CLUSTERPRO® X 1.0 *for Windows*

スタートアップガイド

2011.01.21
第5版

改版履歴

版数	改版日付	内 容
1	2006/09/08	新規作成
2	2006/12/28	ESMPRO/AutomaticRunningController との連携に関する記述を追加 ロゴの変更に対応 誤記や体裁上の修正
3	2007/06/22	内部バージョン9.03で追加されたリソースに関する記述を追加 リリースノートを更新
4	2007/09/28	WebOTX監視リソースに関する記述を追加 リリースノートを更新
5	2011/01/21	注意制限事項を更新 内部バージョン9.0b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。
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。
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® X スタートアップガイド』の構成は、セクション I とセクション II の2部に分かれています。
セクション I では、CLUSTERPRO を初めてご使用になるユーザを対象に、CLUSTERPRO の製品概要と基本的な使用方法について説明します。
セクション II では、CLUSTERPROを導入前のユーザ、および導入後のアップデートを行うユーザを対象に、最新の動作環境情報や制限事項などについて紹介します。

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?」:クラスタシステムの概要について説明します。
第 2 章　「CLUSTERPROについて」: CLUSTERPRO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。

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)
CLUSTERPRO を使用する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、および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 Alert Service 管理者ガイド』(Alert Service Administrator's Guide)
CLUSTERPRO を使用したクラスタシステムに CLUSTERPRO Alert Service 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 X Alert Service 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」および「重要」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]`
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	`c:¥Program files¥CLUSTERPRO`
モノスペースフォント**太字** (courier)	ユーザが実際にコマンドプロンプト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	`clpstat -s [-h host_name]`

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について

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。

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可用性を高めます。

クラスタとは、「群れ」「房」を意味し、その名の通り、「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 (High Availability) クラスタ**

  通常時は一方が現用系として業務を稼動させ、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す。共有ディスク型、ミラーディスク型があります。

- **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、各ノードに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き継ぎはできません。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**

  非常に計算量が多いクラスタのこと。スーパコンピュータを用いて単一の業務を実行するためのクラスタです。全てのノードの CPU を利用し、単一の業務を実行する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、ノード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などのダウンタイムがビジネスに大きな影響を与えてしまうシステムに、HA クラスタの導入が求められています。

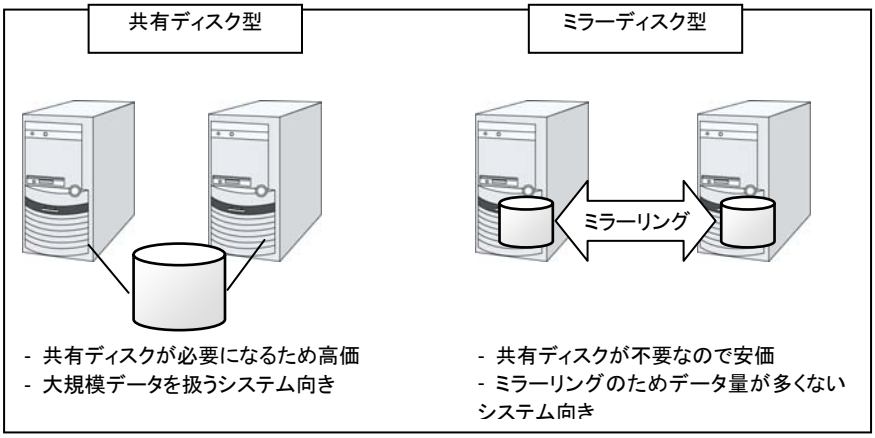


図 1-1 HA クラスタ構成図

HA クラスタは、共有ディスク型とミラーディスク型に分けることができます。次ページから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、SAN 接続の FibreChannel ディスクアレイ装置のように複数のサーバからアクセス可能な外付けディスク（共有ディスク）上に置き、このディスクを介してサーバ間でデータを引き継ぐ形態を共有ディスク型といいます。

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、障害発生時に業務を引き継ぐサーバ(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、業務単位にサーバに付与している IP ではなく別ネットワークの IP アドレス(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。

現用系のダウンによりフェイルオーバが発生すると、共有ディスク上のデータは適切な終了処理が行われないまま待機系に引き継がれることになります。このため、待機系で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。これは一般に、クラスタ化されていないシステムでダウン後の再起動時に行われるのと同様の処理になります。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。

障害発生後は、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、グループのフェイルバックを行わなくても良いです。どうしても、元のサーバで業務を行いたい場合は、グループの移動を実行してください。

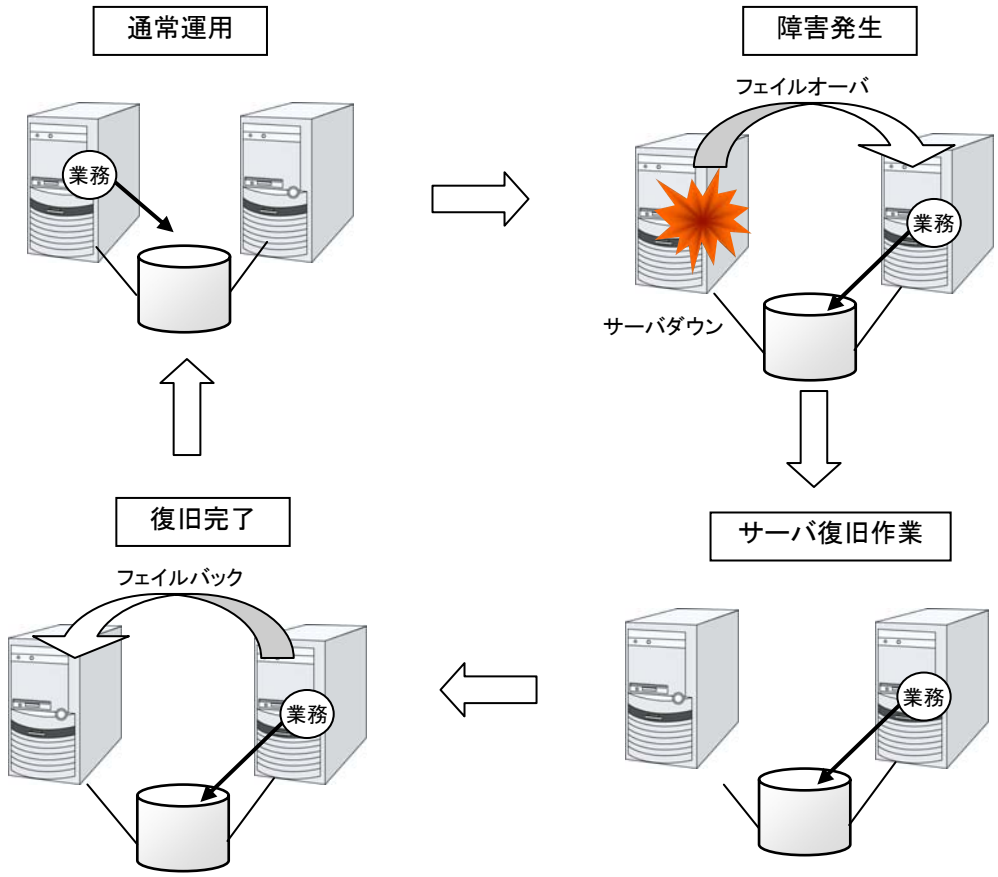


図 1-2 障害発生から復旧までの流れ

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、元のノードの復旧作業が完了してから一旦業務を停止し、元のノードで業務を再開します。フェイルオーバしたグループを元のサーバに戻すことをフェイルバックといいます。

図 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、それぞれのノード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。

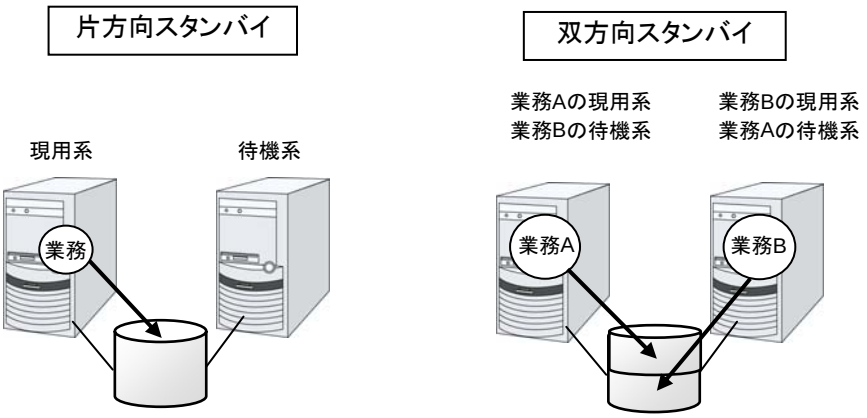


図 1-3 HA クラスタの運用形態

ミラーディスク型

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、同等の機能をより低価格で実現したクラスタシステムをミラーディスク型といいます。

