス
Oracleモニタリソース	CLUSTERPRO X Database Agent 3.0
PostgreSQLモニタリソース	CLUSTERPRO X Database Agent 3.0
MySQLモニタリソース	CLUSTERPRO X Database Agent 3.0
sambaモニタリソース	CLUSTERPRO X File Server Agent 3.0
nfsモニタリソース	CLUSTERPRO X File Server Agent 3.0

Builder、WebManagerの動作OSについて

- 64bitマシン上では「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。

共有ディスクの要件について

- 共有ディスクはSolaris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。

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。

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可能です。
   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。

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。

依存するライブラリ

SUNWlxml

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し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 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。

サーバにファイアフ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]

接続元			接続先		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	icmp	FIP/VIP リソースの重複確認
サーバ	自動割り当て	→	サーバ	XXXX [2]/UDP	内部ログ用通信

［サーバ・WebManager 間］

接続元			接続先		備考
WebManager	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］

接続元			接続先		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］

接続元			接続先		備考
サーバ	自動割り当て	→	ネットワーク警告灯	各製品のマニュアルを参照	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. 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

TCP の通信ポート番号の自動割り当ての範囲は以下のコマンドで確認することができます。

```
#ndd -get /dev/tcp tcp_smallest_anon_port
32768
#ndd -get /dev/tcp tcp_largest_anon_port
65535
```

これは、TCP 通信を行う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、32768～65535 の範囲でアサインされる状態です。

同様に、UDP の通信ポート番号の自動割り当ての範囲は以下のコマンドで確認することができます。

```
#ndd -get /dev/udp udp_smallest_anon_port
32768
#ndd -get /dev/udp udp_largest_anon_port
65535
```

これは、UDP 通信を行う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、32768～65535 の範囲でアサインされる状態です。

