

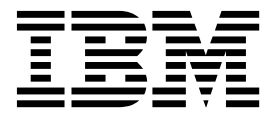
IBM PowerHA SystemMirror for AIX
Standard Edition
バージョン 7.2.2

PowerHA SystemMirror の概 念

IBM

IBM PowerHA SystemMirror for AIX
Standard Edition
バージョン 7.2.2

PowerHA SystemMirror の概 念

The IBM logo, consisting of the letters "IBM" in a bold, sans-serif font, with each letter formed by eight horizontal bars of varying lengths.

お願い

本書および本書で紹介する製品をご使用になる前に、123 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM PowerHA SystemMirror 7.2.2 Standard Edition for AIX および新しい版で明記されていない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： IBM PowerHA SystemMirror for AIX
Standard Edition
Version 7.2.2
PowerHA SystemMirror concepts

© Copyright IBM Corporation 2017.

目次

本書について	v	PowerHA SystemMirror でサポートされるハードウェア	55
強調表示	v	PowerHA SystemMirror クラスター・ソフトウェア	56
AIX での大/小文字の区別	v	PowerHA SystemMirror ノードのソフトウェア・コンポーネント	56
ISO 9000	v	クラスター・マネージャー	57
関連情報	v	PowerHA SystemMirror ソフトウェア・コンポーネント	58
PowerHA SystemMirror の概念	1	補足クラスター・ソフトウェア	65
PowerHA SystemMirror の概念に関する新着情報	1	アプリケーションの可用性の確保	65
新規の機能、および廃止された機能	1	概説: アプリケーションの可用性	65
PowerHA SystemMirror for AIX	1	PowerHA SystemMirror クラスター内の単一障害点の除去	66
AIX の高可用性クラスター化	2	予定されたダウン時間を PowerHA SystemMirror で最小限に抑える	77
PowerHA SystemMirror での Cluster Aware AIX の使用	4	予定外のダウン時間を最小限に抑える	84
可用性のコストと利点の関係	8	テークオーバー時間を最小限に抑える	85
AIX ソフトウェアによる可用性の強化	9	災害時回復を最大限に行う	86
PowerHA SystemMirror クラスターの物理コンポーネント	11	クラスター・イベント	86
PowerHA SystemMirror の目的: 予定されたダウン時間の解消	18	PowerHA SystemMirror クラスター構成	90
PowerHA SystemMirror クラスター・ノード、ネットワーク、およびハートビートの概念	18	スタンバイ構成	90
ノード	18	テークオーバー構成	93
サイト	19	多層アプリケーションを使用したクラスター構成	97
分割ポリシー	20	リソース・グループ・ロケーション依存関係を持つクラスター構成	99
マージ・ポリシー	22	災害時回復のためのサイト間 LVM ミラー構成	100
分割ポリシーおよびマージ・ポリシーのタイブレーカー・オプション	24	動的 LPAR を使用したクラスター構成	101
検疫ポリシー	25	PowerHA SystemMirror 構成プロセスおよび機能	105
PowerHA SystemMirror マルチサイト・ソリューション	28	PowerHA SystemMirror に渡す情報	105
クラスター・ネットワーク	32	PowerHA SystemMirror によってディスクカバーされる情報	106
IP アドレス・テークオーバー	35	クラスター構成オプション: 標準および拡張	106
TCP/IP ネットワークおよび Storage Area Network を使用したハートビート	37	クラスター・セキュリティー	107
PowerHA SystemMirror マルチキャスト	39	インストール・ツール、構成ツール、および管理ツール	107
PowerHA SystemMirror リソースおよびリソース・グループ	41	モニター・ツール	113
クラスター・リソースの識別、およびクラスター・リソースの可用性の保持	41	トラブルシューティング・ツール	119
クラスター・リソースのタイプ	42	クラスター・テスト・ツール	121
クラスター・リソース・グループ	45	特記事項	123
リソース・グループのポリシーおよび属性	48	プライベート・ポリシーに関する考慮事項	125
リソース・グループ依存関係	50	商標	125
サイトおよびリソース・グループ	54	索引	127
ログ・アナライザー	55		

本書について

本書では、PowerHA® SystemMirror® for AIX® ソフトウェアについて紹介します。この情報は、オペレーティング・システム付属の文書 CD にも収録されています。

強調表示

本書では、以下の強調表示規則を使用します。

太字	システムによって名前が事前に定義されているコマンド、サブルーチン、キーワード、ファイル、構造、ディレクトリー、およびその他の項目を示します。また、ユーザーが選択するボタン、ラベル、アイコンなどのグラフィカル・オブジェクトも示します。
イタリック	実際の名前または値をユーザーが指定する必要があるパラメーターを示します。
モノスペース	特定のデータ値の例、画面に表示されるものと同様のテキスト例、プログラマーが作成するものと同様のプログラム・コード部分の例、システムからのメッセージ、実際に入力する必要がある情報などを示します。

AIX での大/小文字の区別

AIX オペレーティング・システムは、すべてケース・センシティブとなっています。これは、英大文字と小文字が区別されるということです。例えば、**ls** コマンドを使用するとファイルをリスト表示できます。LS と入力した場合、そのようなコマンドはないという応答がシステムから返ってきます。同様に、**FILEA**、**FiLea**、および **filea** は、同じディレクトリーにある場合でも、3 つの異なるファイル名です。予期しない処理が実行されないように、常に正しい大/小文字を使用するようにしてください。

ISO 9000

当製品の開発および製造には、ISO 9000 登録品質システムが使用されました。

関連情報

- PowerHA SystemMirror バージョン 7.2.2 for AIX PDF 資料は、『PowerHA SystemMirror 7.2.2 の PDF』のトピックで入手可能です。
- PowerHA SystemMirror バージョン 7.2.2 for AIX リリース・ノートは、『PowerHA SystemMirror 7.2.2 リリース・ノート』のトピックで入手可能です。

PowerHA SystemMirror の概念

ここでは、AIX オペレーティング・システム用の PowerHA SystemMirror ソフトウェアを使用する前に理解しておく必要がある重要な概念について説明します。

PowerHA SystemMirror の概念に関する新着情報

PowerHA SystemMirror の概念のトピック集について、新しい情報または大幅に変更された情報を説明します。

新規情報または変更情報の参照方法

この PDF ファイルでは、左マージンに新規および変更情報を示すリビジョン・バー (1) が表示される場合があります。

2017 年 12 月

このトピック集に対する更新の要約を以下に示します。

- PowerHA SystemMirror バージョン 7.2.2 for AIX のログ分析機能をサポートするために 55 ページの『ログ・アナライザー』トピックが追加されました。

2016 年 12 月

このトピック集に対する更新の要約を以下に示します。

- 以下のトピックが追加されました。
 - 25 ページの『検疫ポリシー』
 - 25 ページの『アクティブ・ノード停止ポリシー』
 - 26 ページの『ディスク・フェンシング・ポリシー』
- 以下のトピックでポリシー情報が更新されました。
 - 20 ページの『分割ポリシー』
 - 22 ページの『マージ・ポリシー』
 - 24 ページの『分割ポリシーおよびマージ・ポリシーのタイ・ブレーカー・オプション』
- 115 ページの『クラスター・ステータス・ユーティリティ (clstat)』トピックで、IP バージョン 6 ループバック・アドレスに関する潜在的な問題についての情報が更新されました。

新規の機能、および廃止された機能

このリリースの新規機能、およびこのリリースで廃止された機能のリストについては、IBM® 製品発表資料を参照してください。また、製品ソフトウェアとともにインストールされる `/usr/es/sbin/cluster/WhatsNewInThisRelease` ファイルにも、追加情報が記載されています。

PowerHA SystemMirror for AIX

ここでは、高可用性とクラスター化の概念を説明するとともに、PowerHA SystemMirror クラスターの図を示して、機能の観点から PowerHA SystemMirror クラスターを説明します。

AIX の高可用性クラスター化

IBM の PowerHA SystemMirror ソフトウェアは、主幹業務アプリケーションのハードウェア障害やソフトウェア障害から迅速に回復できるようにする低コストの商用コンピューティング環境を提供します。

PowerHA SystemMirror ソフトウェアを使用すると、重要なリソースを継続的に使用可能にすることができます。例えば、PowerHA SystemMirror クラスターは、クライアント・アプリケーションにサービスするデータベース・サーバー・プログラムを実行できます。クライアントは、共用外部ディスクに格納されたデータベースにアクセスすることによって、要求に応えるサーバー・プログラムにクエリーを送信します。

このハイ・アベイラビリティ・システムは、カスタム・ソフトウェアと業界標準のハードウェアを組み合わせ、システム、コンポーネント、またはアプリケーションに障害が発生したときにサービスを素早く復元してダウン時間を最小限にします。サービスの復旧は、瞬時というわけではありませんが、短時間 (普通は 30 から 300 秒) のうちに行われます。

PowerHA SystemMirror クラスターでは、このようなアプリケーションの可用性を保証するために、アプリケーションは PowerHA SystemMirror 制御下に置かれます。PowerHA SystemMirror は、クラスター内のコンポーネントに障害が発生しても、アプリケーションがクライアント・プロセスにとって使用可能なままとなるような方法をとります。コンポーネントに障害が発生した場合でも可用性を保証するため、PowerHA SystemMirror はアプリケーションを (そのアプリケーションに確実にアクセスするリソースと共に) クラスター内の別のノードに移します。

高可用性とハードウェア可用性

高可用性 は、単純なハードウェアの可用性と混同されることがあります。耐障害性がある冗長システム (RAID など) や動的スイッチング・テクノロジー (DLPAR など) は、一部のハードウェア障害を回復しますが、複雑なアプリケーションを可用性の高いまま保持するために必要なすべての範囲のエラー検出や回復までには行いません。

現代の複雑なアプリケーションは、以下のすべてのコンポーネントにアクセスできる必要があります。

- ノード (CPU、メモリー)
- ネットワーク・インターフェース (ネットワーク・トポロジーにおける外部デバイスも含む)
- ディスク・デバイスまたはストレージ・デバイス

ダウン時間の原因に関する最近の調査によると、実際のハードウェア障害が原因となったケースは、計画外の停止の中のほんのわずかなパーセンテージに過ぎません。その他には、以下の原因があります。

- オペレーターのエラー
- 環境の問題
- アプリケーションおよびオペレーティング・システムのエラー

単に信頼性が高く回復可能なハードウェアであれば、構成におけるこれらのさまざまな障害すべてを防止できる、というわけではありません。これらの多様な要素の高可用性、その結果としてアプリケーションの高可用性を保つには、以下が必要となります。

- アプリケーションが依存するリソースのアクセスや操作の物理的および論理的手順についての詳細かつ完全な計画。このような手順があれば、最初の段階で障害を回避するのに役立ちます。
- エラーの検出および回復を自動化するモニターおよび回復パッケージ。
- アプリケーションを使用可能なまま保持しながらクラスター構成のハードウェア面およびソフトウェア面を保守するための、よく制御されたプロセス。

高可用性と耐障害性

耐障害性と高可用性は違います。耐障害性の環境ではサービス中断は発生しませんが、コストがきわめて高くなります。それに対し、高可用性の環境では最小限のサービス中断が発生します。

耐障害性テクノロジーでは、プロセッサ、メモリー・ボード、電源装置、入出力サブシステム、記憶サブシステムなどに障害が起きても、特殊ハードウェアを使用してハードウェア障害を検出し、予備のハードウェア・コンポーネントに瞬時に切り替えます。このカットオーバーは確かにシームレスであり、処理を停止せずにサービスを提供できますが、予備のコンポーネントが処理を行っていないため、ハードウェア・コストおよびパフォーマンスの両方に大きな代償が払われています。さらに重要なことは、耐障害性モデルでは、ダウン時間の原因として今のところ最も一般的なソフトウェア障害に対処できないことです。

ハイ・アベイラビリティとは、一連の物理コンポーネントの複製による可用性ではなく、必須のサービスを保証するために協調するシステム規模の共用リソースの集合による可用性を示します。システム、コンポーネント、またはアプリケーションに障害が発生したときに必須のサービスを素早く復元することによりダウン時間を最小限にするために、ハイ・アベイラビリティは、ソフトウェアと業界標準のハードウェアを組み合わせます。サービスは瞬時ではありませんが、一般に 1 分未満で素早く復元されます。

耐障害性のために高いコストを払うよりも、ダウン時間が少しかかっても、ハイ・アベイラビリティを望む導入先は少なくありません。また、大部分の可用性の高い構成では、バックアップ・プロセッサを通常の操作に利用することができます。

ハイ・アベイラビリティ・システムは、障害が発生した場合に、早急に復元しなければならないが短時間の中断であれば許容できるアプリケーションにとっては優れたソリューションです。数秒のダウン時間も許容できないほどタイム・クリティカルなアプリケーションを運用している業界もあります。しかし、データベースがダウンしたとしても、短時間であれば許容できる業界も数多くあります。こうした業界の場合、PowerHA SystemMirror を使用すれば、全体的な冗長性を確保することなく、必要とされるサービスの継続性が得られます。

PowerHA SystemMirror の役割

PowerHA SystemMirror には多くの利点があります。

PowerHA SystemMirror には、ユーザーを支援する以下の機能があります。

- PowerHA SystemMirror 計画プロセスおよび文書には、可用性の高い PowerHA SystemMirror クラスターをインストールし保守するための最良実例のヒントやアドバイスがあります。
- クラスターが稼働中であれば、PowerHA SystemMirror はアプリケーションが依存するすべてのリソースの自動化モニターおよび回復機能を提供します。
- PowerHA SystemMirror には、アプリケーションをクライアントに対して使用可能なまま保持しながらクラスターを保守するツールのフル・セットがあります。

PowerHA SystemMirror を使用すれば、以下のことが可能になります。

- 基本的な 2 ノード・クラスターを迅速かつ簡単にセットアップする。このために標準的なクラスター初期構成 SMIT パス、またはアプリケーション構成アシスト (Smart Assist) を使用します。
- クラスター・テスト・ツールを使用して、ご使用の PowerHA SystemMirror 構成をテストする。ノード、ネットワークがアクセス不可になったときなどの、一連の特定環境下でクラスターがどのように振る舞うかを評価できます。
- PowerHA SystemMirror 環境の SPOF (単一障害点) を除去することにより、アプリケーションのハイ・アベイラビリティを保証する。
- AIX で使用可能なハイ・アベイラビリティ機能を利用する。

- クラスタがコンポーネント障害を処理する方法を管理する。
- クラスタ通信を保護する。
- PowerHA SystemMirror コンポーネントをモニターして、起こり得る問題を診断する。

関連情報:

 PowerHA SystemMirror for AIX

アプリケーション・クラスター

アプリケーション・クラスター は、同じネットワークに属してディスク・リソースを共有する疎結合マシンのグループです。

クラスター内では、複数のサーバー・マシンが協調して、クライアントに一連のサービスおよびリソースを提供します。

2 台以上のサーバーをクラスター化してクリティカル・アプリケーションをバックアップするのは、コスト効果のあるハイ・アベイラビリティの選択肢です。サイトのコンピューティング・パワーをより有効に使用するとともに、ハードウェアまたはソフトウェアの障害が発生した場合も、中断時間を最小限に抑えてクリティカル・アプリケーションを再開できます。

また、アプリケーション・クラスターにより、段階的でスケーラブルな拡張パスが可能です。増大する作業負荷を分担させるために、クラスターにはプロセッサを簡単に追加できます。また、クラスター内の 1 つ以上のプロセッサを、より強力なモデルにアップグレードすることもできます。耐障害性戦略を使用する場合は、プロセッサを 2 つ 追加する必要があります。プロセッサの 1 つは、予備のバックアップとして、通常の操作では処理を行いません。

PowerHA SystemMirror での Cluster Aware AIX の使用

PowerHA SystemMirror は、AIX オペレーティング・システムでサポートされる主要なクラスター機能の最上位に構築されます。PowerHA SystemMirror は、Cluster Aware AIX (CAA) 機能に対応した AIX のすべてのエディションでサポートされます。

CAA および PowerHA SystemMirror では、ディスクおよびノードの追跡に汎用 ID (UID および UUID) が使用されます。UID も UUID も動的に変更することはサポートされていません。通常は、多くの環境において UID と UUID は不変です。ただし、UID や UUID を変更できる既知のシナリオ (オペレーティング・システムの再インストールなど) があります。UID や UUID を変更する場合は、すべての UID と UUID が更新されるように CAA クラスタを除去して作成し直す必要があります。

AIX バージョン 7.2 以降または IBM AIX 7.1 (テクノロジー・レベル 4 適用) 以降では、CAA はネットワーク障害が発生してから 20 秒 (デフォルト値) 後にその障害を検出して処理します。デフォルト値を 20 秒から変更するには、`clmgr modify cluster NETWORK_FAILURE_DETECTION_TIME=<xxx>` コマンドを実行します。xxx は、5 から 590 までの範囲の時間 (秒) です。