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みます。書き込んだデータを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、サーバ間をつなぐケーブルのことで、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。

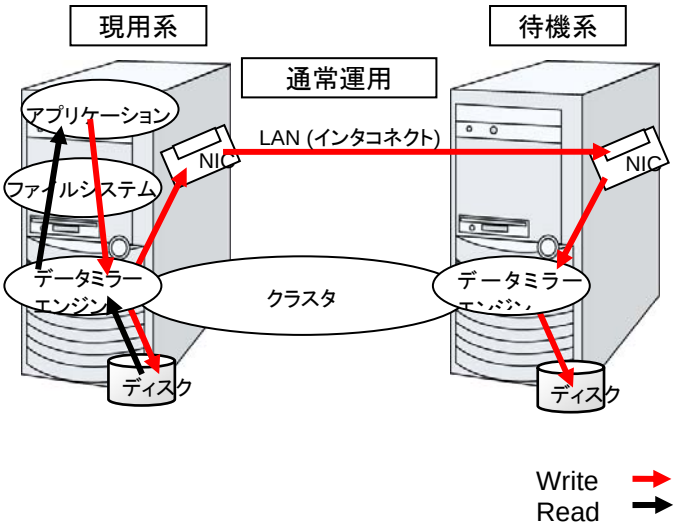


図 1-4 データミラーの仕組み

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、スナップショットバックアップとしてデータを保存する運用が可能です。

HA クラスタの仕組みと問題点

次に、クラスタの実装と問題点について説明します。

システム構成

共有ディスク型クラスタは、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。

ミラーディスク型クラスタは、クラスタサーバ上のデータディスクをネットワーク経由でミラーリングする構成です。サーバ障害時には待機系サーバ上のミラーデータを使用し業務を引き継ぎます。データのミラーリングは I/O 単位で行うため上位アプリケーションから見ると共有ディスクと同様に見えます。

以下の図は、共有ディスク型クラスタの構成例です。

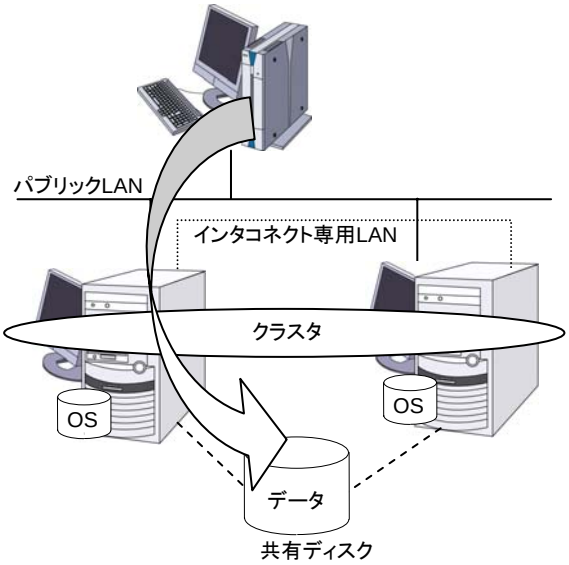


図 1-5 システム構成

フェイルオーバ型クラスタは、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

一方のサーバを運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。

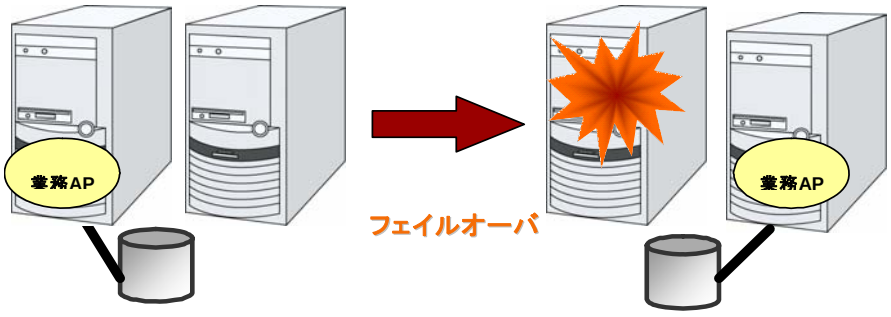


図 1-6 片方向スタンバイクラスタ

同一アプリケーション双方向スタンバイクラスタ

複数のサーバで同じ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。各業務アプリケーションは独立して動作します。フェイルオーバ時には一台のサーバ上で同一業務アプリケーションが複数動作することになりますので、このような運用が可能なアプリケーション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。

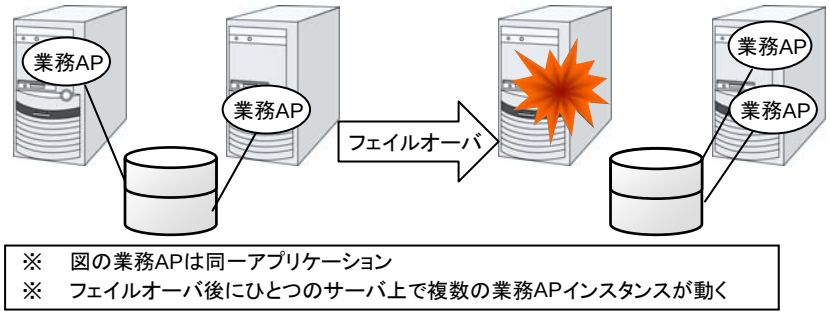


図 1-7 同一アプリケーション双方向スタンバイクラスタ

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。フェイルオーバ時には一台のサーバ上に複数の業務アプリケーションが動作することになりますので、これらのアプリケーションは共存可能でなければなりません。業務単位での負荷分散システムを構築できます。

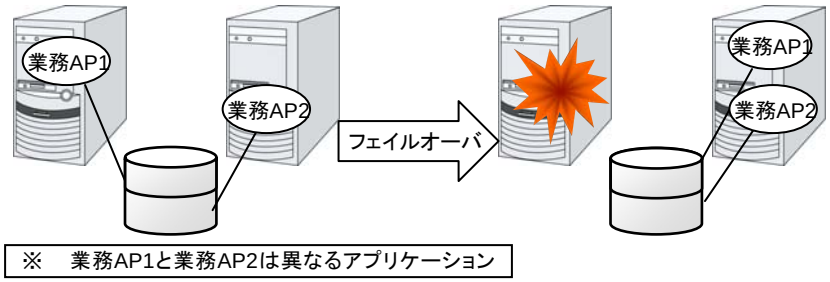


図 1-8 異種アプリケーション双方向スタンバイクラスタ

N サーバ + M 業務 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、正常時には待機系サーバが何も業務を行わないため、無駄なリソースの比率が 1/2 になっていたのですが、この構成の場合無駄なリソースの比率が 1/4 となり、コストの削減ができます。また、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。

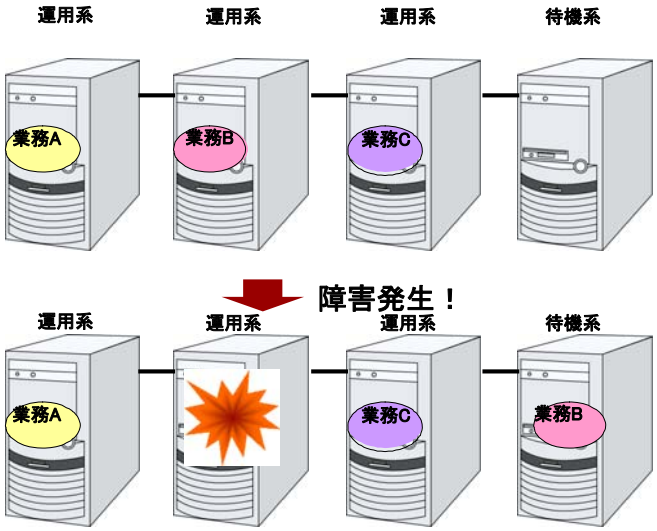


図 1-9 N + N 構成

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のダウンです。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のダウン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。

共有ディスクの排他制御

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、次に説明する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。

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だけではサーバのダウンと区別できません。この状態でサーバダウンとみなし、フェイルオーバ処理を実行し、複数のサーバでファイルシステムを同時にマウントすると、共有ディスク上のデータが破壊されてしまいます。

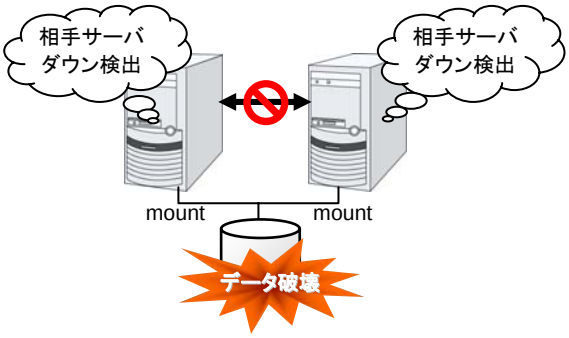


図 1-10 ネットワークパーティション症状

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。この問題を解決するため、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