OS の設定の変更例

TCP ポートの自動割り当て範囲を 30000～65000 に変更する場合、以下のコマンドを実行します。

```
#ndd -set /dev/tcp tcp_smallest_anon_port 30000
#ndd -set /dev/tcp tcp_largest_anon_port 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。

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。

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[ipmitool]コマンドを使用します。
  - 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
  - 強制停止機能
  - 筐体 ID ランプ連携

- [ipmitool]コマンド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。
  - [ipmitool]コマンド自体に関するお問い合わせ
  - [ipmitool]コマンドの動作保証
  - [ipmitool]コマンドの不具合対応、不具合が原因の障害
  - 各サーバの[ipmitool]コマンド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[ipmitool]コマンド対応可否についてはユーザ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[ipmitool]コマンドが動作しない場合がありますので、ご注意ください。

nsupdate,nslookupについて

- 以下の機能でnsupdateとnslookupを使用します。
  - グループリソースのダイナミック DNS リソース(ddns)
  - モニタリソースのダイナミック DNS モニタリソース(ddnsw)
- CLUSTERPROにnsupdateとnslookupは添付しておりません。ユーザ様ご自身で別途nsupdateとnslookupの rpm ファイルをインストールしてください。
- nsupdate、nslookup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。
  - nsupdate、nslookup 自体に関するお問い合わせ
  - nsupdate、nslookup の動作保証
  - nsupdate、nslookup の不具合対応、不具合が原因の障害
  - 各サーバの nsupdate、nslookup の対応状況のお問い合わせ

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。

環境変数

環境変数が 256 個以上設定されている環境では、下記のスクリプトが実行できません。下記の機能またはリソースを使用する場合は、環境変数を 255 個以下に設定してください。

- execリソースが活性/非活性時に実行する開始/停止スクリプト
- カスタムモニタリソースが監視時に実行するスクリプト
- グループリソース、モニタリソース異常検出後の最終動作実行前スクリプト

強制停止機能、筐体IDランプ連携

強制停止機能、筐体 ID ランプ連携を使用する場合、各サーバの BMC の IP アドレス、ユーザ名、パスワードの設定が必須です。ユーザ名には必ずパスワード登録されているものを設定してください。

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
  -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
  -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 keepalive

- シャットダウンストール監視
  -監視方法 keepalive
- 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』の「第 4 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。

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。
- 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1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2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。

スクリプトのコメントなどで取り扱える2バイト系文字コードについて

- CLUSTERPROでは、Solaris環境で編集されたスクリプトはEUC、Window環境で編集されたスクリプトはShift-JISとして扱われます。その他の文字コードを利用した場合、環境によっては文字化けが発生する可能性があります。

仮想マシングループのフェイルオーバ排他属性の設定について

- 仮想マシングループを設定する場合には、フェイルオーバ排他属性には「通常排他」、「完全排他」を設定しないでください。

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。

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 → 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、サポート対象外となります。

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。

- モニタリソースの一時停止後、監視対象リソースを停止させた場合モニタリソースは停止状態となります。そのため、監視の再開はできません。
- モニタリソースを一時停止後、監視対象リソースを停止/起動させた場合、監視対象リソースが起動したタイミングで、モニタリソースによる監視が開始されます。

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]をクリックして最新の内容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。
  次回の自動更新まで待つか、[リロード]をクリックして最新の内容を再取得してください。
- 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 (Cluster Managerの設定モード) について

- 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]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 のメニューバーの[ファイル]-[設定のエクスポート]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]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 のメニューバーの[ファイル]-[設定のエクスポート]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。

サービス起動時間について

CLUSTERPRO の各サービスは、起動時の待ち合わせ処理の有無により時間がかかる場合があります。

- clusterpro_evt
  マスタサーバ以外のサーバは、マスタサーバの構成情報をダウンロードする処理を最大2分間待ち合わせます。マスタサーバが起動済みの場合、通常数秒以内に終了します。
  マスタサーバはこの処理で待ち合わせは発生しません。
- clusterpro_trn
  特に待ち合わせ処理はありません。通常数秒以内に終了します。
- clusterpro
  特に待ち合わせ処理はありませんが、CLUSTERPRO の起動に時間がかかる場合数十秒かかります。通常数秒以内に終了します。
- clusterpro_webmgr
  特に待ち合わせ処理はありません。通常数秒以内に終了します。
- clusterpro_alertsync
  特に待ち合わせ処理はありません。通常数秒以内に終了します。

さらに、CLUSTERPRO デーモン起動後は、クラスタ起動同期待ち処理があり、デフォルト設定では、5 分間の待ち合わせがあります。

これに関しては『リファレンスガイド』の「第 9 章 保守情報 クラスタ起動同期待ち時間について」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 のアップデート手順

X2.1からX3.0へのアップデート

CLUSTERPRO Server パッケージは root ユーザでインストールしてください。

1. オンライン Builder または[clpcfctrl]コマンドを使用して構成情報を取得します。
2. WebManager または[clpcl]コマンドを使用してクラスタを停止します。
3. **svcadm disable *name*** を以下の順序で実行してサービスを無効にします。 *name* には無効にするサービスを指定します。
   - clusterpro_alertsync
   - clusterpro_webmgr
   - clusterpro
   - clusterpro_trn
   - clusterpro_evt
4. CLUSTERPRO のサービスが起動していないことを確認してから、[pkgrm]コマンドを実行してパッケージファイルをアンインストールします。

   ```
   pkgrm NECclusterpro
   ```

5. インストール CD-ROM の媒体を mount します。
6. [pkgadd]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。
   パッケージファイルは CD-ROM 内の `/Solaris/3.0/jp/server` 配下にあります。
   アーキテクチャにより利用するファイルが異なります。アーキテクチャには i686、x86_64 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。

   ```
   pkgadd -d NECclusterpro-<バージョン>-<アーキテクチャ>.pkg
   ```

   CLUSTERPRO のインストールディレクトリは `/opt/nec/clusterpro` です。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るので注意してください。
7. インストール終了後 CD-ROM を umount し、取り除きます。
8. 手順 3～7を全てのサーバで実行します。
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で初期構築用ダイアログが表示されるので、「クラスタ構成情報をインポートする」を選択し、手順 1で取得したクラスタ構成情報を読み込んでください。
12. 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、設定を反映します。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。設定を反映すると自動的に [マネージャ再起動] が実行されます。
13. 全てのサーバで OS 再起動を実行してください。

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 などでつないで、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

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、フェイルオーバグループがフェイルオーバする先のサーバ
(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

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
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

フェイルオーバポリシー
フェイルオーバ可能なサーバリストとその中でのフェイルオーバ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

付録 B　　索引

せ

そ

ち

つ

て

と

な

ね

は

ふ

め

も

り