ここでは、PowerHA SystemMirror ソリューション・スタックを構築するための基盤として使用される Cluster Aware AIX の主なコンポーネントについて説明します。

ハートビートの管理

デフォルトで、PowerHA SystemMirror はハートビートにユニキャスト通信を使用します。また、ユニキャストの代わりにマルチキャスト通信も構成できます。マルチキャストの場合、オプションでマルチキャスト・アドレスを選択するか、Cluster Aware AIX (CAA) が自動的にマルチキャスト・アドレスを割り当てるようにすることができます。クラスターの構成時にマルチキャスト・アドレスを指定したり、ネットワーク環境に基づいた構成時に Cluster Aware AIX (CAA)

によってマルチキャストをセットアップしたりすることができます。クラスター通信は、通信用の複数の冗長パスを介した通信によって行われます。以下の通信用の冗長パスは、クラスターの区分化を起こしにくい堅固なクラスター化基盤を提供します。

TCP/IP ネットワーク

PowerHA SystemMirror と Cluster Aware AIX では、クラスター通信に使用できるすべてのネットワーク・インターフェースを使用します。デフォルトでは、これらのすべてのインターフェースがディスカバーされ、正常性の管理とその他のクラスター通信に使用されます。アプリケーション可用性のために使用しないインターフェースを除去するには、PowerHA SystemMirror 管理インターフェースを使用することができます。また、使用しないインターフェースを PowerHA SystemMirror でプライベート・インターフェースとして定義することもできます。

SAN ベースの通信

CAA は、一定数のアダプターに対するストレージ・エリア・ネットワーク (SAN) ファブリック・ベースのクラスター通信 (ハートビートを含む) をサポートします。このタイプのハートビートはオプションであり、ほとんどの環境で動作しない可能性があります。その原因として、SCSI プロトコルを使用してパケットをクライアント間で移動できるようにするネットワーク・ゾーニング要件があります。

中央クラスター・リポジトリ・ベースの通信

クラスター正常性およびその他のクラスター通信は中央リポジトリ・ディスクを使用して実現されます。PowerHA SystemMirror 7.2.0 以降には、障害が発生したりリポジトリ・ディスクを自動的にバックアップ・リポジトリ・ディスクに交換する Automatic Repository Disk Replacement (ARR) 機能が用意されています。ARR 機能は、PowerHA SystemMirror を使用してバックアップ・リポジトリ・ディスクを構成および識別する場合にのみ使用可能です。

ネットワーク・インターフェース障害検出時間

PowerHA SystemMirror は CAA に依存して、ネットワーク・インターフェースおよびノードをモニターしてそれらの障害を検出します。IBM AIX 7.1 (テクノロジー・レベル 4 適用) 以前では、CAA は固定時間 (5 秒) 以内にネットワーク障害を検出していました。それらのバージョンの AIX オペレーティング・システムでハードウェア障害が発生した場合、その障害は即時に報告されました。このタイプの報告は即時障害処理と呼ばれます。AIX オペレーティング・システムにおけるこの検出/報告処理は、PowerHA SystemMirror バージョン 6.1 が障害を報告および検出する方法とは異なります。PowerHA SystemMirror 6.1 では、完全ネットワーク障害検出時間になるまで障害は示されません。この処理は、寛大障害検出に基づく全待機時間と呼ばれます。

AIX バージョン 7.2 以降または IBM AIX 7.1 (テクノロジー・レベル 4 適用) 以降では、**clmgr** コマンドで `NETWORK_FAILURE_DETECTION_TIME` オプションを使用してネットワーク・インターフェースの障害検出時間を設定できます。`NETWORK_FAILURE_DETECTION_TIME` オプションのデフォルト値は 20 秒です。AIX バージョン 7.2 以降または IBM AIX 7.1 (テクノロジー・レベル 4 適用) 以降では、障害検出時間の全待機期間の後で障害検出処理が行われます。それらのバージョンの AIX オペレーティング・システムでは、即時障害検出処理は使用されません。

`NETWORK_FAILURE_DETECTION_TIME` オプションのデフォルト値を 20 秒から変更するには、**clmgr modify cluster NETWORK_FAILURE_DETECTION_TIME=<xxx>** コマンドを実行します。`xxx` は以下のいずれかの値です。

- 0 この値を指定した場合に、クラスターが同期されると、ネットワーク障害は発生後の 5 秒後に検出され、即時障害検出処理が使用されます。このオプションは IBM AIX 7.1 (テクノロジー・レベル 4 適用) 以前で使用されました。

5 秒から 590 秒まで

この範囲の値を指定した場合に、クラスターが同期されると、指定の時間が経過した後でネットワーク障害が検出され、全待機時間処理が使用されます。

ノード障害検出時間

PowerHA SystemMirror および CAA は、ネットワーク通信およびディスク通信からハートビートが失われたときにクラスター内のパートナー・ノードの障害を検出できます。これらの通信チャネルが失われると、設定された期間だけモニターが有効になります。このモニターはノード障害検出時間と呼ばれます。

ノード障害検出時間を構成するには、以下のいずれかの選択肢を使用できます。

SMIT ノード障害検出時間を構成するには、以下の手順を実行します。

1. コマンド行から **smit sysmirror** と入力します。
2. SMIT インターフェイスで「ユーザー定義クラスター構成」 > 「クラスター・ノードおよびネットワーク」 > 「クラスターの管理」 > 「クラスター・ハートビート設定」を選択し、Enter を押します。
3. 必要なすべてのフィールドに入力して、Enter を押します。

コマンド行

コマンド行で **clmgr modify cluster HEARTBEAT_FREQUENCY=<v1> GRACE_PERIOD=<v2>** コマンドを実行します。v1 と v2 は値 (秒) です。

HEARTBEAT_FREQUENCY オプションはノード通信タイムアウト値です。この値は、CAA が、パートナー・ノードで障害が発生したかどうかを判別するために、次の処理ステップを実行する前にパートナー・ノードからパケットを受け取るまで待機する時間 (秒) です。

HEARTBEAT_FREQUENCY オプションの有効な値は 20 秒から 600 秒までです。デフォルト値は 30 秒です。HEARTBEAT_FREQUENCY オプションの値は、NETWORK_FAILURE_DETECTION_TIME オプションに使用される値より 10 秒多い値でなければなりません。

GRACE_PERIOD オプションは、HEARTBEAT_FREQUENCY オプションに対して指定された時間が経過した後で CAA が待機する追加時間です。GRACE_PERIOD オプションのデフォルト値は 10 秒です。

イベント管理機能の拡張

CAA は詳細なストレージおよびネットワーク・イベントを生成します。PowerHA SystemMirror はこれを使用して、高可用性の管理のために、より高度な意思決定機能を提供します。

ノード全体にわたるストレージの管理

PowerHA SystemMirror では、クラスター内のノード全体に対してより高度なストレージ管理を提供するために、AIX のストレージ・フェンシング機能を使用します。フェンシング機能は、AIX 固有のマルチパス入出力 (MPIO) が構成されているディスクに対してのみサポートされます。

PowerHA SystemMirror は、拡張コンカレント・ボリューム管理方式を使用して共用ディスクを管理します。

注: PowerHA SystemMirror は、所有する共用ボリューム・グループへのアクセス権限がないノードが共用ディスクにアクセスできないようにするために、CAA ストレージ・フレームワーク・フェンシング機能の使用を試行します。このフェンシング機能により、複数のノードから共用ディスクに不注意でアクセスしたことが原因でデータ破損が発生しないようにします。ただし、CAA ストレージ・フレームワーク・フェンシング機能は、AIX 固有の MPIO に対してのみサポートされます。

関連概念:

28 ページの『PowerHA SystemMirror マルチサイト・ソリューション』

PowerHA SystemMirror は、異なるタイプのサイト定義、および高可用性災害時リカバリー (HADR) のサイト固有のポリシーをサポートしています。PowerHA SystemMirror Standard Edition for AIX および PowerHA SystemMirror Enterprise Edition for AIX の両方で、複数のサイトを定義することができます。

29 ページの『PowerHA SystemMirror リンク・クラスター』

各サイトが地理上の異なる場所にある状況には、リンク・クラスターは理想的です。通常、これらのサイトは離れすぎているため、共通のストレージ・エリア・ネットワーク (SAN) をうまく共有できません。それぞれのサイトには、固有でアクティブなリポジトリ・ディスク、および何らかのバックアップ・リポジトリを備えている必要があります。リンク・クラスターはサイト間の通信に必ずユニキャストを使用します。リンク・クラスターは、高可用性災害復旧 (HADR) の有用な一部分です。

30 ページの『PowerHA SystemMirror 拡張クラスター』

拡張クラスターは、それぞれのサイトが地理上の同じ場所にある場合に最適です。通常、すべてのノードは共通のストレージ・エリア・ネットワーク (SAN) 上にあります。少なくとも、アクティブ・リポジトリおよび任意のバックアップ・リポジトリは、どちらのサイトでも稼働可能な、リソース・グループ内の任意のディスクが共有されるように、サイト間で共有されます。拡張クラスターは、サイト間 LVM ミラーリング、および HyperSwap[®] をサポートできます。

14 ページの『クラスター・リポジトリ・ディスク』

PowerHA SystemMirror は、クラスターの構成を管理するために中央リポジトリとして共用ディスクを使用します。このディスクには、標準クラスター・デプロイメント方式または拡張クラスター・デプロイメント方式を使用してクラスター内のすべてのノードがアクセスできなければなりません。あるいは、リンク・クラスター・デプロイメント方式を使用してサイト内のすべてのノードがアクセスできなければなりません。

『中央構成管理』

クラスター構成の重要な要素はクラスター・リポジトリ・ディスクです。クラスター・リポジトリ・ディスクは、クラスター構成データの中央リポジトリとして使用されます。

関連情報:

クラスター・ストレージ通信のセットアップ

中央構成管理

クラスター構成の重要な要素はクラスター・リポジトリ・ディスクです。クラスター・リポジトリ・ディスクは、クラスター構成データの中央リポジトリとして使用されます。

クラスター・リポジトリ・ディスクには、クラスター内のすべてのノードからアクセスできる必要があります。PowerHA SystemMirror クラスター構成に対して 512 MB 以上 460 GB 以下のスペースが必要です。これは冗長性のある高可用性 SAN 構成によってバックアップされなければなりません。中央リポジトリ・ディスクでは、単一障害点が生じるのを避けるために、RAID (Redundant Array of Independent Disks) を適切に構成する必要があります。

クラスター・リポジトリ・ディスクは、クラスター用の特殊なデバイスです。これは固有の特性を持っており、Cluster Aware AIX (CAA) および PowerHA SystemMirror によって独自に管理されます。LVM またはファイルシステム・コマンドおよびインターフェースはクラスター・リポジトリ・ディスクの破損の原因となる可能性があるため、直接使用しないでください。クラスター・リポジトリ・ディスクは CAA によって専用に管理されます。クラスター・リポジトリ・ディスクは、CAA 環境または PowerHA SystemMirror 環境の外部で管理してはなりません。

クラスター・リポジトリは、クラスター管理に対して次の目的のために使用されます。

- CAA クラスターの構成をリポジトリ・ディスク上で一元的に保守および管理します。
- クラスター・ベースの通信では、リポジトリ・ディスクの一部がノード間のハートビートとメッセージ通信のために予約されます。

注: この形式の通信は、他の形式の通信がすべて失敗した場合に使用されます。

関連概念:

4 ページの『PowerHA SystemMirror での Cluster Aware AIX の使用』

PowerHA SystemMirror は、AIX オペレーティング・システムでサポートされる主要なクラスター機能の最上位に構築されます。PowerHA SystemMirror は、Cluster Aware AIX (CAA) 機能に対応した AIX のすべてのエディションでサポートされます。

クラスターの SAN 通信のセットアップ

PowerHA SystemMirror および Cluster Aware AIX では、ホスト間の通信にストレージ・エリア・ネットワーク (SAN) 通信と呼ばれる追加の方法をサポートしています。

SAN 通信は、標準的なデータ・センターに存在する SAN 物理リンクを活用して、クラスター・ノード間で高速な通信を実現します。この通信メカニズムは、クラスター・ノード間のネットワーク通信パスで障害が発生した場合のフォールバック機能を備えています。

SAN 通信は一連のディスク・アダプターに対して機能します。SCSI パケットがそれぞれのホスト間にある各種のスイッチを介してホストからホストへ流れるように SAN ゾーニングをセットアップする必要があります。

SAN 通信のセットアップでは、サポートされるディスク・アダプターで一連のコマンドを呼び出して SAN 通信を可能にします。

VIOS 2.2.0.11 以降では、各 VIOS クライアント上の仮想イーサネット・アダプターを介して仮想ローカル・エリア・ネットワークを確立することによって、論理区画間で SAN 通信を使用することができます。NPIV および vSCSI の両方の環境で、VIOS を介した SAN 通信をセットアップできます。

関連情報:

クラスター・ストレージ通信のセットアップ

仮想イーサネット・アダプターの定義

可用性のコストと利点の関係

PowerHA SystemMirror の可用性には多くのコストと利点があります。

次の図は、可用性テクノロジーのコストと利点を示したものです。

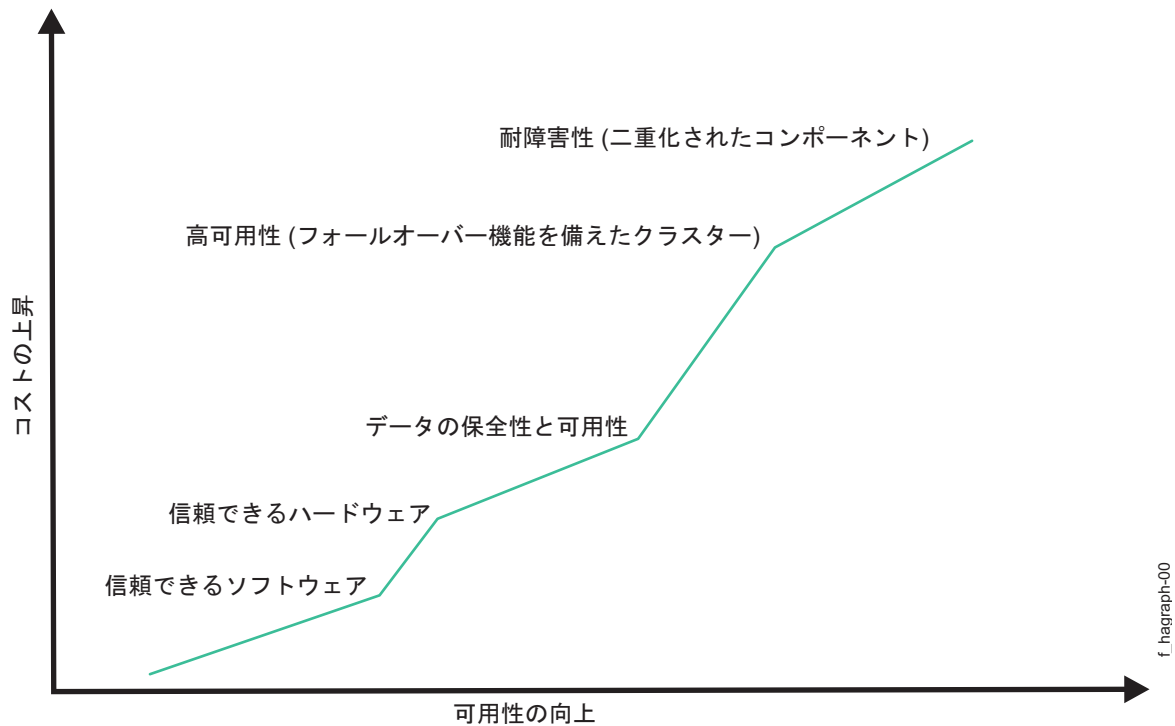


図 1. 可用性テクノロジーのコストと利点

上の図に示すように、可用性は、全か無かの命題ではありません。可用性を 1 つの連続体と考えてみてください。信頼できるハードウェアとソフトウェアが基本レベルの可用性を提供します。RAID デバイスなどの拡張機能がさらに高いレベルの可用性を提供します。ハイ・アベイラビリティ・ソフトウェアは、データおよびアプリケーションへのほぼ途切れることのないアクセスを提供します。耐障害システムでは、システム全体で一定した可用性が保証されますが、コストは高くなります。

AIX ソフトウェアによる可用性の強化

PowerHA SystemMirror は、高性能の UNIX オペレーティング・システムである AIX の機能を利用します。AIX では、セキュリティーおよびシステムの可用性をより向上させる新規機能が追加されました。