共有ディスク型クラスタでは、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。

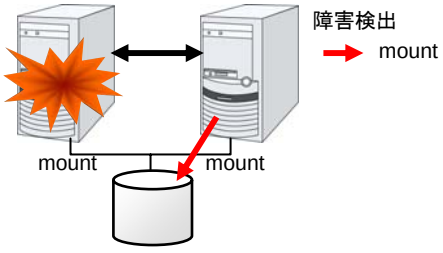


図 1-11 データの引き継ぎ

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、ファイルシステムやデータベースの復旧時間の問題です。引き継ごうとしているファイルは、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、ファイルシステムによっては引き継ぐ際に整合性チェックが必要となりますし、データベースであればロールバック等の処理が必要となります。これは電源障害などでダウンした単体サーバを再起動した場合と同様です。このような復旧処理に長時間を要する場合、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共有ディスクにデータを書き出す際に、通常はファイルシステムを介しての書き出しになりますが、アプリケーションが書き込みを完了していても、ファイルシステムがディスクキャッシュ上に保持しているだけで、共有ディスクへの書き込みを行っていなかった場合、この状態で現用系のサーバがダウンすると、ディスクキャッシュ上のデータは待機系に引き継がれないことになります。このため、障害発生時に確実に待機系に引き継ぐ必要のある大切なデータは、同期書き込みなどにより確実にディスクに書き込む必要があります。これは単体サーバがダウンした際にデータが揮発しないようにするのと同じです。つまり、待機系に引き継がれるのは共有ディスクに記録されたデータのみであり、ディスクキャッシュのようなメモリ上のデータは引き継がれないということを考慮してクラスタシステムを設計する必要があります。

IPアドレスの引き継ぎ

次にクラスタソフトウェアが行うことは、IP アドレスの引き継ぎです。フェイルオーバした際に、IP アドレスを引き継ぐことで、業務がどのサーバで動作しているのか、気にすることなく作業を行うことができます。クラスタソフトウェアは、そのための IP アドレスの引き継ぎを行います。

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、DB のインスタンスをフェイルオーバする場合、障害発生直前の状態で再開されるのではなく、一旦ダウンした状態から再起動した場合と同様にトランザクションのロールバック等が行われ、クライアントからも再接続が必要になります。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。

フェイルオーバについての総括

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

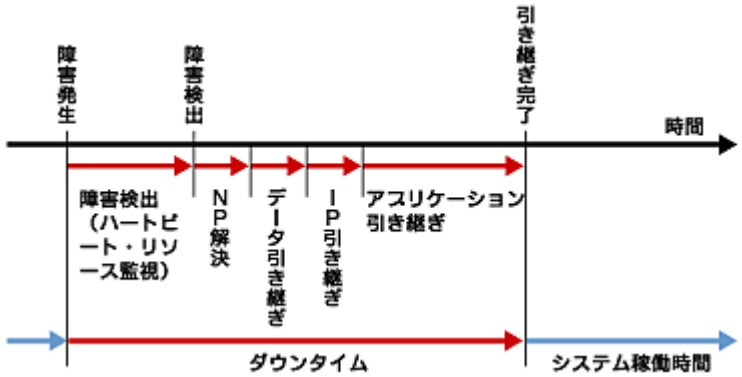


図 1-12 フェイルオーバタイムチャート

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。

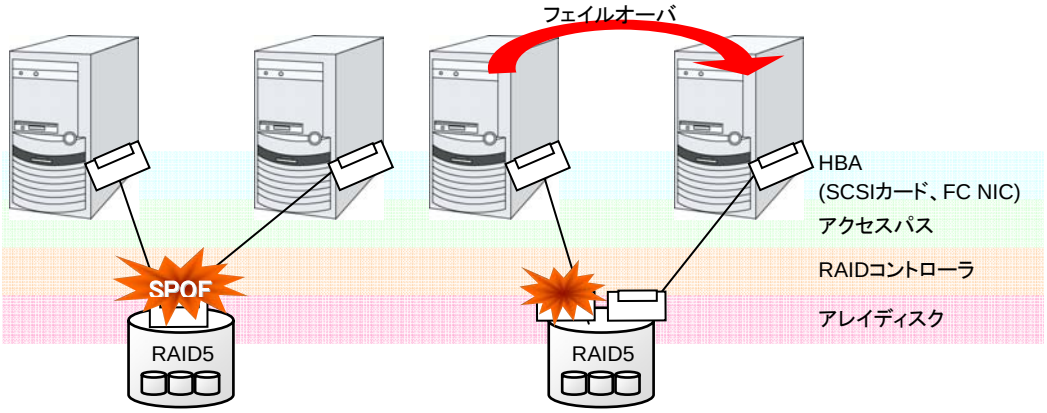
- 共有ディスク
- 共有ディスクへのアクセスパス

- 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。



※HBA: 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。

図 1-13 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、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)の低下
- 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) の低下
- ミラー再同期時間(ミラー再同期が完了するまでフェイルオーバできない)

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のに有効といえます。

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。

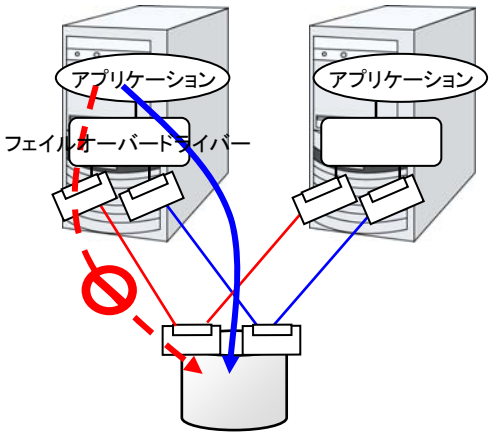


図 1-14 パスフェイルオーバドライバ

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。

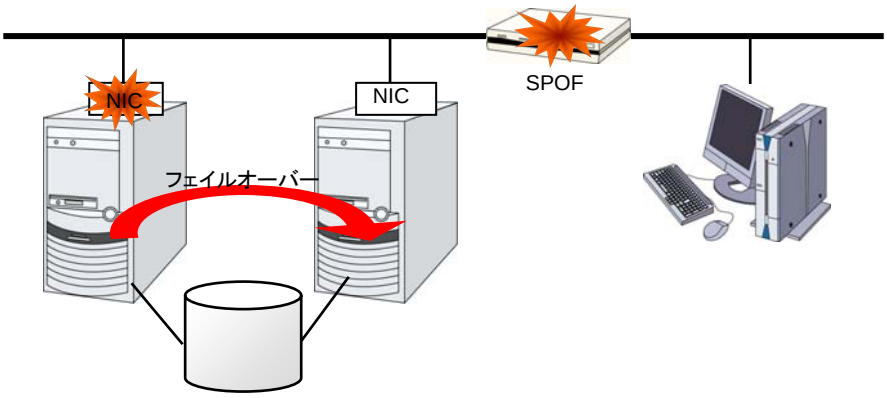


図 1-15 ルータが SPOF となる例

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「一連の状態遷移」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り、長期間運用を続けると障害が発生することがあります。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。

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。注意すべき点を簡単にまとめます。

- 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 について

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、HA クラスタシステムを実現するためのソフトウェアです。

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 の本体です。クラスタシステムを構成する各サーバマシンにインストールします。Server には、CLUSTERPRO の高可用性機能の全てが包含されています。また、WebManager および Builder 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 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 を介して CLUSTERPRO 本体に含まれる Builder を利用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。クラスタを構成するサーバ上には「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。

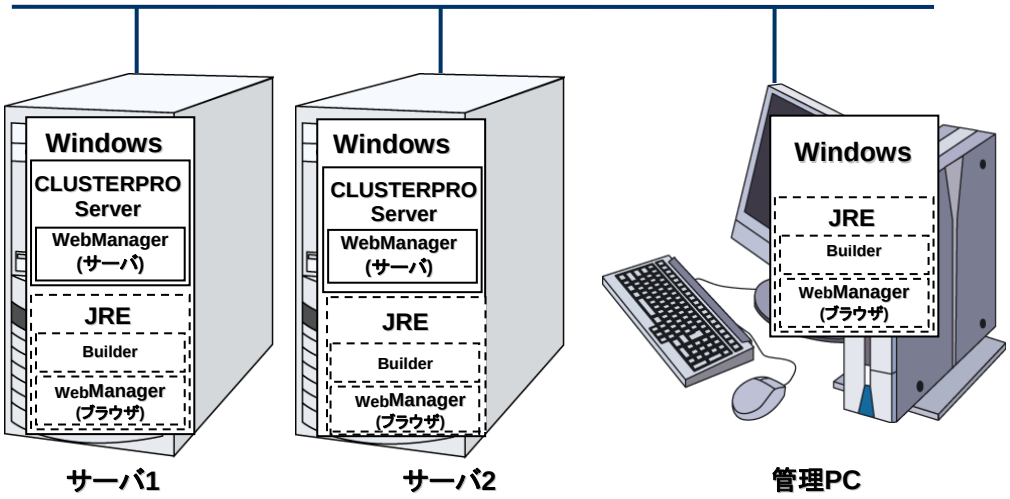


図 2-1 CLUSTERPRO のソフトウェア構成

注: JRE とは、Java Runtime Environment のことです。

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。

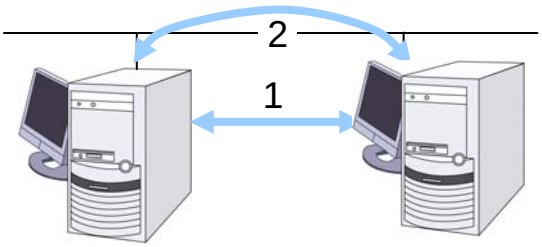
CLUSTERPRO はサーバ監視のために、定期的にサーバ同士で生存確認を行います。この生存確認をハートビートと呼びます。ハートビートは以下の通信パスを使用して行います。

- インタコネクト専用LAN

  クラスタサーバ間通信専用の LAN です。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用いるパスとして使用します。サーバ間の情報交換や、インタコネクトのバックアップ用としても使用します。



1 インタコネクト専用LAN
2 パブリックLAN

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。

- 監視オプションによるアプリケーション/プロトコルのストール/結果異常監視

  別途ライセンスの購入が必要となりますが、データベースアプリケーション(Oracle,DB2等)、プロトコル(FTP,HTTP 等) 、アプリケーションサーバ(Websphere, Weblogic 等)のストール/結果異常監視を行うことができます。詳細は、『リファレンスガイド セクションⅡ』の「第 7 章 モニタリソースの詳細」を参照してください。

- 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 (アプリケーションリソース、サービスリソースと呼びます) により起動し、監視用のリソース (アプリケーション監視リソース、サービス監視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、どのような障害が検出可能なのか、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)

- STOP エラー

◆ 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 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 [1]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、そのプログラムをアプリケーションリソースで起動、アプリケーション監視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。

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。

◆ COM 方式

◆ ping 方式

◆ 共有ディスク方式

◆ COM+共有ディスク方式

◆ ping+共有ディスク方式

◆ 多数決方式

◆ ネットワークパーティション解決しない

[1] 監視オプションで取り扱う、データベースアプリケーション(Oracle,DB2等)、プロトコル(FTP,HTTP等) 、アプリケーションサーバ(Websphere, Weblogic等)については、ストール/結果異常監視を行うことができます。

関連情報: ネットー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 9 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は他サーバからのハートビートの途絶を検出すると、フェイルオーバ開始前に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。

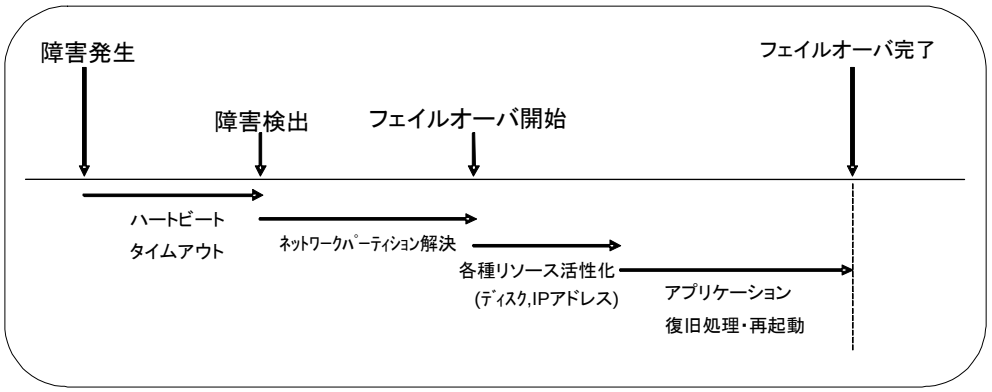


図 2-2 フェイルオーバのタイムチャート

- 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。
  - 業務の負荷等による遅延も考慮して、クラスタプロパティの設定値を調整します。(規定値では 30 秒です。)
- ネットワークパーティション解決
  - 相手サーバからのハートビートの途絶(ハートビートタイムアウト)が、ネットワークパーティション症状によるものか、実際に相手サーバが障害を起こしたのかを確認するための時間です。
  - ネットワークパーティション方式として共有ディスク方式が指定されている場合には、ディスク I/O の遅延を考慮した待ち時間が必要なため、既定値の設定で 30 秒～60 秒程の時間を要します。この所要時間は CLUSTER パーティションへのアクセス時間や、ハートビートタイムアウト値などに連動して変化します。その他の方式の場合、通常はほぼ瞬時に確認が完了します。
- 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資源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。)

◆ アプリケーション復旧処理・再起動

- 業務で使用するアプリケーションの起動に要する時間です。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処理の時間も含まれます。
- 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。

CLUSTERPROで構築する共有ディスク型クラスタのハードウェア構成

共有ディスク型クラスタの 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。

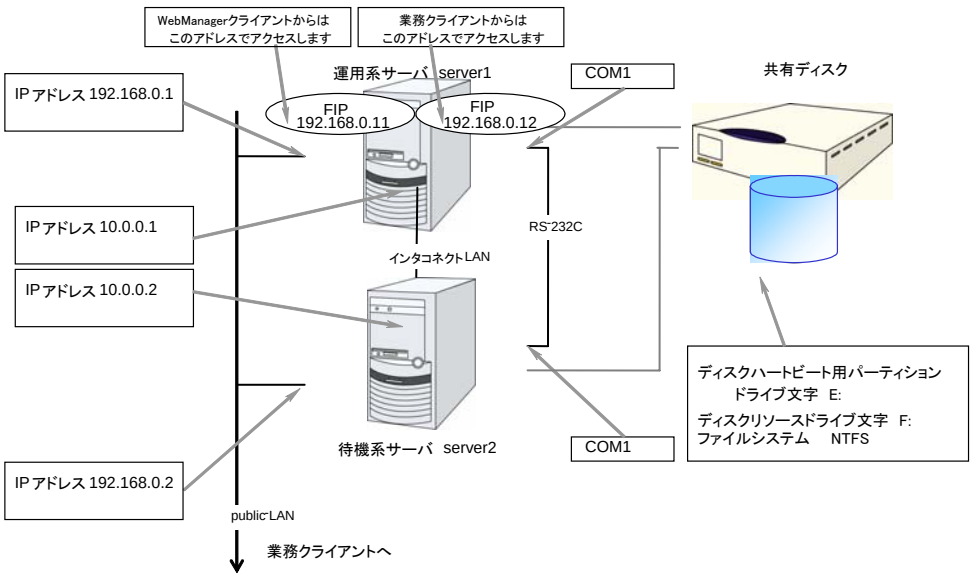


図 2-3 共有ディスク使用時のクラスタ環境のサンプル

上記は、共有ディスク使用時のクラスタ環境のサンプルです。

CLUSTERPROで構築するミラーディスク型クラスタのハードウェア構成

各サーバのディスク上のパーティションをミラーリングすることによって、共有ディスク装置の代替とする構成です。共有ディスク型に比べて小規模で低予算のシステムに向いています。

注: ミラーディスクを使用するには、Replicator オプションをご購入いただく必要があります。

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、インタコネクト(CLUSTERPRO の内部通信用 NIC)で兼用します。