その中には、ミラーリングされるデータの可用性の向上、およびワークロード・マネージャーの機能拡張が含まれます。このワークロード・マネージャーは、リソースの可用性をクリティカル・アプリケーションに動的に提供するので、混合しているワークロードの問題を解決するのに役立ちます。PowerHA SystemMirror は、システムにダウン時間を生じさせることなく、水平および垂直の両方のスケーラビリティを実現します。AIX オペレーティング・システムは、(データのバックアップやシステム管理のための) 予定されたダウン時間と (ハードウェアやソフトウェアの障害による) 予定外のダウン時間による影響を低減することによってシステム可用性を向上させるように設計された多数のフィーチャーを備えており、Capacity on Demand (CoD) の使用によるハードウェア・リソース管理における柔軟性を提供します。

AIX オペレーティング・システムには以下のフィーチャーがあります。

- ジャーナル・ファイルシステムと拡張ジャーナル・ファイルシステム
- ディスク・ミラーリング
- プロセス制御
- DLPAR および CoD
- ワークロード・パーティション

ジャーナル・ファイルシステムと拡張ジャーナル・ファイルシステム

AIX 固有のファイルシステムであるジャーナル・ファイルシステム (JFS) は、データベース・ジャーナリング技法を使用して、その構造の整合性を維持します。システムまたはソフトウェアに障害が発生しても、ファイルシステムは管理不能な状態のままにはなりません。重要な障害の後でファイルシステムを再構築するときに、AIX は JFS ログを使用して、ファイルシステムを最後に整合していた状態に復元します。したがって、ジャーナリングを使用すれば、標準の UNIX ファイルシステム整合性検査 (fsck) ユーティリティよりも短時間で復旧が行えます。さらに、AIX では拡張ジャーナル・ファイルシステム (JFS2) を使用できます。

ディスク・ミラーリング

ディスク・ミラーリング・ソフトウェアは、データの整合性とオンライン・バックアップ機能を提供します。このソフトウェアは、ディスクの 1 つが故障してもデータにアクセスできるように最大 3 つのデータ・コピーを別々のディスクに保持することによってディスク障害によるデータ喪失を防止します。ディスク・ミラーリングは、アプリケーションに透過的なものです。ミラーリングされるディスクと従来型ディスクはアプリケーションからは同じように見えるので、アプリケーションを変更する必要はありません。

プロセス制御

AIX システム・リソース・コントローラー (SRC) は、主なプロセスのモニターおよび制御を行います。SRC は、プロセスの異常終了を検知し、その終了をログに記録して、通知プログラムにメッセージを渡し、そのプロセスをバックアップ・プロセッサ上で始動し直すことができます。

動的 LPAR 管理

PowerHA SystemMirror は LPAR 間でアプリケーション・リソースを移動できます。また、リソース最適化高可用性 (ROHA) 機能を使用して、必要な動的リソース調整を実行することができます。ROHA は、IBM Power Systems™ サーバーで使用できるフィーチャーを使用して、以下のタイプのハードウェア・リソースを動的に管理します。

- Capacity on Demand (CoD) 機能 (On/Off CoD および Enterprise Pool CoD を含む) は、フレーム (CEC) レベルでメモリーおよび CPU リソースを管理します。
- DLPAR 機能は、論理区画レベルでメモリーおよび CPU リソースを管理します。

ROHA 機能により、PowerHA SystemMirror を使用して各アプリケーションのリソース量を最適化することができます。例えば、テークオーバー中にハードウェア・リソース (CPU およびメモリー) はアクティブ・ノードから動的に解放され、動的に獲得されてスタンバイ・ノードに割り振られます。

ワークロード・パーティション

ワークロード・パーティションを使用すれば、同じオペレーティング・システム環境内で複数のアプリケーション・インスタンスを実行できます。その一方で、その環境を分離し、ひいてはインスタンスを分離して保護することもできます。

Cluster Aware AIX (CAA)

PowerHA SystemMirror は AIX のクラスター化機能を使用して、高度なハイ・アベイラビリティ・ソリューションを実現します。CAA は、ハートビートと正常性の管理を含む AIX カーネルを提供します。PowerHA SystemMirror は詳細なストレージおよびネットワーク・イベントについてモニターし、AIX オペレーティング・システムのクリティカルな状態を処理します。PowerHA SystemMirror はハードウェア・コンポーネントをディスカバーできるため、クラスターの管理とデプロイが容易になります。

PowerHA SystemMirror クラスターの物理コンポーネント

PowerHA SystemMirror は無中断の処理に不可欠な一連のリソースを特定することにより、高可用性環境を提供します。また、これらのリソースを使用可能にするために各ノードが連携する際に使用するプロトコルを定義します。

PowerHA SystemMirror は、連携するプロセッサ間の関係 (つまり、あるプロセッサがサービスを提供できなくなった場合にそのサービスを提供する別のプロセッサ) を定義することによって、クラスタリング・モデルを拡張したものです。次の図で示すとおり、PowerHA SystemMirror クラスタは次の物理コンポーネントで構成されています。

- ノード
- 共用外部ディスク装置
- ネットワーク
- ネットワーク・インターフェース
- クラスタ・リポジトリ・ディスク
- クライアント

PowerHA SystemMirror ソフトウェアでは、これらの物理コンポーネントを組み合わせることで広範囲なクラスタ構成を構築することができ、お客様の処理要件に適合するクラスタを構築する柔軟性を提供しています。この図は、PowerHA SystemMirror クラスタの一例を示したものです。プロセッサの数、選択したネットワークやディスク・テクノロジーなどによって、PowerHA SystemMirror クラスタは、この図に示すものとはかなり違う場合もあります。

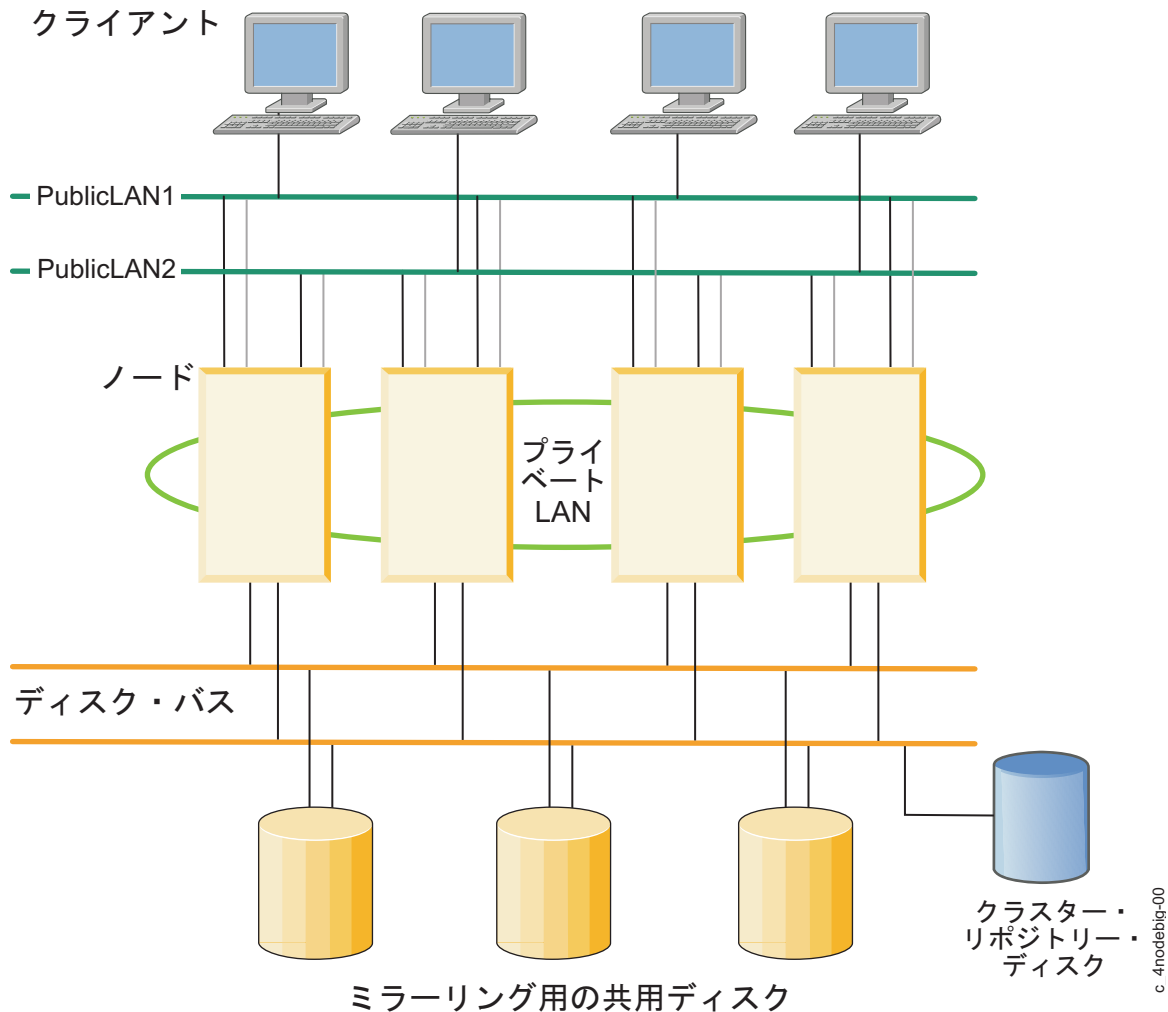


図 2. PowerHA SystemMirror クラスターの例

PowerHA SystemMirror クラスター構成は、PowerHA SystemMirror ソフトウェアでサポートされるクラスター構成の例を示します。

関連概念:

90 ページの『PowerHA SystemMirror クラスター構成』

この章では、PowerHA SystemMirror ソフトウェアがサポートする各種のサンプル・クラスター構成を提供します。

PowerHA SystemMirror ノード

ノードは PowerHA SystemMirror クラスターのコアを形成します。ノードは、AIX、PowerHA SystemMirror ソフトウェア、およびアプリケーション・ソフトウェアを実行するプロセッサです。

これらのサーバーをクラスター化してクリティカル・アプリケーションをバックアップするのは、コスト効率に優れたハイ・アベイラビリティの選択肢です。ハードウェアまたはソフトウェアの障害が発生しても中断時間を最小限に抑えてクリティカル・アプリケーションの実行を再開できるようになるため、企業は自社のコンピューティング・パワーを有効に利用できるようになります。

PowerHA SystemMirror クラスターでは、各ノードは固有の名前により識別されます。ノードは、リソースの集合 (ディスク、ボリューム・グループ、ファイルシステム、ネットワーク、ネットワーク・アドレ

ス、およびアプリケーション)を所有できます。一般的には、ノードは共用外部ディスク上のデータにアクセスするサーバー・アプリケーションまたはバック・エンド・アプリケーションを実行します。

関連概念:

65 ページの『アプリケーションの可用性の確保』

このセクションでは、PowerHA SystemMirror ソフトウェアがクラスター・コンポーネントの可用性を確保することで、どのようにアプリケーションの可用性を確保するかについて説明します。PowerHA SystemMirror は、すべてのキー・システム・コンポーネントについて単一障害点を除去し、ほとんどのルーチン・クラスター保守タスクについてスケジュールされたダウン時間の必要を除去します。

41 ページの『PowerHA SystemMirror リソースおよびリソース・グループ』

ここでは、文書全体で使用され、PowerHA SystemMirror ユーザー・インターフェースでも使用されるリソース関連の概念および定義について説明します。

共用外部ディスク装置

各ノードは、1 つ以上の共用外部ディスク装置にアクセスできなければなりません。共用外部ディスク装置は、複数のノードに物理的に接続されたディスクです。

共用ディスクは、主幹業務データを格納し、通常はデータ冗長度を保持するためにミラーリングまたは RAID 構成になります。PowerHA SystemMirror クラスター内のノードには、オペレーティング・システムおよびアプリケーションのバイナリーを保管する内部ディスクも必要です。ただし、これらのディスクは共用されません。使用するディスクのタイプにより、PowerHA SystemMirror ソフトウェアは、共用外部ディスク装置へのアクセス方式として、非コンカレント・アクセスとコンカレント・アクセスの 2 つをサポートしています。

- 非コンカレント・アクセス環境では、任意の時点で 1 つの接続だけがアクティブであり、アクティブな接続のあるノードがディスクを所有します。1 つのノードに障害が生じ、現在ディスクを所有しているノードがクラスターから離れると、ディスク・テークオーバーが起こり、クラスターに残ったノードが共用ディスクの所有権を引き継ぎます。この標準クラスター構成は、ほとんどのアプリケーションによって使用されます。
- コンカレント・アクセス環境では、共用ディスクは同時に複数のノードにアクティブに接続されます。したがって、1 つのノードに障害が発生してもディスク・テークオーバーは必要ありません。このタイプのアクセスは、複数のノードから共用データへの同時アクセスを管理および調整することができるアプリケーションによってのみ使用されます。

そのような環境では、クラスター内で定義されたすべてのノードがコンカレント・アクセスの一部である、またはクラスターのサブセットのみが共用ディスクへのアクセス権を持つことができることに注意してください。2 つめのケースの場合、共用ディスク・アクセス権を持つノードのみにリソース・グループを構成します。この 2 つの方式の違いについて詳しくは、セクション『共用外部ディスク・アクセス』を参照してください。

関連概念:

62 ページの『共用外部ディスク・アクセス』

PowerHA SystemMirror ソフトウェアは、2 つの方式の共用外部ディスク・アクセス (非コンカレントとコンカレント) をサポートします。

ネットワーク

PowerHA SystemMirror ソフトウェアは、AIX オペレーティング・システムの独立した階層化コンポーネントであり、TCP/IP ベースのあらゆるネットワークで使用できるよう設計されています。

PowerHA SystemMirror クラスター内のノードは、ネットワークを使用して次のことを行います。

- クライアントがクラスター・ノードにアクセスできるようにする
- クラスター・ノード間のハートビート・メッセージの交換を可能にする
- (コンカレント・アクセス環境では) データへのアクセスを逐次化する

PowerHA SystemMirror ソフトウェアは、イーサネットおよび Storage Area Networks を使用してテスト済みです。

ネットワークのタイプ

PowerHA SystemMirror ソフトウェアでは、これらのネットワークが TCP/IP サブシステムに基づく通信インターフェースを使用する (TCP/IP ベース) か Storage Area Network を使用するかによって、2 種類の通信ネットワークが定義されます。

TCP/IP ベースのネットワーク

TCP/IP プロトコルを使用して、2 つ以上のサーバー・ノードに接続し、オプションでこれらのクラスター・ノードへのクライアント・アクセスを可能にします。デフォルトで、PowerHA SystemMirror はハートビートにユニキャスト通信を使用します。ネットワークがマルチキャストをサポートするように構成されている場合は、オプションとしてマルチキャスト通信の使用を選択できます。

Storage Area Network (SAN)

PowerHA SystemMirror クラスター・ノード間でメッセージおよびハートビート・トラフィックを制御するための通信を提供します。これは、TCP/IP ネットワークの代替通信パスです。

クラスター・リポジトリ・ディスク

PowerHA SystemMirror は、クラスターの構成を管理するために中央リポジトリとして共用ディスクを使用します。このディスクには、標準クラスター・デプロイメント方式または拡張クラスター・デプロイメント方式を使用してクラスター内のすべてのノードがアクセスできなければなりません。あるいは、リンク・クラスター・デプロイメント方式を使用してサイト内のすべてのノードがアクセスできなければなりません。

System Management Interface Tool (SMIT) インターフェースを使用して、クラスター・リポジトリ・ディスクとバックアップ・リポジトリ・ディスクを管理および構成します。

標準クラスターおよび拡張クラスターでは、クラスターごとに少なくとも 1 つのアクティブ・リポジトリ・ディスクが必要です。標準クラスターおよび拡張クラスターのクラスターごとに、最大 6 つのバックアップ・リポジトリ・ディスクを指定できます。リンク・クラスターでは、サイトごとに 1 つのアクティブ・リポジトリ・ディスクが必要です。リンク・クラスターのサイトごとに、最大 6 つのバックアップ・リポジトリ・ディスクを指定できます。

Cluster Aware AIX を使用する環境で、クラスター・リポジトリ・ディスクは次の目的のために使用されます。

- クラスター全体の構成管理。
- クラスターのメッセージングおよびハートビート。リポジトリ・ディスクは、ノード間の通信に対する別の冗長パスとして使用されます。

クラスター・リポジトリ・ディスクには、512 MB 以上かつ 460 GB 以下のディスク・スペースが割り振られている必要があります。

リポジトリー・ディスクとして選択したディスクに予約ポリシーが指定されていないことを確認してください。ディスクの予約ポリシーを確認するには、`lsattr -El hdisk - a reserve_policy` コマンドを実行します。ディスクの予約ポリシーを `no_reserve` に変更する場合は、`chdev -a reserve_policy=no_reserve -l hdisk` コマンドを実行します。

ディスクをリポジトリー・ディスクとして割り当てた後で、そのディスクを他の目的で使用することはできません。リポジトリー・ディスクとして使用するディスクにユーザー・データが含まれていないことを確認してください。PowerHA SystemMirror によってそのディスクがリポジトリー・ディスクとして識別されるときに、ディスク上のすべての情報が消去されます。

ディスクをリポジトリー・ディスクとして使用する計画を立てる際は、1 次リポジトリー・ディスクに障害が起きた場合に使用できるバックアップ・ディスクまたは交換用ディスクについて計画する必要があります。バックアップ・ディスクは、1 次ディスクと同じサイズおよびタイプにする必要がありますが、異なる物理ストレージ・ディスクであってもかまいません。バックアップ・ディスク情報を使用して、管理手順および文書を更新します。リポジトリー・ディスクの喪失または障害は、クラスターによって容認されます。基幹業務は、リポジトリー・ディスクなしで継続されます。しかし、リポジトリー・ディスクに関する問題は速やかに修正しなければなりません。PowerHA SystemMirror 7.2.0 以前を使用している場合は、障害が発生したリポジトリー・ディスクを、バックアップ・リストにある別のリポジトリー・ディスクに手動で交換する必要があります。稼働中のリポジトリー・ディスクを新しいディスクと交換して、サイズを増やしたり異なるストレージ・サブシステムに変更することもできます。

アクティブ・リポジトリー・ディスクの障害発生時に Cluster Aware AIX (CAA) の Automatic Repository Disk Replacement (ARR) 機能を使用するように PowerHA SystemMirror 7.2.0 以降を構成できます。ARR 機能は自動的に、障害が発生したリポジトリー・ディスクをバックアップ・リポジトリー・ディスクのうちの 1 つに交換します。リストの最初にあるバックアップ・リポジトリー・ディスクによって、障害の発生したリポジトリー・ディスクが置き換えられます。

すべてのリポジトリー・ディスクに AIX 固有のマルチパス入出力 (MPIO) を構成する必要があります。AIX MPIO により、PowerHA SystemMirror は、リポジトリー・ディスクを使用してより適切にディスク障害をモニターして対応できるようになります。

ミラー・プールおよびリポジトリー・ディスク

PowerHA SystemMirror および論理ボリューム・マネージャー (LVM) を使用して異なるロケーションにわたるミラー・プールを管理しているときに、リポジトリー・ディスクで障害が発生した場合、障害が発生したロケーションのディスクは新しいリポジトリー・ディスクとして使用できません。

関連概念:

29 ページの『PowerHA SystemMirror リンク・クラスター』

各サイトが地理上の異なる場所にある状況には、リンク・クラスターは理想的です。通常、これらのサイトは離れすぎているため、共通のストレージ・エリア・ネットワーク (SAN) をうまく共有できません。それぞれのサイトには、固有でアクティブなリポジトリー・ディスク、および何らかのバックアップ・リポジトリーを備えている必要があります。リンク・クラスターはサイト間の通信に必ずユニキャストを使用します。リンク・クラスターは、高可用性災害復旧 (HADR) の有用な一部分です。

30 ページの『PowerHA SystemMirror 拡張クラスター』

拡張クラスターは、それぞれのサイトが地理上の同じ場所にある場合に最適です。通常、すべてのノードは共通のストレージ・エリア・ネットワーク (SAN) 上にあります。少なくとも、アクティブ・リポジトリーおよび任意のバックアップ・リポジトリーは、どちらのサイトでも稼働可能な、リソース・グループ内の任意のディスクが共有されるように、サイト間で共有されます。拡張クラスターは、サイト間 LVM ミラーリング、および HyperSwap をサポートできます。

4 ページの『PowerHA SystemMirror での Cluster Aware AIX の使用』

PowerHA SystemMirror は、AIX オペレーティング・システムでサポートされる主要なクラスター機能の最上位に構築されます。PowerHA SystemMirror は、Cluster Aware AIX (CAA) 機能に対応した AIX のすべてのエディションでサポートされます。

関連情報:

リポジトリ・ディスクのプランニング

SMIT によるリポジトリ・ディスクの交換

リポジトリ・ディスクおよびマルチパス:

AIX オペレーティング・システムは、クラスターの構成中に特別な方法でリポジトリ・ディスクを管理します。

AIX クラスター・サービスは、AIX デバイス構成処理中にリポジトリ・ディスクを認識し、そのディスク上で特別なアクションを実行することができます。例えば、リポジトリ・ディスクはクラスター・サービスによって使用され、ホスト間でディスク・ベースの通信を可能にします。ディスク通信には、ディスクへの書き込みを行う 1 つのホストと、ポーリング機構を通してメッセージを取得する 1 つまたは複数の別のホストが含まれます。

また、クラスター・サービスはリポジトリ・ディスクの正常性に関連した詳細なイベントを生成します。リポジトリ・ディスクに障害が発生すると、CAA は、障害が発生したリポジトリ・ディスクを自動的にバックアップ・リポジトリ・ディスクに交換する Automatic Repository Replacement (ARR) 機能を使用します。ARR 機能は、PowerHA SystemMirror を使用してバックアップ・リポジトリ・ディスクを指定した場合のみ使用可能です。

以下のタイプのディスクをリポジトリ・ディスクとして構成することができます。

AIX マルチパス・ディスク (AIX ローカル MPIO)

これらのディスクは、自動的に作成してリポジトリ・ディスクとして使用することができます。これは、リポジトリ・ディスクとして使用する場合の推奨ディスク・タイプです。

仮想 SCSI (vSCSI) ディスク

これらのディスクは、仮想入出力サーバー (VIOS) を使用して、クライアント論理区画 (LPAR) に対して vSCSI ディスクとしてマップすることができます。vSCSI ディスクをリポジトリ・ディスクとして使用する場合、このディスクは、同様の vSCSI 方式を使用して別の PowerHA SystemMirror ノードまたは別の Cluster Aware AIX (CAA) ノードに対してマップする必要があります。

サード・パーティーのマルチパス・ディスク

これらのディスクは AIX マルチパスの概念のガイドラインに従っていますが、独自のマルチパス・デバイス・ドライバーおよびソフトウェアを提供します。関連するオブジェクト・データ・マネージャー (ODM) 情報が使用可能な場合、これらのディスクをリポジトリ・ディスクとして構成することができます。EMC PowerPath および Hitachi Dynamic Link Manager (HDLM) によって管理されるディスクは、この方式を使用してリポジトリ・ディスクとして構成できます。

MPIO ソフトウェアまたはサード・パーティーのマルチパス・ソフトウェアをアップグレードするには、**clmgr offline cluster STOP_CAA=yes** コマンドを入力して、CAA クラスター・サービスを停止する必要があります。

リポジトリー・ディスクの障害:

リポジトリー・ディスクの障害に対して正しい計画を立て、リポジトリー・ディスクの障害に関連する問題を修正するために必要な事項を計画する必要があります。

リポジトリー・ディスクの障害は PowerHA SystemMirror クラスタによって容認されます。クラスタ内のノードがリポジトリー・ディスク、またはリポジトリー・ディスクへのアクセスでエラーを検出すると、そのクラスタは制限モードに入ります。この操作モードでは、ほとんどのトポロジー関連操作が使用できません。例えば、ノードの追加やノードのクラスタとの結合は行えません。ただし、重要なクラスタ機能は実行できます。例えば、アクティブ・ノードからスタンバイ・ノードへのリソース・グループの移動は可能です。

リポジトリー・ディスクで障害が発生した場合、管理者にディスク障害が通知されます。障害が解決されるまで、PowerHA SystemMirror は管理者にリポジトリー・ディスクの障害を通知し続けます。リポジトリー・ディスクで障害が発生したときに通知を受け取ったりリポジトリー・ディスクで障害が発生したときのイベント処理をカスタマイズしたりするには、『イベント前処理とイベント後処理の構成』トピックを参照してください。

PowerHA SystemMirror および Cluster Aware AIX (CAA) は稼働中のリポジトリー・ディスク交換をサポートしており、その機能を使用すれば、障害が発生したりリポジトリー・ディスクでも正常に機能しているリポジトリー・ディスクでも交換することができます。CAA はクラスタ情報を新しいディスクに追加し直し、そのディスクをリポジトリー・ディスクとして使用し始めます。

PowerHA SystemMirror 7.2.0 以降は、Automatic Repository Disk Replacement (ARR) 機能をサポートしています。ARR は CAA を使用し、障害が発生したりリポジトリー・ディスクを自動的にバックアップ・リポジトリー・ディスクに交換します。ARR 機能は、PowerHA SystemMirror を使用してバックアップ・リポジトリー・ディスクを構成した場合のみ使用可能です。

ARR 機能を使用するには、ご使用の環境が以下の要件を満たしている必要があります。

- クラスタまたはサイトに、識別されたバックアップ・リポジトリー・ディスクが存在する。
- PowerHA SystemMirror バージョン 7.2.0 以降がインストールされている。
- 次のいずれかのバージョンの AIX オペレーティング・システムがインストールされている。
 - AIX バージョン 7.1.4 以降
 - AIX バージョン 7.2.0 以降

CAA は、入出力エラーの検査、およびディスクがアクティブ状態かどうかの検査を行うことにより、リポジトリー・ディスクの障害をモニターします。これらの検査による確認は定期的に行われており、リポジトリー・ディスクの読み取り時または書き込み時に毎回実行されるわけではありません。テストが目的であっても、リポジトリー・ディスクに直接書き込まないでください。非同期的にリポジトリー・ディスクへ直接書き込むと、オペレーティング・システムおよび CAA の動作が突然中断される原因となるため、予測不能な結果が生じる場合があります。

関連情報:

dd コマンド

リポジトリー・ディスクのプランニング

SMIT によるリポジトリー・ディスクの交換

クライアント

クライアントは、ローカル・エリア・ネットワークを介してクラスター内のノードにアクセスできるプロセッサです。

クライアントはそれぞれ「フロントエンド」またはクライアント・アプリケーションを実行し、クラスター・ノード上で実行されるサーバー・アプリケーションを照会します。PowerHA SystemMirror ソフトウェアは、クラスター・ノード上のクリティカル・データおよびアプリケーションに可用性の高い環境を提供します。PowerHA SystemMirror ソフトウェアは、クライアント自身の可用性を高めるものではありません。AIX クライアントは、クラスター情報 (Clinfo) サービスを使用して、クラスター・イベントの通知を受け取ることができます。Clinfo は、クラスター状況情報を表示する API を備えています。PowerHA SystemMirror は、クラスター状況ユーティリティー `/usr/es/sbin/cluster/clstat` を提供します。それは、Clinfo に基づいており、主要クラスター・コンポーネントの状況 (クラスター自体、クラスター内のノード、ノードに接続されているネットワーク・インターフェース、および各ノード上のリソース・グループ) を報告します。`clstat` ユティリティーのクラスター状況インターフェースは、Web ベース、Motif ベース、ASCII ベースのバージョンを含んでいます。

関連概念:

61 ページの『クラスター情報プログラム』

クラスター情報プログラム (Clinfo) の `clinfo` デーモンは、SNMP ベースのモニターです。

PowerHA SystemMirror の目的: 予定されたダウン時間の解消

ハイ・アベイラビリティ・クラスタリング・ソフトウェアの主な目的は、保守や再構成作業時に必要とされるリソースのサービス停止時間を最小限に抑え、できれば停止時間をゼロにすることにあります。

PowerHA SystemMirror ソフトウェアは、実行中のクラスターの動的再構成を可能にすることによって可用性を最適化します。ノードの追加や削除、またはリソース・グループに関与するノードの優先順位の変更などのような通常のクラスター保守タスクは、クラスター・サービスを停止、再始動しなくても、アクティブ・クラスターに適用できます。さらに Cluster Single Point of Control (C-SPOC) 機能を使用すれば、PowerHA SystemMirror クラスターのオンライン状態を維持したまま構成を変更できます。C-SPOC により、クラスター内にある共有ボリューム・グループ、ユーザー、およびグループに対する変更を 1 つのノードから行えるようになるため、クラスター管理が容易になります。行った変更は、他のクラスター・ノードに透過的に伝搬されます。

PowerHA SystemMirror クラスター・ノード、ネットワーク、およびハートビートの概念

このセクションでは、文書全体にわたって使用され、PowerHA SystemMirror ユーザー・インターフェースの中でも使用される、主要なクラスター・トポロジーに関する概念および定義について説明します。

ノード

ノードは、AIX と PowerHA SystemMirror ソフトウェアの両方を実行するプロセッサです。

ノードは、リソースの集合 (ディスク、ボリューム・グループ、ファイルシステム、ネットワーク、ネットワーク IP アドレス、およびアプリケーションなど) を共有できます。PowerHA SystemMirror ソフトウェアは、1 つのクラスターで最大 16 個のノードをサポートします。PowerHA SystemMirror クラスターでは、各ノードは固有の名前により識別されます。一般に、PowerHA SystemMirror ではノード名と

ホスト名は同じ名前になります。ノードは、PowerHA SystemMirror クラスターのコア物理コンポーネントとして機能します。ノードおよびハードウェアについては、セクション『ノード』を参照してください。次の 2 種類のノードが定義されます。

- サーバー・ノードは PowerHA SystemMirror クラスターのコアを構成します。サーバー・ノードは、共用外部ディスク上のデータにアクセスするサービスまたはバックエンド・アプリケーションを実行します。
- クライアント・ノードは、サーバー・ノードによって提供されるサービスからデータを検索するフロントエンド・アプリケーションを実行します。クライアント・ノードで PowerHA SystemMirror ソフトウェアを実行することにより、ノードの正常性をモニターし、障害に対処することができます。

サーバー・ノード

サーバー・ノード・クラスターは、通常は共用外部ディスク上のデータにアクセスするアプリケーションを実行します。サーバー・ノードは PowerHA SystemMirror デモンを実行し、リソースの高可用性を保持します。一般に、アプリケーションが実行され、ストレージはこれらのノード間で共用され、クライアントはサービス IP アドレスを通してサーバー・ノードに接続します。

クライアント・ノード

全ハイ・アベイラビリティ・ソリューションには、通常は各サーバーによって提供されるサービスを使用するクライアント・マシンが含まれています。クライアント・ノードは、ナীবおよびインテリジェントという 2 つのカテゴリーに分けることができます。

- ナীব・クライアントは、クラスターを 1 つのエンティティと見なします。サーバーに障害が発生した場合、そのクライアントを再始動する必要 (あるいは、少なくともサーバーに再接続する必要) があります。
- これに対して、インテリジェント・クライアントはクラスター対応型です。クラスター対応クライアントは、サーバーの障害時に適切に対処し、別のサーバーへ接続して、おそらくユーザーに障害の発生を気付かせることはありません。このようなインテリジェント・クライアントはクラスターの状態についての知識が必要です。

PowerHA SystemMirror は、(サブシステムやノード障害の変化などの) クラスターの状態変化のダイナミック・クラスター構成レポートと通知の両方を提供することによってクラスター・パラダイムを拡張します。

サイト

site に属する 1 つ以上のサーバー・ノードのグループを定義できます。

サイトは、ノードやネットワークのように、PowerHA SystemMirror ソフトウェアに認識されるコンポーネントの 1 つになります。PowerHA SystemMirror は、2 つのサイトに分割されたクラスターをサポートします。

特定の LVM ミラー・プールを物理サイトに関連付けることにより、分割サイトの論理ボリューム・マネージャー (LVM) ミラーリングを構成することができます。次に、各サイトに存在する物理ボリュームを PowerHA SystemMirror に指定した場合、LVM ミラー・プールとしてサイトからボリュームを選択すると、C-SPOC はそのサイトの情報を表示します。クラスターの検査時に、PowerHA SystemMirror は、ミラーの定義とサイトの定義の整合性を確認するために追加のチェックを実行します。

さらに、PowerHA SystemMirror Enterprise Edition の拡張距離機能は、災害復旧の 2 つの別々のソフトウェア・ソリューションを提供します。これらのソリューションは、PowerHA SystemMirror クラスタが 2 つのサイトで拡張された距離間で操作できるようにします。

PowerHA SystemMirror Enterprise Edition for Metro Mirror は、ピアツーピア・リモート・コピー (PPRC) を使用して、災害時回復目的でリモート・サイトにデータをコピーする DS8000[®] ボリューム、DS6000[™] ボリューム、および IBM TotalStorage Enterprise Storage Server[®] (ESS) ボリュームのデータ可用性を増大させます。PowerHA SystemMirror Enterprise Edition for Metro Mirror は、PPRC のフォールオーバーおよびフォールバック機能と PowerHA SystemMirror クラスタ管理を利用して、災害復旧時のダウン時間と回復時間を短縮します。

PPRC が、サイト間のデータ・ミラーリングに使用されるとき、サイト間の物理距離は、ESS ハードウェアの機能に制限されます。

PowerHA SystemMirror Enterprise Edition for Geographic Logical Volume Manager (GLVM) は、GLVM を使用して、災害時回復の目的でリモート・サイトにデータをコピーする IBM ボリュームのデータ可用性を向上させます。PowerHA SystemMirror Enterprise Edition for GLVM は、以下のコンポーネントを利用して、災害復旧の際のダウン時間と復旧時間を短縮します。