CLUSTERPRO で構築するデータミラー型クラスタのハードウェア構成は、下図のような構成になります。

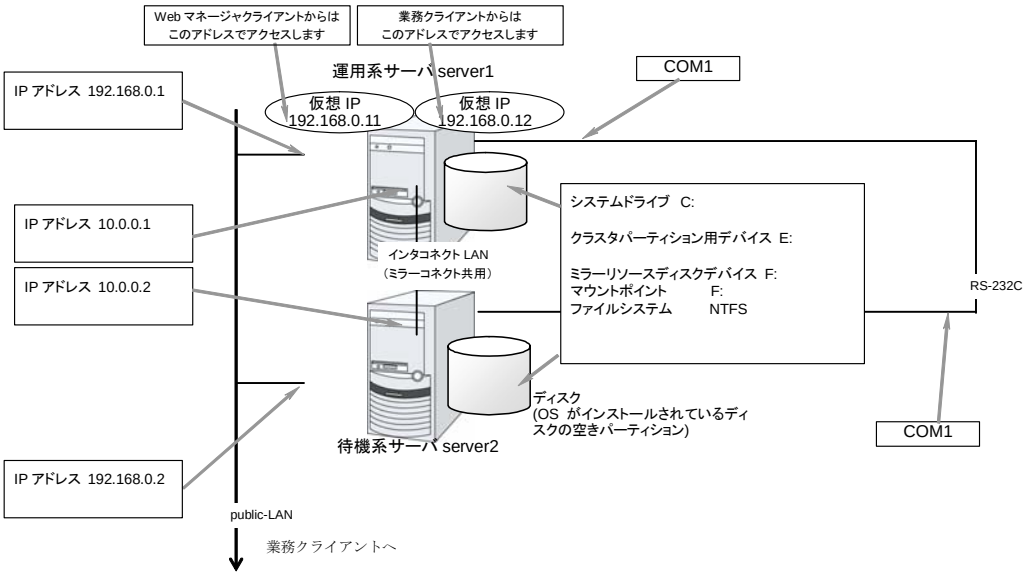


図 2-4 ミラーディスク使用時のクラスタ環境のサンプル

上記は、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)です。

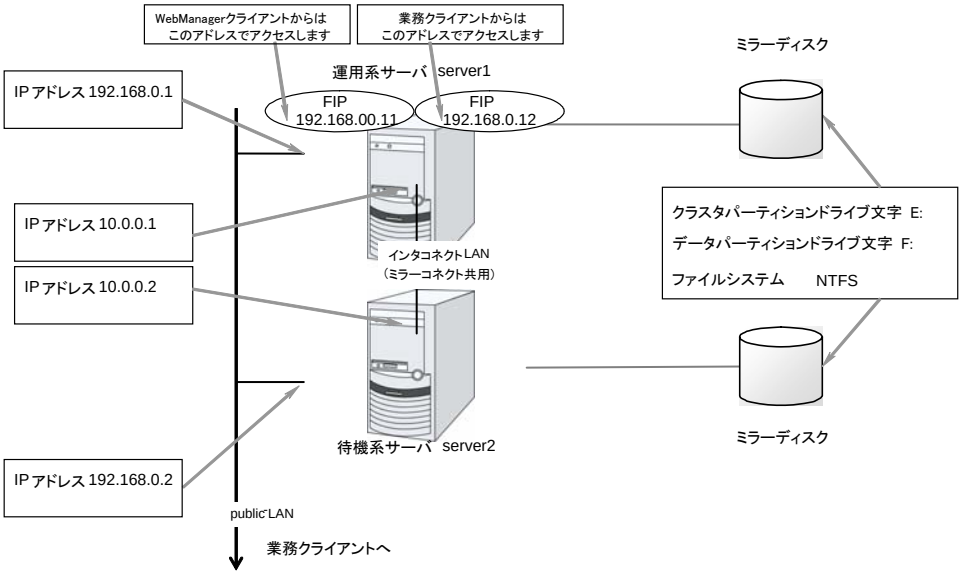


図 2-5 ミラーディスク使用時のクラスタ環境のサンプル

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)です。

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
  一群のサーバをまとめたクラスタシステムです。
- 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
  仮想の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
  仮想サーバの持つ資源(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、監視する側とされる側のリソースを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。

各リソースの詳細については、『リファレンスガイド セクションⅡ』を参照してください。

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。

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。

- COM ネットワークパーティション解決リソース
  COM 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。
- DISK ネットワークパーティション解決リソース
  DISK 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。共有ディスク構成の場合のみ利用可能です。
- 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。
- 多数決ネットワークパーティション解決リソース
  多数決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。

- アプリケーションリソース (appli)
  アプリケーション(ユーザ作成アプリケーションを含む)を起動／停止するための仕組みを提供します。
- 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。
- ミラーディスクリソース (md)
  ローカルディスク上の特定のパーティションのミラーリングとアクセス制御を行う機能を提供します。ミラーディスク構成の場合のみ利用可能です。
- レジストリ同期リソース (regsync)
  クラスタを構成するサーバ間でアプリケーションやサービスを同一設定で動作させるために、複数サーバの特定レジストリを同期する仕組みを提供します。
- スクリプトリソース (script)
  ユーザ作成スクリプト等のスクリプト(BAT)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 (sd)
  共有ディスク上の特定のパーティションのアクセス制御を行う機能を提供します。共有ディスク装置が接続されている場合にのみ利用可能です。
- サービスリソース (service)
  データベースや Web 等のサービスを起動／停止するための仕組みを提供します。
- プリントスプーラリソース (spool)
  プリントスプーラをフェイルオーバするための機能を提供します
- 仮想コンピュータ名リソース (vcom)
  仮想的なコンピュータ名を提供します。クライアントからは一般のコンピュータ名と同様にアクセス可能です。

- 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。
- CIFS リソース (cifs)
  共有ディスク/ミラーディスク上のフォルダを共有公開するための機能を提供します。
- NAS リソース (nas)
  ファイルサーバ上の共有フォルダをネットワークドライブとしてマウントするための機能を提供します。

注:

ミラーディスクリソースを使用するためには、『CLUSTERPRO X Replicator』 のライセンスが必要です。

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。

- アプリケーション監視リソース (appliw)
  アプリケーション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。
- ディスク RW 監視リソース (diskw)
  ファイルシステムへの監視機構を提供します。また、ファイルシステム I/O ストール時に意図的な STOP エラーまたは、HW リセットによりフェイルオーバを実施する機能を提供します。共有ディスクのファイルシステムへの監視にも利用できます。
- フローティング IP 監視リソース (fipw)
  フローティング IP リソースで起動した IP アドレスの監視機構を提供します。
- IP監視リソース (ipw)
  ネットワークの疎通を監視する機構を提供します
- ミラーディスク監視リソース (mdw)
  ミラーディスクの監視機構を提供します。
- ミラーコネクト監視リソース (mdnw)
  ミラーコネクトの監視機構を提供します。
- NIC Link Up/Down監視リソース (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。
- マルチターゲット監視リソース (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。
- レジストリ同期監視リソース (regsyncw)
  レジストリ同期リソースによる同期処理の監視機構を提供します。
- ディスク TUR 監視リソース (sdw)
  SCSI の TestUnitReady コマンドにより共有ディスクへのアクセスパスの動作を監視する機構を提供します。FibreChannel の共有ディスクに対しても使用できます。
- サービス監視リソース (servicew)
  サービス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。

- プリントスプーラ監視リソース (spoolw)
  プリントスプーラリソースで起動したプリントスプーラの監視機構を提供します。
- 仮想コンピュータ名監視リソース (vcomw)
  仮想コンピュータ名リソースで起動した仮想コンピュータの監視機構を提供します。
- 仮想 IP 監視リソース (vipw)
  仮想 IP リソースで起動した IP アドレスの監視機構を提供します。
- CIFS 監視リソース (cifsw)
  CIFS リソースで公開した共有フォルダの監視機構を提供します。
- NAS 監視リソース (nasw)
  NAS リソースでマウントしたネットワークドライブの監視機構を提供します。
- DB2 監視リソース (db2w)
  IBM DB2 データベースへの監視機構を提供します。
- ODBC 監視リソース (odbcw)
  ODBC でアクセス可能なデータベースへの監視機構を提供します。
- Oracle 監視リソース (oraclew)
  Oracle データベースへの監視機構を提供します。
- PostgreSQL 監視リソース (psqlw)
  PostgreSQL データベースへの監視機構を提供します。
- SQL Server 監視リソース (sqlserverw)
  SQL Server データベースへの監視機構を提供します。
- FTP 監視リソース (ftpw)
  FTP サーバへの監視機構を提供します。
- HTTP 監視リソース (httpw)
  HTTP サーバへの監視機構を提供します。
- IMAP4 監視リソース (imap4w)
  IMAP サーバへの監視機構を提供します。
- POP3 監視リソース (pop3w)
  POP サーバへの監視機構を提供します。
- SMTP 監視リソース (smtpw)
  SMTP サーバへの監視機構を提供します。
- Tuxedo 監視リソース(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 監視リソース (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 監視リソース(wlsw)
  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 監視リソース (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。

注:

DB2 監視リソース、ODBC 監視リソース、Oracle 監視リソース、PostgreSQL 監視リソース、SQL Server 監視リソースを使用するためには、『CLUSTERPRO X Database Agent』 のライセンスが必要です。

FTP 監視リソース、HTTP 監視リソース、IMAP4 監視リソース、POP3 監視リソース、SMTP 監視リソースを使用するためには、『CLUSTERPRO X Internet Server Agent』 のライセンスが必要です。

Tuxedo 監視リソース、Weblogic 監視リソース、Websphere 監視リソース、WebOTX 監視リソースを使用するためには、『CLUSTERPRO X Application Server Agent』 のライセンスが必要です。

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。

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 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。

ハードウェア動作環境

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

必要スペック

CLUSTERPRO Server に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 (ミラーディスクを使用する場合)
- CD-ROMドライブ

オフラインで Builder を使用する場合は、クラスタ構成情報のやりとりのため以下のいずれかが必要です(オンラインの場合は不要)。

- FDドライブ,USBメモリなどのリムーバブルメディア
- CLUSTERPRO Server をインストール済のサーバマシンとファイルを共有する手段

CLUSTERPRO Server の動作環境

対応OS

CLUSTERPRO は、下記の OS に対応しています。

IA32 版

OS	Replicator サポート	備考
Microsoft Windows Server 2003, Standard Edition SP1以降	○	
Microsoft Windows Server 2003, Enterprise Edition SP1以降	○	
Microsoft Windows Server 2003, Standard Edition R2	○	
Microsoft Windows Server 2003, Enterprise Edition R2	○	

x86_64 版

OS	Replicator サポート	備考
Microsoft Windows Server 2003, Standard x64 Edition SP1以降	○	
Microsoft Windows Server 2003, Enterprise x64 Edition SP1以降	○	
Microsoft Windows Server 2003, Standard x64 Edition R2	○	
Microsoft Windows Server 2003, Enterprise x64 Edition R2	○	

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版	60MB	32MB + 4MB×ミラーリソース数	20MB	600MB	
x86_64版	110MB	32MB + 4MB×ミラーリソース数	40MB	600MB	

ミラーディスクリソースに必要なメモリサイズ、およびディスクサイズです。

非同期方式に変更時やキューサイズ変更時は、構成時に指定したサイズのメモリが追加で必要になります。また、ミラーディスクへの I/O に対応してメモリを使用するため、ディスク負荷の増加にともない使用するメモリサイズも増加します。

Builder の動作環境

CLUSTERPRO Builder を動作させるために必要な環境について記載します。

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	IE6 SP2	日本語/英語
Microsoft Windows Vista™	IE7	日本語/英語
Microsoft Windows Server 2003 SP1以降	IE6 SP1	日本語/英語
Microsoft Windows Server 2003 R2	IE6 SP1	日本語/英語

注: Builder は 64bit、x86_64 上では動作しません。クラスタ構成を作成、変更するには 32bit の OS 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(TM) Runtime Environment
Version 5.0 Update 6 (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 実行環境を除く)

対応するCLUSTERPROのバージョン

Builder のバージョンと CLUSTERPRO 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。

Builderバージョン	CLUSTERPRO X Server 内部バージョン
1.0.0-1	9.00
1.0.0-1	9.01
1.0.2-1	9.02
1.1.0-1	9.03
1.1.0-1	9.04
1.1.2-1	9.05
1.1.2-1	9.06
1.1.4-1	9.07
1.1.4-1	9.08
1.1.4-1	9.09
1.1.7-1	9.0a
1.1.7-1	9.0b

WebManager の動作環境

CLUSTERPRO WebManager を動作させるために必要な環境について記載します。

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 SP2	IE6 SP2	日本語/英語
Microsoft Windows Vista™	IE7	日本語/英語
Microsoft Windows Server 2003 SP1	IE6 SP1	日本語/英語
Microsoft Windows Server 2003 R2	IE6 SP1	日本語/英語

注: Builder は 64bit 、x86_64 上では動作しません。クラスタ構成を作成、変更するには 32bit の 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 Update 6 (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　300KB 以上(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。

最新バージョン

2011 年 1 月時点での CLUSTERPRO X 1.0 for Windows の最新内部バージョンは 9.0b です。
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。

CLUSTERPRO の内部バージョンは、WebManager で確認してください。
WebManager のツリービューからサーバのアイコンを選択すると、そのサーバの内部バージョンがリストビューに表示されます。

内部バージョンが 9.0a 以前の場合、アップデート CPRO-XW010-10 を適用することにより 9.0b にバージョンアップすることができます。アップデートの適用手順と、アップデートにより修正される障害情報については、アップデート手順書を参照してください。

機能強化情報

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	9.02	ESMPRO/AutomaticRunningController との連携機能を追加しました。
2	9.03	以下のグループリソースとモニタリソースを追加しました。 グループリソース： cifs, nas モニタリソース： cifsw, nasw, tuxw, wlsw, wasw
3	9.03	仮想コンピュータ名の DNS への動的登録が可能になりました。
4	9.03	非同期モードのミラーディスクがミラーコネクト通信に使用する通信帯域と履歴ファイル格納フォルダに格納する一時ファイルのサイズの上限設定が可能になりました。
5	9.03	ミラーディスクリソースでミラーリングするデータパーティションのサイズを調整する clpvolsz コマンドを追加しました。
6	9.03	clprsc コマンドによりグループリソースの起動・停止をバッチファイル等から制御できるようになりました。
7	9.03	WebManager と Builder が Windows Vista + IE7 及び Java™ Runtime Environment Version 6.0 に対応しました。
8	9.03	IPv6 と IPv4 の混在環境に対応しました（クラスタサーバ間通信には IPv4 を用いる必要があります）。
9	9.03	CLUSTERPRO X Application Server Agent 1.0 for Windows に対応しました。
10	9.05	以下のモニタリソースを追加しました。 モニタリソース： otxw
11	9.0b	グループリソースの活性/非活性異常検出による再起動回数およびモニタリソースの異常検出による再起動回数をリセットするコマンド (clpregctrl コマンド) を追加しま

		した。
12	9.0b	clpmonctrl コマンドにモニタリソースの回復動作の回数を表示およびリセットするオプションを追加しました。

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の注意事項

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

- x86_64 のマシン上でBuilderおよび、WebManagerを動作させるには 32bit用のJava Runtimeを使用する必要があります。

ミラーディスクの要件について

- ダイナミックディスクは使用できません。ベーシックディスクを使用してください。
- GPT形式のディスクは使用できません。
- ミラーディスク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ディスク上の配置には特に制限はありませんが、データパーティションのサイズはバイト単位で完全に一致している必要があります。またクラスタパーティションには17MB以上の容量が必要です。
- データパーティションを拡張パーティション上の論理パーティションとして作成する場合は、両サーバとも論理パーティションにしてください。基本パーティションと論理パーティションでは同じサイズを指定しても実サイズが若干異なることがあります。
- 負荷分散のため、クラスタパーティションとデータパーティションは別のディスク上に作成することを推奨します（同じディスク上に作成しても動作に支障はありませんが、非同期ミラーの場合やミラーリングを中断している状態での書き込み性能が若干低下します）。
- ミラーリソースでミラーリングするデータパーティションを確保するディスクは、両サーバでディスクのタイプを同じにしてください。

  例)

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

- 「ディスクの管理」などで確保したパーティションサイズは、ディスクの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。このため、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と、データパーティションのサイズを完全に一致させることができない場合があります。このような問題を避けるため、データパーティションを確保するディスクは、RAID構成なども含め両サーバでHW構成を一致させることを推奨します。
- 両サーバでディスクのタイプやジオメトリを揃えらない場合は、ミラーディスクリソースを設定する前にclpvolszコマンドにより両サーバのデータパーティションの正確なサイズを確認し、もしサイズが一致しない場合は再度clpvolszコマンドを使用して大きいほうのパーティションを縮小してください。
- RAID構成のディスクをミラーリングする場合、ディスクアレイコントローラのキャッシュをWRITE THRUにすると書き込み性能の低下が大きくなるため、WRITE BACKでの使用をお勧めします。ただし、WRITE BACKで使用する場合は、バッテリーを搭載したディスク

アレイコントローラを用いるか、UPSを併用する必要があります。

- OSのページファイルがあるパーティションは、ミラーリングできません。

共有ディスクの要件について

- ダイナミックディスクは使用できません。ベーシックディスクを使用してください。
- ソフトウェアRAID(ストライプセット、ミラーセット、パリティ付ストライプセット)やボリュームセットは使用できません。
- GPT形式のディスクは使用できません。

NIC Link Up/Down監視リソース

NIC のボード、ドライバによっては、必要な DeviceIoControl()がサポートされていない場合がごく稀にあります。その場合には このモニタリソースは使用できません。このモニタリソースを使用する場合は、試用版ライセンス等を使用して事前に動作確認を行ってください。