AIX および PowerHA SystemMirror Enterprise Edition for GLVM のデータ・ミラーリングと同期化。標準および拡張の両方のコンカレント・ボリューム・グループを、GLVM ユーティリティを使用して地理的にミラーリングすることができます。

最大 4 つの XD_data データ・ミラーリング・ネットワークを持つ TCP/IP ベースの無制限距離のネットワーク・サポートを構成できます。

コンポーネント障害の発生時に、PowerHA SystemMirror はミラーリングされたデータのコピーをローカル・サイトとリモート・サイトのいずれからも利用できるようにします。コンカレントおよび非コンカレントの両方のリソース・グループを、PowerHA SystemMirror クラスタ内で GLVM を使用して構成できます。ただし、サイト間のポリシーをコンカレントにすることはできません。

関連情報:

計画ガイド

Metro Mirror: プランニングおよび管理ガイド

Geographic LVM: プランニングおよび管理ガイド

分割ポリシー

一群のノードがクラスタ内のその他のノードと通信できない場合、サイト間でクラスタ分割イベントが発生することがあります。例えば、リンク・クラスタ内では、2 つのサイト間のすべての通信リンクで障害が発生すると分割が発生します。クラスタ分割イベントにより、クラスタは 2 つ以上の区画に分割されます。

PowerHA SystemMirror を使用して、クラスタ分割イベントへの対応を指定する分割ポリシーを構成できます。

分割ポリシーを構成する場合は、以下のオプションを選択できます。

なし このオプションは、クラスタ分割イベントが検出されたときにアクションが実行されないように指示します。クラスタ分割イベントにより作成された各区画は、独立したクラスタになります。各区画は、他の区画に関係なくワークロードを開始できます。共有ボリューム・グループが使用中の場合は、データ破損につながるおそれがあります。このオプションはデフォルト設定です。

その理由は、代わりのポリシーを作成するには手動による構成が必要なためです。ご使用の環境が HyperSwap for PowerHA SystemMirror を使用するように構成されている場合は、このオプションを使用しないでください。

タイ・ブレイカー

このオプションを使用すれば、ディスクまたは NFS ファイルを指定できます。

ディスクをタイ・ブレイカーに対して指定した場合、各区画は、タイ・ブレイカー・ディスクにロックをかけることによりタイ・ブレイカー・ディスクを取得しようとし、ディスクをタイ・ブレイカーに対して指定すると、クラスターにあるすべてのノードからアクセスできる SCSI ディスクが使用されます。このディスクにロックをかけることができない区画は、アクション・プランで指定されたとおりにリブートされます。

NFS ファイルをタイ・ブレイカーに対して指定する場合、NFS マウントは、選択された NFS サーバーのクラスター内の各ノードに存在しなければなりません。NFS ファイルを最初に予約した区画が引き続き機能します。この NFS ファイルにロックをかけることができない区画は、アクション・プランで指定されたとおりにリブートされます。

手動 このオプションは、クラスターが分割されたときに問題を手動で修正することを示します。

区画内の各ノードは、クラスター・サービスの実行を継続するか、クラスター・サービスを復旧するか (これはノードを再始動します) を選択するメッセージを表示します。このオプションを指定した場合は、入力が必要となる試行の回数および試行の頻度を指定できます。また、入力を必要とする試行の回数に達した場合に入力データを何も指定していないとき、その後発生するデフォルトのアクションを指定することもできます。

クラスター分割イベントが発生したときに「手動」オプションを指定するリンク・クラスターに関しては、次のメッセージが表示されます。

```
Broadcast message from root@e08m138.ausprv.stglabs.ibm.com (tty) at 04:09:48 ...
A cluster split has been detected.
You must decide if this side of the partitioned cluster is to continue.
To have it continue, enter
    /usr/es/sbin/cluster/utilities/cl_sm_continue
To have the recovery action - Reboot - taken on all nodes on this partition, enter
    /usr/es/sbin/cluster/utilities/cl_sm_recover
LOCAL_PARTITION 1 e08m138 OTHER_PARTITION 2 e08m140
```

この例では、「手動」オプションを使用して、手動応答を分割イベントまたはマージャー・イベントが待機しているのかどうかを「問題判別ツール」 > 「分割またはマージへの手動応答」 > 「必要なすべての手動応答の表示」SMIT メニューから確認することができます。

拡張クラスターおよび標準クラスターに対して「手動」オプションを使用する場合は、ご使用の環境で以下のバージョンのソフトウェアが実行されていなければなりません。

- IBM AIX 7.2 with Technology Level 1 またはそれ以降
- PowerHA SystemMirror バージョン 7.2.1 またはそれ以降

注: 「手動」オプションが使用されるいずれのタイプのクラスターに関しても、指定された試行回数に達した後は、何も入力しないのであれば、最低のノード ID を持つ区画が、運用が継続される区画として選択されます。

関連概念:

22 ページの『マージ・ポリシー』

クラスターには、クラスター分割ポリシーに応じて、それぞれ独立して稼働する 2 つの区画がある場合があります。PowerHA SystemMirror を使用して、区画間の通信が復元した後で、再度、区画を一緒に運用できるようにするマージ・ポリシーを構成できます。

24 ページの『分割ポリシーおよびマージ・ポリシーのタイ・ブレイカー・オプション』
タイ・ブレイカー・オプションを使用して、分割ポリシーおよびマージ・ポリシーで使用される SCSI ディスクまたはネットワーク・ファイルシステム (NFS) ファイルを指定できます。

29 ページの『PowerHA SystemMirror リンク・クラスター』
各サイトが地理上の異なる場所にある状況には、リンク・クラスターは理想的です。通常、これらのサイトは離れすぎているため、共通のストレージ・エリア・ネットワーク (SAN) をうまく共有できません。それぞれのサイトには、固有でアクティブなリポジトリ・ディスク、および何らかのバックアップ・リポジトリを備えている必要があります。リンク・クラスターはサイト間の通信に必ずユニキャストを使用します。リンク・クラスターは、高可用性災害復旧 (HADR) の有用な一部分です。

関連情報:

分割とマージのポリシーの構成

マージ・ポリシー

クラスターには、クラスター分割ポリシーに応じて、それぞれ独立して稼働する 2 つの区画がある場合があります。PowerHA SystemMirror を使用して、区画間の通信が復元した後で、再度、区画を一緒に運用できるようにするマージ・ポリシーを構成できます。

マージ・ポリシーを構成する場合は、以下のオプションを選択できます。

多数決

一番大きい番号のノードがある区画がオンラインを継続します。同じ数のノードが各区画にある場合は、最低のノード ID を持つ区画が選択されます。オンライン状態を継続しない区画は、選択したアクション・プランで指定されているようにしてリブートされます。拡張クラスターで「多数決」オプションが使用される場合は、ご使用の環境で以下のいずれかの AIX オペレーティング・システム・バージョンが稼働していなければなりません。

- IBM AIX 7.1 (テクノロジー・レベル 4 適用) またはそれ以降
- AIX バージョン 7.2 またはそれ以降

タイ・ブレイカー

タイ・ブレイカーにはディスクまたは NFS ファイルを使用できます。ディスクをタイ・ブレイカーに使用する場合、各区画は、タイ・ブレイカー・ディスクにロックをかけることによりタイ・ブレイカーを取得しようとします。タイ・ブレイカーは、クラスター内のすべてのノードからアクセス可能な SCSI ディスクです。このディスクにロックをかけられない区画では、選択したアクション・プランで指定されたように、リブートされるか、クラスター・サービスが再始動されます。

NFS ファイルをタイ・ブレイカーに使用する場合、NFS マウントは、選択した NFS サーバーのクラスター内の各ノードに存在する必要があります。NFS ファイルを最初に予約した区画が引き続き機能します。この NFS ファイルにロックをかけられない区画では、選択したアクション・プランで指定されたように、リブートされる、クラスター・サービスが再始動されます。このオプションを使用する場合は、分割ポリシー構成でもこのタイ・ブレイカー・オプションを使用する必要があります。

手動 このポリシー・オプションでは、運用を継続するサイトをマージ・イベント中にユーザーが選択する必要があります。

区画内の各ノードは、クラスター・サービスの実行を継続するか、クラスター・サービスを復旧するか (これはノードを再始動します) を選択するメッセージを表示します。このオプションを指定した場合は、入力が必要となる試行の回数および試行の頻度を指定できます。また、入力を必要とする試行の回数に達した場合に入力データを何も指定していないとき、その後が発生するデフォルトのアクションを指定することもできます。

クラスター分割イベントが発生したときに「手動」オプションを指定するリンク・クラスターに関しては、次のメッセージが表示されます。

```
Broadcast message from root@e08m138.ausprv.stglabs.ibm.com (tty) at 04:09:48 ...
A cluster split has been detected.
You must decide if this side of the partitioned cluster is to continue.
To have it continue, enter
    /usr/es/sbin/cluster/utilities/cl_sm_continue
To have the recovery action - Reboot - taken on all nodes on this partition, enter
    /usr/es/sbin/cluster/utilities/cl_sm_recover
LOCAL_PARTITION 1 e08m138 OTHER_PARTITION 2 e08m140
```

この例では、「手動」オプションを使用して、手動応答を分割イベントまたはマージャー・イベントが待機しているのかどうかを「問題判別ツール」 > 「分割またはマージへの手動応答」 > 「必要なすべての手動応答の表示」SMIT メニューから確認することができます。

拡張クラスターおよび標準クラスターに対して「手動」オプションを使用する場合は、ご使用の環境で以下のバージョンのソフトウェアが実行されていなければなりません。

- IBM AIX 7.2 with Technology Level 1 またはそれ以降
- PowerHA SystemMirror バージョン 7.2.1 またはそれ以降

注: 「手動」オプションが使用されるいずれのタイプのクラスターに関しても、指定された試行回数に達した後は、何も入力しないのであれば、最低のノード ID を持つ区画が、運用が継続される区画として選択されます。

なし このオプションは、クラスター・マージ・イベントが発生したときにアクションが実行されないように指示します。マージが行われた後でデータが壊れないようにするために、運用が継続されない区画ノードを手動でリブートする必要があります。このオプションは、**clmgr** コマンドからのみ選択できます。マージ・ポリシーに対して「なし」を指定する場合は、分割ポリシーに対して「なし」を選択する必要があります。

拡張クラスターおよび標準クラスターに対して「なし」オプションを使用する場合は、ご使用の環境で以下のバージョンのソフトウェアが実行されていなければなりません。

- IBM AIX 7.2 with Technology Level 1 またはそれ以降
- PowerHA SystemMirror バージョン 7.2.1 またはそれ以降

関連概念:

20 ページの『分割ポリシー』

一群のノードがクラスター内のその他のノードと通信できない場合、サイト間でクラスター分割イベントが発生することがあります。例えば、リンク・クラスター内では、2 つのサイト間のすべての通信リンクで障害が発生すると分割が発生します。クラスター分割イベントにより、クラスターは 2 つ以上の区画に分割されます。

24 ページの『分割ポリシーおよびマージ・ポリシーのタイ・ブレイカー・オプション』

タイ・ブレイカー・オプションを使用して、分割ポリシーおよびマージ・ポリシーで使用される SCSI ディスクまたはネットワーク・ファイルシステム (NFS) ファイルを指定できます。

29 ページの『PowerHA SystemMirror リンク・クラスター』

各サイトが地理上の異なる場所にある状況には、リンク・クラスターは理想的です。通常、これらのサイトは離れすぎているため、共通のストレージ・エリア・ネットワーク (SAN) をうまく共有できません。それぞれのサイトには、固有でアクティブなリポジトリ・ディスク、および何らかのバックアップ・リポジトリを備えている必要があります。リンク・クラスターはサイト間の通信に必ずユニキャストを使用します。リンク・クラスターは、高可用性災害復旧 (HADR) の有用な一部分です。

関連情報:

分割とマージのポリシーの構成

分割ポリシーおよびマージ・ポリシーのタイ・ブレーカー・オプション

タイ・ブレーカー・オプションを使用して、分割ポリシーおよびマージ・ポリシーで使用される SCSI ディスクまたはネットワーク・ファイルシステム (NFS) ファイルを指定できます。

タイ・ブレーカー・ディスクまたは NFS ファイルは、クラスター内のサイトが相互に通信できなくなると使用されます。このように通信障害が発生すると、クラスターでサイトが 2 つの独立した区画に分割されます。クラスター通信リンクが応答しないために障害が発生した場合は、両方の区画がタイ・ブレーカー・ディスクまたは NFS ファイルをロックしようとします。タイ・ブレーカー・ディスクを取得する区画は、引き続き機能しますが、他方の区画では、選択したアクション・プランに応じてリブートされるかクラスター・サービスが再始動されます。

タイ・ブレーカーとして識別されるディスクまたは NFS マウントされたファイルには、クラスター内のすべてのノードからアクセスできなければなりません。

クラスターの一部であった区画が、通信障害後にオンラインに復帰する場合、その区画は、タイ・ブレーカー・ディスクまたは NFS ファイルを保持した区画と通信できなければなりません。オンラインに復帰する区画がタイ・ブレーカー・ディスクまたは NFS ファイルと通信できない場合、その区画はクラスターに結合されません。タイ・ブレーカー・ディスクまたは NFS ファイルは、構成内のすべてのノードがクラスターに再結合されると解放されます。

分割/マージ回復処理用にタイ・ブレーカー・ディスクを構成する場合は、ディスクも **devsrc** コマンド (AIX オペレーティング・システムの一部) によってサポートされていなければなりません。タイ・ブレーカー・ディスクの選択に使用される SMIT インターフェースで、この要件に適合しないディスクがすべてフィルタリングされて除外されます。EMC ディスクはこの要件に適合しません。

EMC PowerPath ディスクはタイ・ブレーカー・ディスクとしての使用にはサポートされていません。

クラスターに 3 つ以上のノードがある場合のタイ・ブレーカー・オプションに関する制限

3 つ以上のノードがあるクラスターでクラスターが分割されると、タイ・ブレーカー・ディスクまたは NFS ファイルは、運用が継続される区画によって予約されます。運用が継続される区画から、1 つのノードがタイ・ブレーカー・ディスクまたは NFS ファイルに対して予約を行います。このノードでタイ・ブレーカー・ディスクまたは NFS ファイルに対する予約が解放されずに障害が発生すると、残りのノードはタイ・ブレーカー・ディスクまたは NFS ファイルに対する予約を取得できず、運用は継続されません。したがって、すべてのノードがリブートされます。

クラスターの分割が発生したら、タイ・ブレーカー・ディスクの予約が解放されるように、障害の発生したノードの問題を迅速に解決する必要があります。

関連概念:

20 ページの『分割ポリシー』

一群のノードがクラスター内のその他のノードと通信できない場合、サイト間でクラスター分割イベントが発生することがあります。例えば、リンク・クラスター内では、2 つのサイト間のすべての通信リンクで障害が発生すると分割が発生します。クラスター分割イベントにより、クラスターは 2 つ以上の区画に分割されます。

22 ページの『マージ・ポリシー』

クラスターには、クラスター分割ポリシーに応じて、それぞれ独立して稼働する 2 つの区画がある場合があります。PowerHA SystemMirror を使用して、区画間の通信が復元した後で、再度、区画を一緒に運用できるようにするマージ・ポリシーを構成できます。

検疫ポリシー

検疫ポリシーでは、クリティカルなリソース・グループをホストしていて以前はアクティブだったノードがクラスター分割イベントまたはノード障害の発生後に隔離されます。検疫ポリシーでアプリケーション・データが壊れたり失われたりすることはありません。

検疫ポリシーには、アクティブ・ノード停止ポリシーとディスク・フェンシングという 2 つのオプションがあります。

アクティブ・ノード停止ポリシー

クラスターが分割されると、アプリケーションをホストしている LPAR が完全に静止してからスタンバイ LPAR がアプリケーションを再びオンラインにするように、以前アクティブだった LPAR がアクティブ・ノード停止ポリシーによって停止されます。

このプロセスにより、アプリケーションはクラスター内で一度に 1 つのノードでのみ作動するようになります。PowerHA SystemMirror は、ハードウェア管理コンソール (HMC) からコマンドを実行してアクティブ LPAR をオフラインにします。LPAR ポリシーを構成して AIX オペレーティング・システムにおいて自動的にリポートが行われるようにしてある場合は、手動再始動用に PowerHA SystemMirror リソース・グループ・ブート時ポリシーがセットアップされるように構成する必要があります。

次の図は、HMC ベースのアクティブ・ノード停止ポリシー実装環境を示したものです。その図では、スタンバイ LPAR は HMC 上でコマンドを実行してアクティブ LPAR を停止します。そのコマンドによりアクティブ LPAR が正常にオフラインになると、スタンバイ LPAR がスタンバイ・ノード上でリソース・グループを開始します。