ミラーリソースのwrite性能について

ミラーリソースのディスクミラーリングには同期ミラーと非同期ミラーの2種類の方式があります。

同期ミラーの場合、ミラーリング対象のデータパーティションへの書き込み要求毎に、両サーバのディスクへの書き込みを実施し、その完了を待ち合わせます。各サーバへの書き込みは並行して実施されますが、他サーバのディスクへの書き込みはネットワークを介して実施されるため、ミラーリングしない通常のローカルディスクに比べ書き込み性能が低下します。特にネットワークの通信速度が低く遅延が大きい遠隔クラスタ構成などの場合は大幅に性能が低下することになります。

非同期ミラーの場合、自サーバへの書き出しは即時実行しますが、他サーバへの書き出しは一旦ローカルキューに保存し、バックグラウンドで書き出します。他サーバへの書き出しの完了を待ち合わせないため、ネットワーク性能が低い場合も書き込み性能が大きく低下することはありません。ただし、非同期ミラーの場合も書き込み要求毎に更新データをキューに保存するため、ミラーリングしない通常のローカルディスクや共有ディスクに比べると、書き込み性能が低下します。このため、ディスクへの書き込み処理に高い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) には共有ディスクの使用を推奨します。

また、非同期ミラーの場合、書き込み順序は保証されますが、現用系サーバがダウンした場合に最新の更新分が失われる可能性があります。このため、障害発生直前の情報を確実に引き継ぐ必要がある場合は、同期ミラーか共有ディスクを用いる必要があります。

非同期ミラーの履歴ファイルについて

非同期モードのミラーディスクでは、メモリ上のキューに記録しきれない書き込みデータは、履歴ファイル格納フォルダとして指定されたフォルダに履歴ファイルとして一時的に記録されます。この履歴ファイルは、履歴ファイルのサイズ制限を設定していない場合、指定されたフォルダに制限なく書き出されます。このような設定の場合、回線速度が業務アプリケーションのディスク更新量に比べて低すぎると、リモートサーバへの書き込み処理がディスク更新に追いつかず、履歴ファイルでディスクが溢れてしまいます。このため、遠隔クラスタ構成でも業務APのディスク更新量に合わせて十分な速度の通信回線を確保する必要があります。

また、長時間通信帯域が狭まったりディスク更新が連続して発生し、履歴ファイル格納フォルダが溢れた場合に備え、履歴ファイルの書き出し先に指定するドライブには十分な空き容量を確保し、履歴ファイルサイズ制限を設定するか、システムドライブとは別のドライブを指定する必要があります。

複数の非同期ミラー間のデータ整合性について

非同期モードのミラーディスクでは、現用系のデータパーティションへの書き込みを、同じ順序で待機系のデータパーティションにも実施します。

ミラーディスクの初期構築中やミラーリング中断後の復帰（コピー）中以外は、この書き込み順序が保証されるため、待機系のデータパーティション上にあるファイル間のデータ整合性は保たれます。

しかし、複数のミラーディスクリソース間では書き込み順序が保証されませんので、例えばデータベースのデータベースファイルとジャーナル（ログ）ファイルのように、一方のファイルが他方より古くなるとデータの整合性が保てないファイルを複数の非同期ミラーディスクに分散配置すると、サーバダウン等でフェイルオーバした際に業務アプリケーションが正常に動作しなくなる可能性があります。

このため、このようなファイルは必ず同一の非同期ミラーディスク上に配置してください。

マルチブートについて

他のブートディスクで起動すると、ミラーや共有ディスクのアクセス制限が外れてしまい、ミラーディスクの整合性保証や共有ディスクのデータ保護ができなくなるため、これらのリソースを使用している場合はマルチブートを使用しないでください。

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。

ファイルシステムについて

OS をインストールするパーティション、共有ディスクのディスクリソースとして使用するパーティション、ミラーディスクリソースのデータパーティションのファイルシステムは NTFS を使用してください。

通信ポート番号

CLUSTERPRO では、デフォルトで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。これらのポート番号には、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］

From			To		備考
サーバ	自動割り当て [2]	→	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	29003/UDP	アラート同期
サーバ	自動割り当て	→	サーバ	29004/TCP	ディスクエージェント間通信
サーバ	自動割り当て	→	サーバ	29005/TCP	ミラードライバ間通信
サーバ	29106/UDP	→	サーバ	29106/UDP	ハートビート

［サーバ・クライアント間］

From			To		備考
クライアント	自動割り当て	→	サーバ	29007/TCP 29007/UDP	クライアントサービス通信

［サーバ・WebManager 間］

From			To		備考
WebManager	自動割り当て	→	サーバ	29003/TCP	http 通信

［統合 WebManager でブラウザが接続しているサーバ・管理対象のサーバ間］

From			To		備考
ブラウザが接続しているサーバ	自動割り当て	→	サーバ	29003/TCP	http 通信

ミラーコネクト監視リソースを使用する場合、CLUSTERPRO はサーバ間で ping による疎通確認を行うため、icmp パケットを通すように設定する必要があります。ミラーコネクト監視リソースを使用する場合は、ファイアウォールの設定を変更して、サーバ間で ping による疎通確認ができるようにしてください。

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。タイムサーバ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

- CLUSTERPROによるアクセス制限を行っていない状態で、共有ディスクに接続されたサーバを複数起動すると、共有ディスク上のデータが破壊される危険があります。アクセス制限をかける前は、必ずいずれか一台のみ起動するようにしてください。
- ネットワークパーティション解決方式として共有ディスク方式を用いる場合、DISKネットワークパーティション解決リソースが使用する17MB以上のRAW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）を共有ディスク上に作成してください。
- ディスクリソースとしてサーバ間のデータ引き継ぎに使用するパーティション（切替パーティション）はNTFSでフォーマットしてください。

[2] 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。

- 共有ディスク上の各パーティションには、全てのサーバで同一のドライブ文字を設定してください。
- 共有ディスク上のパーティション作成やフォーマットは、いずれか一台のサーバからのみ行います。各サーバで再作成・再フォーマットを行う必要はありません。ただし、ドライブ文字は各サーバで設定する必要があります。
- サーバの再インストール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ォーマットは行わないでください。パーティションの確保やフォーマットを行うと共有ディスク上のデータは削除されます。

ミラーディスク用のパーティションについて

- ミラー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)として、17MB以上のRAWパーティションを各サーバのローカルディスクに作成してください。
- ミラーリング対象のパーティション(データパーティション)を各サーバのローカルディスクに作成し、NTFSでフォーマットしてください（既存のパーティションをミラーリングする場合、パーティションを作り直す必要はありません）。
- データパーティションのサイズは、両サーバで等しくなるように設定してください。
- クラスタパーティションとデータパーティションには、両サーバで同じドライブ文字を設定してください。

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 [3]。

- 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

ネットワークの確認

- インタコネクトやミラー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。
- ipconfigコマンドやpingコマンドを使用してネットワークの状態を確認してください。
  - public-LAN (他のマシンと通信を行う系)
  - インタコネクト専用 LAN(CLUSTERPRO のサーバ間接続に使用する系)
  - ミラーコネクト LAN(インタコネクトと共用)
  - ホスト名
- クラスタで使用するフローティングIPリソースのIPアドレスは、OS側への設定は不要です。

[3] BOOT時に選択するOSが一つしかない場合、起動待ち時間を設定しても無視される場合があります。この場合、boot.iniファイルを編集して、[Operating System]セクションに2つ目のエントリを追加してください。2つ目のエントリは1つ目のエントリのコピーで問題ありません。

ESMPRO/AutomaticRunningControllerとの連携について

ESMPRO/AutomaticRunningController(以降 ESMPRO/AC と称す)と連携動作させる場合は、CLUSTERPRO の構築/設定に次の留意事項があります。これらが満たされていないと、ESMPRO/AC との連携機能が正しく動作しないことがあります。

- ESMPRO/ACと連携動作させるためには、CLUSTERPRO Xに、内部Ver9.02(アップデート管理番号:CPRO-XW010-01)以降の適用が必要です。
- x64 EditionのOS上ではESMPRO/ACとの連携機能は動作しません。
- ネットワークパーティション解決リソースとして、DISK方式のリソースのみを単独で指定することはできません。DISK方式を指定する場合は、必ずPING方式、COM方式など、他のネットワークパーティション解決方式のリソースと組み合わせて指定してください。
- ディスクTUR監視リソースを作成する際は、「最終動作」の設定値はデフォルト(「何もしない」)から変更しないでください。
- ディスクRW監視リソースを作成する際、「ファイル名」の設定値に共有ディスク上のパスを指定する場合は、「監視タイミング」の設定値はデフォルト(活性時)から変更しないでください。
- 復電後再起動した際、次のアラートがCLUSTERPROのマネージャ上にエントリされることがあります。上記の設定により、実際の動作に支障はありませんので無視してください。
  - ID:18
    モジュール名:nm
    メッセージ:リソース<DiskNPのリソース名>の起動に失敗しました。(サーバ名:xx)
  - ID:1509
    モジュール名:rm
    メッセージ:監視 <ディスクTUR監視リソース名> は異常を検出しました。 (4 : デバイスオープンに失敗しました。監視先ボリュームのディスク状態を確認してください。)
- ESMPRO/ACの設定方法、留意事項等については、「CLUSTERPRO X for Windows PPガイド」のESMPRO/ACの章の記述を参照してください。