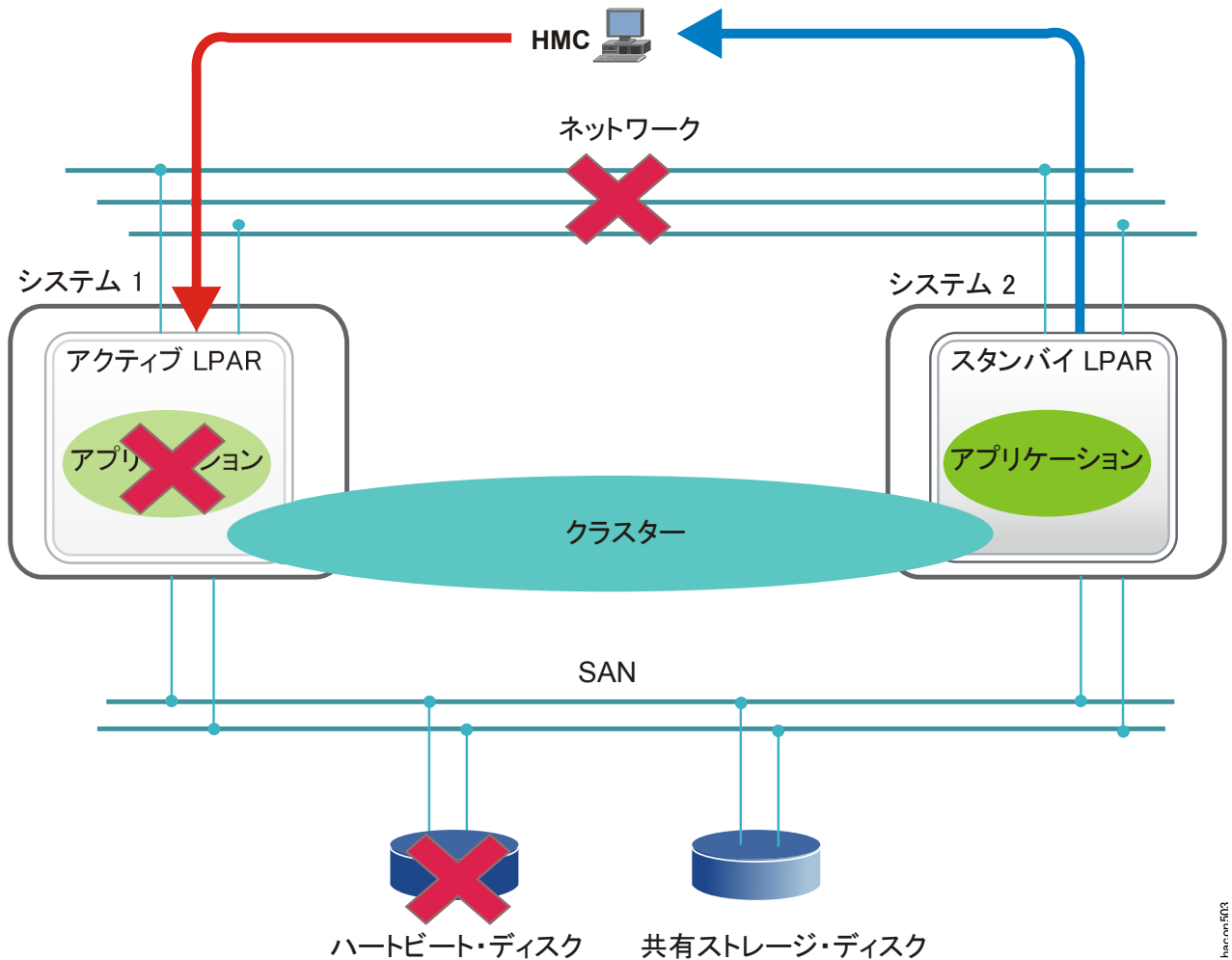


図 3. アクティブ・ノード停止ポリシー

注: エラーが発生してアクティブ LPAR がオフラインにならない場合、スタンバイ LPAR はエラー状態になり、問題修正のために手動で介入しなければなりません。

ディスク・フェンシング・ポリシー

PowerHA SystemMirror は、ディスク・フェンシング・ポリシーを使用して分割クラスターに対する保護を提供しています。ディスク・フェンシング・ポリシーは、SCSI-3 予約機能を使用して、ワークロードをホストしている問題のあるノードを、クラスターから分離します (締め出します)。このシナリオでは、問題のあるノードで実行されていたワークロードがスタンバイ LPAR で開始されます。

フェンシング・プロセスにより、スタンドアロン・ノードによるディスクへのアクセス、およびデータの保護状態の維持が確実に行われるようになります。ディスク・フェンシング・ポリシーはアクティブ/パッシブ・デプロイメント・モデルに対してサポートされています。ディスク・フェンシングにより、書き込み権限をクラスターにある 1 つのノードのみに限定してワークロードを実行できるようになります。

PowerHA SystemMirror は、任意のリソース・グループに属するすべてのボリューム・グループのディスクを登録します。

次の図は、ディスク・フェンシング・ポリシーが使用されているときにクラスターが分割されると何が起こるのかを示しています。その図では、スタンバイ LPAR は共有ストレージ・ディスクと通信し、アクティブ LPAR ディスクに対するアクセス権限を取り消すように要求します。共有ストレージ・ディスクは、以

前にアクティブだった LPAR が再始動されても、そのアクティブ LPAR からの書き込みアクセスをすべてブロックします。PowerHA SystemMirror がアクティブ LPAR 内のリソース・グループのディスクを締め出すことができる場合に、スタンバイ LPAR はアプリケーションやリソース・グループをオンラインにします。アクティブ LPAR 内のディスクを締め出すときにスタンバイ LPAR でエラーが発生すると、アプリケーションはオンラインにならないため、問題を修正してからリソース・グループを手動でオンラインに戻す必要があります。

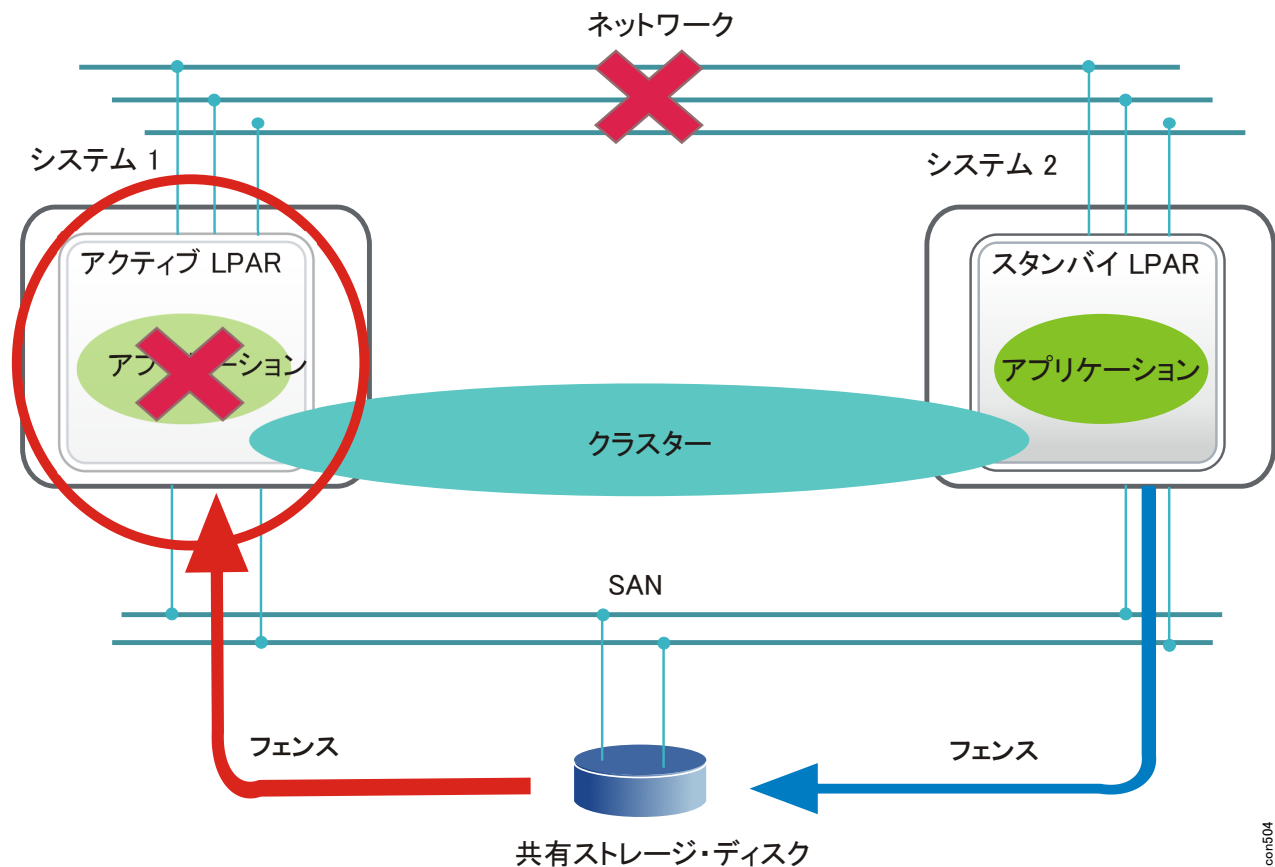


図 4. ディスク・フェンシング・ポリシー

PowerHA SystemMirror では以下の主要な属性がディスク・フェンシング・ポリシーに適用されます。

- ディスク・フェンシングはアクティブ/パッシブ・クラスター環境にのみ適用されます。
- 始動ポリシー「使用可能なすべてのノードでオンライン」が使用されるリソース・グループに対してディスク・フェンシングはサポートされていません。
- ディスク・フェンシングはクラスター・レベルでサポートされています。そのため、ディスク・フェンシング・ポリシーはクラスター・レベルで有効/無効にできます。
- ディスク・フェンシングでは、リソース・グループに含まれるボリューム・グループに属するすべてのディスクが管理されます。
- ディスク・フェンシングは相互テークオーバー構成でサポートされています。クラスターに複数のリソース・グループが存在する場合は、最もクリティカルなリソース・グループとなるリソース・グループを 1 つ選択する必要があります。クラスター分割イベントが発生すると、クリティカルなリソース・グループの相対位置によって、いずれの区画の運用が継続されるのかが決まります。運用が継続される区画になるのは、クラスター分割イベントが発生する前にクリティカルなリソース・グループが実行されていなかった区画です。

PowerHA SystemMirror は、クラスターから得られる可能な限り多くの情報を使用して LPAR の正常性を判別します。例えば、アクティブ LPAR がクラッシュしそうになっていた場合、PowerHA SystemMirror は、そのアクティブ LPAR がオフラインになる前にメッセージをスタンバイ LPAR に送信します。そのような通知によって、スタンバイ LPAR はアクティブ LPAR がオフラインになったことを確実に認識するようになり、さらにスタンバイ LPAR はアプリケーションを確実にオンラインにできるようになります。

注: 場合によっては、スタンバイ LPAR は、アクティブ LPAR がハートビートを送信していないことを認識していても、アクティブ LPAR の現状を判別できないことがあります。その場合、スタンバイ LPAR は、「ノード障害検出タイムアウト」フィールドで指定された期間、待機した後で、アクティブ LPAR で障害が発生したことを示します。このとき、スタンバイ LPAR は、リソース・グループをオンラインにする前にすべてのディスクを締め出します。ディスクが 1 つでも正しく締め出されない場合、リソース・グループはオンラインになりません。

PowerHA SystemMirror マルチサイト・ソリューション

PowerHA SystemMirror は、異なるタイプのサイト定義、および高可用性災害時リカバリー (HADR) のサイト固有のポリシーをサポートしています。PowerHA SystemMirror Standard Edition for AIX および PowerHA SystemMirror Enterprise Edition for AIX の両方で、複数のサイトを定義することができます。

PowerHA SystemMirror 管理インターフェースを使用して、以下のマルチサイト・ソリューションを作成することができます。

拡張クラスター

同じ地理的ロケーションにあるサイトのノードを含みます。拡張クラスターは、ストレージ複製管理による HADR をサポートしません。

リンク・クラスター

異なる地理的ロケーションにあるサイトのノードを含みます。リンク・クラスターは、サイト間 LVM ミラーリングおよび HyperSwap をサポートします。

次の表は、拡張クラスターとリンク・クラスターの相違を示しています。

表 1. 拡張クラスターとリンク・クラスターの相違

機能	拡張クラスター	リンク・クラスター
サイト通信	マルチキャスト	ユニキャスト
リポジトリー・ディスク	共用	分離
クラスター通信	<ul style="list-style-type: none"> • ネットワーク • ストレージ・エリア・ネットワーク (SAN) • ディスク 	ネットワーク
クロス間論理ボリューム・マネージャー・ミラーリング	使用可能	SAN で使用可能
HyperSwap	使用可能	SAN で使用可能
HyperSwap によるコンカレント・リソース・グループ	使用可能	SAN で使用可能

関連概念:

4 ページの『PowerHA SystemMirror での Cluster Aware AIX の使用』

PowerHA SystemMirror は、AIX オペレーティング・システムでサポートされる主要なクラスター機能の最上位に構築されます。PowerHA SystemMirror は、Cluster Aware AIX (CAA) 機能に対応した AIX のすべてのエディションでサポートされます。

PowerHA SystemMirror リンク・クラスター

各サイトが地理上の異なる場所にある状況には、リンク・クラスターは理想的です。通常、これらのサイトは離れすぎているため、共通のストレージ・エリア・ネットワーク (SAN) をうまく共有できません。それぞれのサイトには、固有でアクティブなリポジトリ・ディスク、および何らかのバックアップ・リポジトリを備えている必要があります。リンク・クラスターはサイト間の通信に必ずユニキャストを使用します。リンク・クラスターは、高可用性災害復旧 (HADR) の有用な一部分です。

Cluster Aware AIX (CAA) は、リンク・クラスターを使用して、地理的に遠く離れた複数の (異なる都市にある) サイトをサポートできます。リンク・クラスターは、個々の CAA クラスターを、異なる場所にあるサイトにリンクします。サイト間のリンクが、ハートビートおよびクラスター通信に使用されます。

リンク・クラスターは、同じ都市内の異なる建物の間など、地理的に短い距離間の通信に使用することができます。

次の表は、分割ポリシー・オプションおよびマージ・ポリシー・オプションに対してサポートされているリンク・クラスター構成を示したものです。例えば、リンク・クラスターでは、構成に分割ポリシー値「なし」およびマージ・ポリシー値「多数決」または「優先順位」を指定できます。ただし、構成に分割ポリシー値「タイ・ブレーカー」とマージ・ポリシー値「手動」は指定できません。

表 2. 分割ポリシーおよびマージ・ポリシーのリンク・クラスター・オプション

分割ポリシー・オプション	マージ・ポリシー・オプション			
	多数決	優先順位	タイ・ブレーカー	手動
なし	X	X		
タイ・ブレーカー			X	
手動				X

次の図は、リンク・クラスターの各サイトが別々の都市にあり、そのサイトごとに独自のリポジトリ・ディスクがあることを示しています。

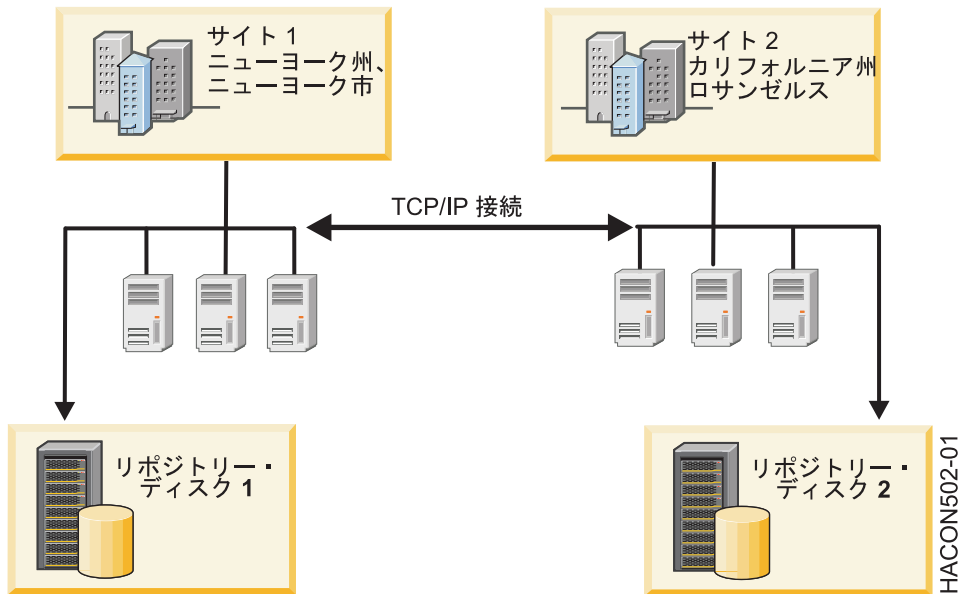


図 5. 代表的なリンク・クラスター構成

関連概念:

4 ページの『PowerHA SystemMirror での Cluster Aware AIX の使用』

PowerHA SystemMirror は、AIX オペレーティング・システムでサポートされる主要なクラスター機能の最上位に構築されます。PowerHA SystemMirror は、Cluster Aware AIX (CAA) 機能に対応した AIX のすべてのエディションでサポートされます。

『PowerHA SystemMirror 拡張クラスター』

拡張クラスターは、それぞれのサイトが地理上の同じ場所にある場合に最適です。通常、すべてのノードは共通のストレージ・エリア・ネットワーク (SAN) 上にあります。少なくとも、アクティブ・リポジトリおよび任意のバックアップ・リポジトリは、どちらのサイトでも稼働可能な、リソース・グループ内の任意のディスクが共有されるように、サイト間で共有されます。拡張クラスターは、サイト間 LVM ミラーリング、および HyperSwap をサポートできます。

14 ページの『クラスター・リポジトリ・ディスク』

PowerHA SystemMirror は、クラスターの構成を管理するために中央リポジトリとして共用ディスクを使用します。このディスクには、標準クラスター・デプロイメント方式または拡張クラスター・デプロイメント方式を使用してクラスター内のすべてのノードがアクセスできなければなりません。あるいは、リンク・クラスター・デプロイメント方式を使用してサイト内のすべてのノードがアクセスできなければなりません。

PowerHA SystemMirror 拡張クラスター

拡張クラスターは、それぞれのサイトが地理上の同じ場所にある場合に最適です。通常、すべてのノードは共通のストレージ・エリア・ネットワーク (SAN) 上にあります。少なくとも、アクティブ・リポジトリおよび任意のバックアップ・リポジトリは、どちらのサイトでも稼働可能な、リソース・グループ内の任意のディスクが共有されるように、サイト間で共有されます。拡張クラスターは、サイト間 LVM ミラーリング、および HyperSwap をサポートできます。

ご使用の環境で拡張クラスターを使用する場合、単一の Cluster Aware AIX (CAA) がクラスター内の複数のサイト間でデプロイされます。

IP 通信が失敗し、環境で拡張クラスターを使用している場合、そのクラスター内のすべてのノードがディスク・ハートビート機能を使用して通信でき、環境を操作可能に保つことができます。

拡張クラスターは、以下の要件を満たしている必要があります。

- サイト通信でユニキャストまたはマルチキャストの IP アドレスを使用できます。
- サイト間で少なくとも 1 つのリポジトリ・ディスクを共有します。

次の表は、分割ポリシー・オプションおよびマージ・ポリシー・オプションに対してサポートされている拡張クラスター構成を示したものです。例えば、拡張クラスターでは、構成に分割ポリシー値「なし」およびマージ・ポリシー値「多数決」を指定できます。ただし、構成に分割ポリシー値「タイ・ブレーカー」とマージ・ポリシー値「多数決」は指定できません。

表 3. 分割ポリシーおよびマージ・ポリシーの拡張クラスター・オプション

分割ポリシー・オプション	マージ・ポリシー・オプション	
	多数決	タイ・ブレーカー
なし	X	
タイ・ブレーカー		X

次の図は、異なる建物にある 2 つのサイトを持つ拡張クラスターの例です。これらのサイトは同じ都市内の地理的に近い場所にあり、リポジトリ・ディスクを共有しています。

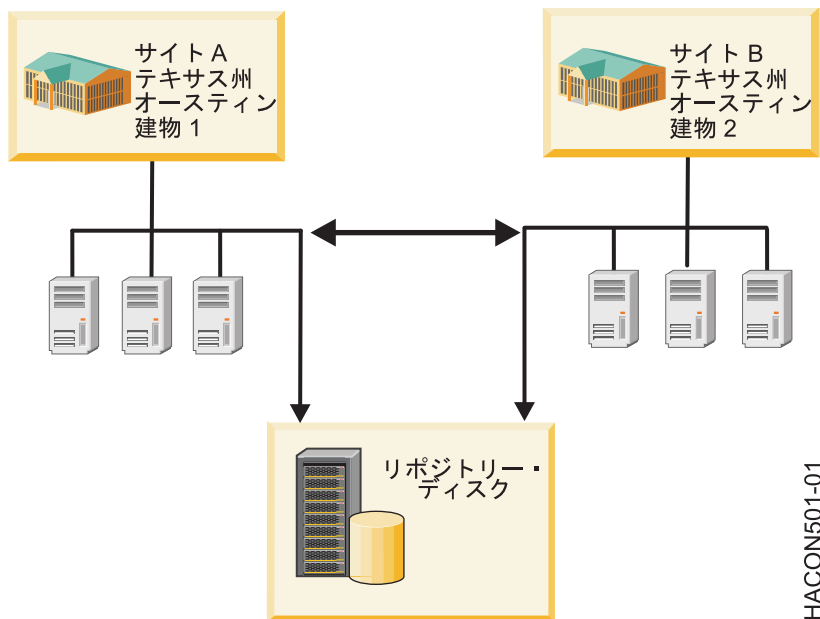


図 6. 代表的な拡張クラスター構成

関連概念:

29 ページの『PowerHA SystemMirror リンク・クラスター』
 各サイトが地理上の異なる場所にある状況には、リンク・クラスターは理想的です。通常、これらのサイトは離れすぎているため、共通のストレージ・エリア・ネットワーク (SAN) をうまく共有できません。それぞれのサイトには、固有でアクティブなリポジトリ・ディスク、および何らかのバックアップ・リポジトリを備えている必要があります。リンク・クラスターはサイト間の通信に必ずユニキャストを使用しま

す。リンク・クラスターは、高可用性災害復旧 (HADR) の有用な一部分です。

4 ページの『PowerHA SystemMirror での Cluster Aware AIX の使用』

PowerHA SystemMirror は、AIX オペレーティング・システムでサポートされる主要なクラスター機能の最上位に構築されます。PowerHA SystemMirror は、Cluster Aware AIX (CAA) 機能に対応した AIX のすべてのエディションでサポートされます。

14 ページの『クラスター・リポジトリ・ディスク』

PowerHA SystemMirror は、クラスターの構成を管理するために中央リポジトリとして共用ディスクを使用します。このディスクには、標準クラスター・デプロイメント方式または拡張クラスター・デプロイメント方式を使用してクラスター内のすべてのノードがアクセスできなければなりません。あるいは、リンク・クラスター・デプロイメント方式を使用してサイト内のすべてのノードがアクセスできなければなりません。

クラスター・ネットワーク

クラスター・ノードは通信ネットワークを通して相互に通信します。

ネットワーク上のノードにある物理ネットワーク・インターフェース・カードの 1 つに障害が発生すると、PowerHA SystemMirror は、同一ノード上の他の物理ネットワーク・インターフェース・カードにトラフィックを転送することによって、そのノードへの通信を保持します。そのノードへの「接続」に失敗した場合、PowerHA SystemMirror はアクセス可能な別のノードにリソースを転送します。

さらに、クラスター・ノード自体の正常性を定期的にチェックするため、RSCT はクラスター・ネットワークを通してノード間にハートビートを送信します。PowerHA SystemMirror がノードからのハートビートを検出しなければ、ノードに障害が発生したとみなされ、リソースは自動的に他のノードに転送されます。

クラスター内のノード間に複数の通信パスを構成することを強くお勧めします。通信ネットワークが複数あれば、クラスターの区分化、つまり各区画内のノードが独自のエンティティを形成することを防ぎます。区分化されたクラスターでは、各区画のノードは同期をとらずに同じデータに同時にアクセスすることが可能です。そうすると、ノードによって異なるデータが表示される可能性があります。

関連概念:

37 ページの『TCP/IP ネットワークおよび Storage Area Network を使用したハートビート』

ハートビート は、ノード間で送信される一種の通信パケットです。ハートビートはノード、ネットワーク、およびネットワーク・インターフェースの正常性をモニターしたり、クラスターの区分化を防止するために使用します。

物理ネットワークおよび論理ネットワーク

物理ネットワークは、2 つ以上の物理ネットワーク・インターフェースを接続します。

前のセクションで述べているとおり、複数の TCP/IP ベースのネットワークの構成は、クラスターの区分化を防ぐのに役立ちます。PowerHA SystemMirror は、必要に応じてストレージ・ネットワークを越えて通信します。この追加の通信パスは、TCP/IP ベースのネットワーク接続がクラスター・ノード間で輻輳状態または切断状態になったとき、追加の通信パスを提供することによりクラスターの区分化を防ぐのに役立ちます。

注: クラスター内の物理ネットワークで、ネットワーク間のパケットの送信に外部ネットワークング・デバイスを使用する場合は、以下の点を考慮してください。PowerHA SystemMirror クラスターを構成するとき、PowerHA SystemMirror は特定の物理ネットワーク上に定義されたすべてのインターフェースへの接続およびアクセスを検証します。ただし、PowerHA SystemMirror には、クラスター・ノード間のネ

ネットワーク・パスにブリッジやルーターなどの外部ネットワーク・デバイスがあることを判断する手段がありません。ネットワークに外部ネットワーク・デバイスがある場合は、使用するデバイスが PowerHA SystemMirror クラスターで単一障害点とならないようにするため、そのデバイスが高可用性かつ冗長であることを確認してください。

論理ネットワークは、2 つ以上の論理ネットワーク・インターフェースまたはデバイスに接続する物理ネットワークの一部です。論理ネットワーク・インターフェースまたはデバイスは、オペレーティング・システムによって認識されるソフトウェア・エンティティです。物理ネットワーク・インターフェース/デバイスと論理ネットワーク・インターフェース/デバイスの間には 1 対 1 のマッピングがあります。各論理ネットワーク・インターフェースは、同じ論理ネットワーク上のそれぞれの論理ネットワーク・インターフェースと相互にパケットを交換できます。

論理ネットワーク上の論理ネットワーク・インターフェースのサブセット同士が、同じ物理ネットワークを共用していて、相互に通信する必要がある (その他とは通信しない) 場合は、サブネットが使用されます。サブネット・マスクは、ある論理ネットワーク・インターフェースが同じ論理ネットワーク上の他の論理ネットワーク・インターフェースにパケットを送信できるかどうかを決定する IP アドレスの一部を定義します。

PowerHA SystemMirror における論理ネットワーク:

PowerHA SystemMirror には、独自の類似した論理ネットワーク概念があります。

PowerHA SystemMirror ネットワークのすべての論理ネットワーク・インターフェースは、PowerHA SystemMirror パケットを相互に直接交換できます。各論理ネットワークは、固有の名前によって識別されます。PowerHA SystemMirror クラスター構成に対して自動ディスカバリー機能を使用した場合、PowerHA SystemMirror は、PowerHA SystemMirror 論理ネットワークをディスカバーするたびに、そのネットワークに名前 (*net_ether_01* など) を割り当てます。PowerHA SystemMirror 論理ネットワークには、1 つ以上のサブネットが含まれている場合があります。RSCT は論理サブネット間のパケット送信を行います。

関連概念:

55 ページの『PowerHA SystemMirror でサポートされるハードウェア』

さまざまなタイプの IBM ハードウェアを PowerHA SystemMirror と組み合わせて使用すれば、基本レベルの高可用性環境を実現できます。

ローカル・ネットワーク障害とグローバル・ネットワーク障害:

クラスター・ネットワークで障害が発生すると、PowerHA SystemMirror はネットワーク障害イベントを使用して、このようなケースを処理します。PowerHA SystemMirror は、ネットワーク障害イベントを監視し、ローカル・ネットワーク障害イベントおよびグローバル・ネットワーク障害イベントという 2 つのタイプのネットワーク障害イベントを区別します。

ローカル・ネットワーク障害

ローカル・ネットワーク障害とは、1 つのノードが PowerHA SystemMirror 論理ネットワーク経由でパケットを送信または受信できない PowerHA SystemMirror イベントです。このイベントが発生するのは、例えば、特定の PowerHA SystemMirror 論理ネットワークに参加しているノードのすべてのネットワーク・インターフェース・カードに障害が起きた場合などです。ローカル・ネットワーク障害が発生しても、そのネットワークはまだ他のノードが使用しています。ローカル・ネットワーク障害に対処するため、PowerHA SystemMirror は選択的に (そのネットワーク上の) リソースをノードから別のノードに移動します。この操作は選択フォールオーバーと呼ばれます。

グローバル・ネットワーク障害

グローバル・ネットワーク障害とは、あらゆるノードが PowerHA SystemMirror 論理ネットワーク経由でパケットを送信または受信できない PowerHA SystemMirror イベントです。このイベントが発生するのは、例えば、物理ネットワークが損傷した場合などです。グローバル・ネットワークとグローバル・ネットワーク障害イベントを区別することが重要です。グローバル・ネットワークは、PowerHA SystemMirror ネットワークを組み合わせたものです。グローバル・ネットワーク障害イベントは、グローバル・ネットワークに限らず、論理 PowerHA SystemMirror ネットワークに接続されているすべてのノードに影響する障害を指します。

PowerHA SystemMirror 通信インターフェース

PowerHA SystemMirror 通信インターフェースとは、PowerHA SystemMirror に定義した論理ネットワーク・インターフェース、サービス IP アドレス、およびサービス IP ラベルのグループです。

PowerHA SystemMirror 通信インターフェースが結合して、IP ベース・ネットワークを作成します。

PowerHA SystemMirror 通信インターフェースとは、次のものの組み合わせです。

- 論理ネットワーク・インターフェースは、AIX が物理ネットワーク・インターフェース・カードのポート (例えば、en0) を解決する名前です。
- サービス IP アドレスは、アプリケーションなどのサービスが提供される時、あるいはクライアント・ノードが通信するときに経由する IP アドレスです (例えば、129.9.201.1)。
- サービス IP ラベルとは、サービス IP アドレスにマップするラベル (例えば、`/etc/hosts` ファイル内のホスト名、または、`node_A_en_service` のようなサービス IP アドレスと論理的に等価な値) です。

PowerHA SystemMirror の通信インターフェースは次のように使用されます。

- 通信インターフェースとは、IP ベース・ネットワークおよびネットワーク・インターフェース・カード (NIC) を指します。共通の物理ネットワークに接続する NIC は、PowerHA SystemMirror が使用する論理ネットワークに結合されます。
- 各 NIC は、いくつかの TCP/IP アドレスをホストすることができます。クラスターを構成するときに、PowerHA SystemMirror がモニターする IP アドレス (基本 IP アドレスまたはブート IP アドレス)、および PowerHA SystemMirror が高可用性を保持する IP アドレス (サービス IP アドレス) を定義する必要があります。
- PowerHA SystemMirror のハートビートは、通信インターフェース間で起こります。PowerHA SystemMirror は、Cluster Aware AIX (CAA) のハートビート機能を使用して、ネットワーク・インターフェースおよび IP アドレスをモニターします。CAA は、作成されたネットワーク・トポロジーを RSCT に渡すのに対し、RSCT は障害通知を PowerHA SystemMirror に提供します。

PowerHA SystemMirror 非 IP 通信デバイス

PowerHA SystemMirror は、IP 通信ができないネットワーク・デバイスもモニターします。これらのデバイスには、Storage Area Network および共用ディスクが含まれます。

PowerHA SystemMirror でのサブネット経路指定要件

サブネット経路は、別の論理ネットワーク上のアドレスへ論理ネットワークを介してパケットを送信するために、サブネットで定義されるパスを定義します。

AIX では、カーネル経路指定テーブルで、同じ宛先に対する複数の経路を追加できます。一致する複数の経路の基準が等しい場合は、いくつかのサブネット経路のうちの 1 つを使用して経路指定が実行されます。PowerHA SystemMirror のサブネット経路指定については、以下の点を考慮する必要があります。

- PowerHA SystemMirror は、同じサブネット経路を共有する複数の論理ネットワーク・インターフェースを識別することはできません。論理ネットワーク・インターフェースが他のインターフェースと経路を共有する場合、PowerHA SystemMirror にはその正常性を判断する手段はありません。ネットワーク経路の詳細については、**route** コマンドに関する AIX のマニュアル・ページを参照してください。
- IP ベース・ネットワークには、ネットワーク管理者から、あるいは TCP/IP 要件に基づいて、さまざまな制約が課せられることがよくあります。サブネットおよび経路も、PowerHA SystemMirror を構成し運用する上で守らなければならない制約です。

注: PowerHA SystemMirror が各インターフェースをモニターできるように、ノード上の各通信インターフェースが固有のサブネットに属するように構成してください。これはすべての場合において厳守すべき要件というわけではなく、いくつかの要因によって変わってきます。例えば、仮想ネットワーク環境などで、ノードごとに 1 つのネットワーク・インターフェースしかない場合は、1 つのサブネット経路であるため、ブートとサービス IP アドレスが同じサブネット上に存在する可能性があります。固有のサブネットが必要な場合、これは PowerHA SystemMirror のクラスターの検証によりエラーのフラグが立ちます。