CLUSTERPRO の構成情報作成時

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。

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。
実際に業務で使用する際には、「何もしない」は設定しないように注意してください。

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モニタリソースへのポーリング時間を算出し、モニタ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。

ディスク監視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。30秒より小さい値を設定すると、CLUSTERPROのパフォーマンスに影響を与えるおそれがあります。

ハートビートリソースの設定について

- カーネルモード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(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)。
- ハートビートタイムアウト時間はOS再起動の所要時間より短くする必要があります。この

条件を満たさない場合、クラスタ内の一部のサーバがリブートした際に、それを他のサーバが正しく検出できず、リブート後に動作異常が発生する場合があります。

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。

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、アプリケーションリソースなど) 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、WebManager やコマンドによる以下の操作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。

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

ミラーディスク使用時は、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するとミラーブレイクが発生します。

ネットワークパーティション状態からの復旧

ネットワークパーティションが発生している状態では、クラスタを構成するサーバ間で互いの状態が確認できないため、この状態でグループの操作(起動/停止/移動)を行ったり、サーバを再起動すると、サーバ間でクラスタの状態についての認識にずれが生じます。このように異なる状態認識のサーバが複数起動している状態でネットワークが復旧すると、その後のグループ操作が正しく動作しなくなりますので、ネットワークパーティション状態にある間は、ネットワークから切り離された(クライアントと通信できない)方のサーバはシャットダウンするか、CLUSTERPRO Server サービスを停止しておき、ネットワークが復旧してから再起動してクラスタに復帰してください。万一、複数のサーバが起動した状態でネットワークが復旧した場合は、クラスタの状態認識が異なるサーバを再起動することにより、正常状態に復帰できます。

なお、ネットワークパーティション解決リソースを使用している場合は、ネットワークパーティションが発生しても、通常はいずれかの(あるいは全ての)サーバが緊急シャットダウンして、互いに通信できないサーバが複数起動するのを回避します。緊急シャットダウンされたサーバを手動で再起動した場合や、緊急シャットダウン時の動作を再起動に設定していた場合も、再起動したサーバは再度緊急シャットダウンされます(Ping 方式や多数決方式の場合はCLUSTERPRO Server サービスが停止されます)。ただし、DISK 方式で複数のディスクハートビート用パーティションを使用している場合、ディスクパス障害によりディスクを介した通信ができない状態でネットワークパーティションが発生すると、両サーバが保留状態で動作を継続する場合があります。

WebManagerについて

- WebManager の「クライアントデータ更新方法」の設定が、「Polling」に設定されている場合、WebManagerで表示される内容は定期的に更新され、状態が変化しても即座には表示に反映されません。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。
  WebManager の「クライアントデータ更新方法」の設定が、「Polling」に設定されている場合、次回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内容を再取得してください。「Realtime」に設定されている場合、自動的に最新の内容に更新されます。
- 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、Webブラウザを一旦終了し、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。
  Javaのキャッシュ (ブラウザ側のキャッシュではありません) をクリアし、ブラウザを起動してください。
- Javaのアップデートを行なった場合、起動している全てのブラウザを一旦終了してください。
  Javaのキャッシュ(ブラウザ側のキャッシュではありません)をクリアして、ブラウザを起動

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 1.0 for Windows 以外の Builder
  - CLUSTERPRO for Linux のトレッキングツール
  - CLUSTERPRO for Windows Value Edition 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。

CLUSTERPRO Disk Agentサービスについて

CLUSTERPRO Disk Agent サービスは停止しないでください。停止した場合、手動での起動はできません。OS を再起動し CLUSTERPRO Disk Agent サービスを起動しなおす必要があります。

ミラー構築中のクラスタ構成情報の変更について

ミラー構築中(初期構築を含む)はクラスタ構成情報を変更しないでください。クラスタ構成情報を変更した場合、ドライバが不正な動作を行う場合があります。

chkdskコマンドとデフラグについて

ディスクリソースで制御している共有ディスク上の切替パーティションや、ミラーディスクリソースでミラーリングしているデータパーティションに対して、chkdsk コマンドやデフラグを実行する場合、リソースが起動済みのサーバで実行する必要があります。起動していない状態では、アクセス制限により実行できません。

また、chkdsk コマンドを修復モード(/f オプション)で実行する場合、対象パーティション上のファイルやフォルダが開かれていると実行が失敗するため、フェイルオーバグループを停止し、対象のディスクリソース/ミラーディスクリソースを単体起動した状態で実行します。もし対象パーティションに対して監視を行うディスク RW 監視リソースがある場合は、このモニタリソースを一時停止しておく必要があります。

インデックスサービスについて

インデックスサービスのカタログに共有ディスク/ミラーディスク上のディレクトリを作成して、共有ディスク/ミラーディスク上のフォルダに対してインデックスを作成する場合、インデックスサービスを手動起動に設定して、共有ディスク/ミラーディスクの活性後に起動するようにCLUSTERPRO から制御する必要があります。インデックスサービスを自動起動にしていると、インデックスサービスが対象ボリュームを OPEN することにより、その後の活性化処理においてマウント処理が失敗し、アプリケーションやエクスプローラからのディスクアクセスが「パラメータが間違っています」（エラーコード 87）というエラーで失敗します。

旧バージョンとの互換性

旧バージョン互換機能について

以下の機能を使用する場合、クラスタ名、サーバ名、グループ名は、従来バージョンの命名規則に従って設定する必要があります。

- CLUSTERPRO Alert Service (通報アイコン)
- CLUSTERPROクライアント
- ESMPRO/AC連携機能
- ESMPRO/SM連携機能
- 仮想コンピュータ名リソース
- 互換API
- 互換コマンド

従来バージョンの命名規則は以下の通りです。

- クラスタ名
  - 15文字以内
  - 使用可能な文字は、半角英数字、ハイフン(-)、アンダーバー(_)です。
  - PRNなどのDOS入出力デバイス名は指定しないでください。
  - 大文字、小文字を区別しません。
- サーバ名
  - 15文字以内
  - 使用可能な文字は、半角英数字、ハイフン(-)、アンダーバー(_)です。
  - 大文字、小文字を区別しません。
- グループ名
  - 15文字以内
  - 使用可能な文字は、半角英数字、ハイフン(-)、アンダーバー(_)です。
  - PRNなどのDOS入出力デバイス名は指定しないでください。
  - 大文字、小文字を区別しません。

互換APIについて

互換 API は、CLUSTERPRO Ver8.0 以前で使用可能であった API を指します。互換 API は CLUSTERPRO X でも使用可能ですが、以下の制限事項があります。

下記に示すリソースのみ対応しています。その他のリソースは設定しても互換 API から参照することはできません。
- ディスクリソース
- ミラーディスクリソース
- 仮想コンピュータ名リソース
- 仮想 IP リソース
- プリントスプーラリソース

クラスタ名、サーバ名、グループ名は、従来バージョンの規則に従い設定する必要があります。従来バージョン規則外の名称を指定された場合は、互換 API で参照することはできません。

Builder で指定されたリソース名を使用して、互換 API を使用することはできません。

クラスタイベントの発生タイミングは、完全互換ではありません。イベントの種類は同じですが、通知されるイベントの数、順序は従来バージョンと異なる場合があります。

常駐プロセスから互換 API を使用している場合、[CLUSTERPRO Server]サービスの停止→再起動時に、ArmTerminateApi → ArminitializeApi を実行し、互換 API の再初期化を行う必要があります。原則として、スクリプトリソースの開始・終了スクリプトでプロセスを起動・停止するように設定してください。

Ver3.0 互換 I/F は使用できません。

スクリプトファイルについて

旧バージョンで使用していたスクリプトファイルを移植する場合、環境変数名の最初の ”ARMS_” を ”CLP_” に置換してください。

例） IF "*%ARMS_EVENT%*" == "START" GOTO NORMAL

↓

IF "*%CLP_EVENT%*" == "START" GOTO NORMAL

付録

- 付録 A　用語集
- 付録 B　索引

付録 A　　用語集

英数字

CLUSTERパーティション　ミラーディスクに設定するパーティション。ミラー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)

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[4]　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

[4] 仮想IPアドレスはwindows版でのみ使用する概念になります。

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
ミラーディスクに設定するデータ用のパーティション
(関連) CLUSTER 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
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

ミラーコネクト
データミラー型クラスタ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