サービス IP ラベルおよびアドレス

サービス IP ラベルは、サービス IP アドレスにマップするラベルです。これは、クライアント・ノードとサーバー・ノードの間で通信を確立するために使用されます。

データベース・アプリケーションなどのサービスは、サービス IP ラベルを使用して確立された接続を使用して提供されます。サービス IP ラベルは、リソースとしてリソース・グループに入れることができます。これにより、PowerHA SystemMirror は、1 つのノード内で、または IP アドレス・テークオーバーが構成されている場合は (障害発生時に別のノードにそのリソースを転送することにより) クラスター・ノード間で、そのリソースの正常性をモニターしたり、そのリソースの高可用性を保持したりできるようになります。

IP エイリアス

IP エイリアスとは、ネットワーク・インターフェース・カード上に最初から構成されていた IP ラベルまたは IP アドレスに加えて、そのネットワーク・インターフェース・カード (NIC) に構成される IP ラベルまたは IP アドレスです。

IP エイリアスは、PowerHA SystemMirror でサポートされる AIXAIX 機能です。AIX オペレーティング・システムは、NIC 上の複数の IP エイリアスをサポートします。NIC の各 IP エイリアスは、別々のサブネット上に構成できます。

IP エイリアスは、PowerHA SystemMirror で IP アドレス・テークオーバーのためのサービス・アドレスとして使用されます。

以下のトピックでは、PowerHA SystemMirror が、サービス IP ラベルのリカバリーに使用されるメカニズムに応じてサービス IP ラベルをどのように通信インターフェースにバインドするかについて説明します。

IP アドレス・テークオーバー

IP アドレス・テークオーバーは、最初の NIC の障害発生時に、サービス IP ラベルを他のノードの他のネットワーク・インターフェース・カード (NIC) に移動することによって、サービス IP ラベルを回復するメカニズムです。

1 つのノード上の物理ネットワーク・インターフェース・カードに障害が発生し、同じノードの同じネットワーク上にアクセス可能な物理ネットワーク・インターフェース・カードが他にない場合、IPAT が発生し

ます。そのため、同じノード内のこれらの NIC の IP ラベルをスワッピングできず、PowerHA SystemMirror はバックアップ・ノード上の NIC を用いてサービス IP アドレスをリカバリーするために IPAT を使用します。IP アドレス・テークオーバーでは、障害発生後に IP アドレスをリカバリーすることにより IP アドレスの高可用性を維持します。PowerHA SystemMirror は IP エイリアスを介して IPAT と呼ばれるメソッドを使用します。

IPAT およびサービス IP ラベル

IPAT はサービス IP ラベルを操作します。

IP エイリアスを介した IPAT を使用すると、サービス IP ラベルまたは IP アドレスは、同じネットワーク・インターフェースにエイリアスとして付けられるか、追加アドレスとして追加されます。つまり、同じネットワーク・インターフェース上に同時に複数の IP アドレスまたは IP ラベルが構成されます。この構成では、ユーザーが定義する IP アドレスまたはラベルはすべて異なるサブネット上に構成される必要があります。ただし、ノードごとに 1 つのネットワーク・インターフェースがある場合を除きます。この方法を使用すればハードウェアを節約できますが、追加のサブネットが必要になります。

IP エイリアスによる IP アドレス・テークオーバー

AIX オペレーティング・システムの IP エイリアス・ネットワーク機能を使用して、特定タイプのネットワーク上で IP アドレス・テークオーバーを構成できます。

IP エイリアスをネットワーク・インターフェースに定義すると、同じネットワーク・インターフェース上に複数の IP ラベルと IP アドレスを作成できます。IP エイリアスによる IPAT では、数多くのタイプのネットワーク上で使用可能な無償のアドレス解決プロトコル (ARP) 機能が使用されます。

サービス IP ラベルを含むリソース・グループが 1 次ノードからターゲット・ノードにフォールオーバーすると、使用可能な NIC 上で、基本 IP アドレスに加えて、サービス IP ラベルがエイリアス・アドレスとして追加 (および除去) されます。このため、単一の NIC が、その NIC 上の複数のサービス IP ラベルをエイリアスとしてサポートできます。したがって、同一のノードで複数のリソース・グループを同時にホストすることが可能です。

同じネットワークに接続された同じノード上に複数のインターフェースがあり、それらのインターフェースがイーサネット・アグリゲーションに結合されていない場合、すべてのブート・アドレスは異なるサブネット上になければなりません。また、永続アドレスまたはサービス・アドレスは、ブート・アドレスとは異なるサブネット上になければなりません。

IP エイリアスを使用すると、同じネットワーク・インターフェース上に複数のサービス・ラベルを共存させることができるため、クラスター内で使用する物理ネットワーク・インターフェース・カードの数を減らすことができます。フォールオーバー時に、PowerHA SystemMirror は、使用可能なネットワーク・インターフェース・カード間で平等にエイリアスを分配します。

サービス IP ラベル・エイリアスの配布設定

PowerHA SystemMirror は IP エイリアス・メソッドを介して IP アドレス・テークオーバー (IPAT) を使用し、リソース・グループ内のサービス IP ラベルの高可用性を保ちます。

クラスター始動時には、PowerHA SystemMirror はデフォルトで最小ロードの原則を使用して、ネットワーク上のすべての使用可能なブート・インターフェースを介してすべてのサービス IP ラベル・エイリアスを配布します。PowerHA SystemMirror は、最小数のエイリアスまたは永続 IP ラベルが既に割り当てられているインターフェースに、新規サービス・アドレスを割り当てます。

ただし、場合によっては、始動時のみでなくその後のクラスター・イベントでも、他のタイプの割り当てを指定したり、ラベルを特定の方法で継続して割り当てたりする必要があります。

例えば、すべてのサービス IP ラベル・エイリアスを、該当ノードの永続 IP ラベルを現在ホストしているのと同じブート・インターフェースに割り当ててこともあります。このオプションは、VPN ファイアウォールの構成の際に、1 つのインターフェースのみに外部接続を許可し、接続を有効にするのと同じインターフェース上にすべての IP ラベル (永続 IP ラベルおよびサービス IP ラベル・エイリアス) を配置しなければならない場合に便利です。

PowerHA SystemMirror の制御下にあるサービス IP ラベルのエイリアスの配布設定を構成できます。

サービス IP ラベル・エイリアスの配布設定はネットワーク全体の属性で、クラスター内のノードの物理ネットワーク・インターフェース・カードに対するサービス IP ラベル・エイリアスの配置を制御します。サービス IP ラベル・エイリアスの配布設定の構成では、以下を行います。

- ノードで既に割り当てられている永続 IP ラベルを考慮しながら、クラスター内のサービス IP ラベルのロード・バランシングをカスタマイズできます。
- 指定した設定に従ってエイリアス・サービス IP ラベルを再配分するために PowerHA SystemMirror を有効にします。
- VPN ファイアウォールの外部接続要件に適した配布設定のタイプを構成できます。
- サービス IP ラベルは別のネットワーク・インターフェースに移行する可能性もありますが、PowerHA SystemMirror では、指定した配布設定に応じてラベルが継続して割り当てられるようにします。言い換えると、配布設定は、始動とその後のクラスター・イベント (フォールオーバー、フォールバック、または同じノード上でのインターフェースの変更など) の間、維持されます。例えば、複数のラベルを同一のインターフェースにマップするよう指定した場合、最初に構成されたサービス IP ラベルが別のノードに移ったとしても、ラベルは同じインターフェースにマップされた状態を維持します。
- 配布設定は、許容されるネットワーク・インターフェースがクラスター内で有効である限りは実行されます。PowerHA SystemMirror は、設定が条件に合わない場合でも、常にサービス IP ラベルをアクティブに保ちます。

関連概念:

116 ページの『永続ノード IP ラベル』

永続的ノード IP ラベル は、PowerHA SystemMirror クラスター・サービスがノード上で停止している場合でも、そのノードに接続できるようにする有用な管理ツールです。

関連情報:

Administering PowerHA SystemMirror

PowerHA SystemMirror プランニング・ガイド

TCP/IP ネットワークおよび Storage Area Network を使用したハートビート

ハートビート は、ノード間で送信される一種の通信パケットです。ハートビートはノード、ネットワーク、およびネットワーク・インターフェースの正常性をモニターしたり、クラスターの区分化を防止するために使用します。

PowerHA SystemMirror クラスターは、障害を認識して、それに対応するために、クラスターの正常性を継続的に検査しなければなりません。このようなチェックの一部がハートビート機能で提供されます。

各クラスター・ノードは、一定の間隔で他のクラスター・ノードにハートビート・メッセージを送信し、また他のノードから一定の間隔でハートビート・メッセージを受信します。メッセージが受信されなくなると、PowerHA SystemMirror は障害が発生したと認識します。ハートビートは、次の経路で送信できます。

- TCP/IP ネットワーク
- Storage Area Network
- クラスター・リポジトリ・ディスク

Cluster Aware AIX (CAA) は、使用可能なすべての TCP/IP ネットワークおよびストレージ・エリア・ネットワーク (SAN) 上で、ハートビート通信を使用します。TCP/IP ネットワークおよび SAN ネットワークに障害が発生した場合、CAA は、リポジトリ・ディスクを代替のハートビート・メカニズムとして使用しようとしています。バックアップのリポジトリ・ディスクのハートビート・パスは、**lscluster** コマンドの出力に、**dpcom** インターフェースとして表示されます。TCP/IP ネットワークおよび SAN ネットワークが機能している場合、**lscluster -i** コマンドは、**dpcom** インターフェースを **restricted** として表示します。

ハートビート機能は、ノード間の特定のパスを使用するように構成されます。これにより、ハートビートは、クラスター・ノード自体の正常性だけでなく、すべての PowerHA SystemMirror ネットワークおよびネットワーク・インターフェースの正常性をモニターできるようになります。

ハートビート・パスは CAA によって自動的にセットアップされます。Point-to-Point パスおよびディスク・パスを PowerHA SystemMirror 構成の一部として構成するオプションもあります。

関連概念:

32 ページの『クラスター・ネットワーク』

クラスター・ノードは通信ネットワークを通して相互に通信します。

インターネット・プロトコル・ネットワークを使用したハートビート

PowerHA SystemMirror は Cluster Aware AIX (CAA) に依存して、インターネット・プロトコル・ネットワークを介してクラスター・ノード間にハートビートを提供します。デフォルトで、CAA はハートビートにユニキャスト通信を使用します。ネットワークがマルチキャストをサポートするように構成されている場合は、オプションとしてマルチキャスト通信の使用を選択できます。

PowerHA SystemMirror 7.1.2 以降は、インターネット・プロトコル・バージョン 6 (IPv6) をサポートしています。ただし、IPv6 マルチキャスト・アドレスを明示的に指定することはできません。CAA は、インターネット・プロトコル・バージョン 4 (IPv4) マルチキャスト・アドレスから派生する IPv6 マルチキャスト・アドレスを使用します。IPv6 マルチキャスト・アドレスを判別するには、**0xFF05** の標準プレフィックスを、論理 OR 演算子を使用して IPv4 アドレスの 16 進の等価部と結合します。例えば、IPv4 マルチキャスト・アドレスが **228.8.16.129** または **0xE4081081** であるとし、標準プレフィックスを用いて論理 OR 演算で変換すると、**0xFF05:: | 0xE4081081** となります。したがって、結果の IPv6 マルチキャスト・アドレスは **0xFF05::E408:1081** です。

CAA は、(PowerHA SystemMirror トポロジーで構成されたインターフェースだけでなく) 構成済みのすべての AIX インターフェースに対して、ハートビートをデフォルトで提供します。

注: 特定のネットワーク上の CAA ハートビートを停止するには、**public** の代わりに **private** と定義します。この注記は、IPv4 インターフェースと IPv6 インターフェースの両方に適用されます。

関連概念:

34 ページの『PowerHA SystemMirror でのサブネット経路指定要件』

サブネット経路は、別の論理ネットワーク上のアドレスへ論理ネットワークを介してパケットを送信するために、サブネットで定義されるパスを定義します。

関連情報:

IP ベース・ネットワークの変更手順

ストレージ・エリア・ネットワークを使用したハートビート

Cluster Aware AIX (CAA) は、複数のクラスター・ノードにまたがる共通ストレージ・エリア・ネットワーク (SAN) を介してハートビートを自動的に交換します。このように接続すれば、単一の TCP/IP ベース・ネットワークを使用するクラスターに代替のハートビート・パスを提供できます。


PowerHA SystemMirror マルチキャスト


PowerHA SystemMirror は、Cluster Aware AIX (CAA) と呼ばれる、AIX オペレーティング・システムの一部として組み込まれたクラスター正常性管理層を使用します。CAA は、AIX カーネル・レベルのコードを使用し、ファイバー・チャネル・アダプター、および中央リポジトリ経由のディスク・ベースのメッセージングを用いて、ネットワークを介してハートビートを交換します。

デフォルトで、CAA はハートビートにユニキャスト通信を使用します。ネットワークがマルチキャストをサポートするように構成されている場合は、オプションとしてマルチキャスト通信の使用を選択できます。クラスター通信では、マルチキャスト・アドレスを手動で構成することも、CAA に自動的にマルチキャスト・アドレスを選択させることもできます。

クラスターを最初に作成する時点ではマルチキャスト通信は使用されませんが、クラスターの通常の操作に必要です。マルチキャスト通信がクラスター内のすべてのノード間で、ネットワークを介して転送可能であることをテストし、検査する必要があります。

関連情報:

 Administratively Scoped IP Multicast

 マルチキャスト・アドレスの範囲および割り当て

 マルチキャスト・ルーター情報

マルチキャスト・パケット通信

マルチキャストは、ノードがグループを形成してメッセージを交換する、アドレス指定の形式です。

グループ内の 1 つのノードによって送信されるマルチキャスト・メッセージは、グループ内のその他すべてのノードで受信されます。この機能によって、効率的なクラスター通信が可能になります。例えば、クラスター内のノードは単一のマルチキャスト・パケットを送信して、重要なイベントについて他のノードに通知することができます。

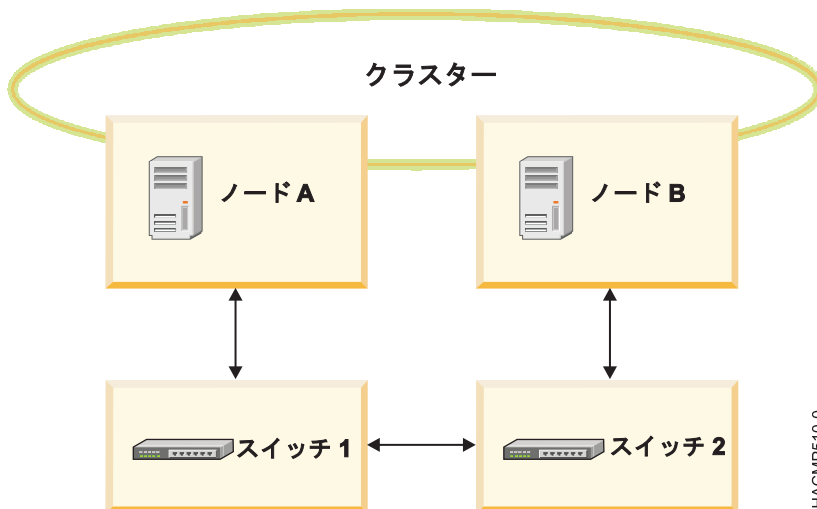
マルチキャスト・ネットワーク・スイッチ

スイッチはさまざまなノードおよびネットワーク・セグメントを接続し、ネットワーク・データを正しいノードに送信します。スイッチは、OSI (Open Systems Interconnection) モデルのデータ・リンク層 (レイヤー 2) でデータを処理および経路指定する、マルチポート・ネットワーク・ブリッジです。一部のスイッチは、ネットワーク層 (レイヤー 3) でデータを処理することもできます。

通常、データ・センター環境では、複数のノードがスイッチによって相互接続されています。スイッチがカスケードされると、マルチキャスト・パケットがスイッチを介して 1 つのノードから送信され、クラスター

一内の最終的な宛先ノードに到達するまで次々に別のスイッチに送られます。スイッチは、通常のネットワーク通信とは異なる方法でマルチキャスト・パケットを処理します。そのため、ネットワークが正しくセットアップされていないと、マルチキャスト・パケットのスイッチからスイッチへの通信は行われないことがあります。

次の図は、ノード A がスイッチ 1 とスイッチ 2 を介してノード B に接続されているクラスターの一部である構成を示しています。マルチキャスト・パケットがノード A とノード B の間を流れるためには、スイッチ 1 とスイッチ 2 の間のネットワーク接続でマルチキャスト通信が使用可能になっている必要があります。



IGMP (Internet Group Management Protocol)

Internet Group Management Protocol (IGMP) は、ノード (受信側) がマルチキャスト・ルーター (IGMP 照会側) に対し、そのノードが特定のマルチキャスト・トラフィックを受信する意図があることを伝えられる通信プロトコルです。

IGMP は、以下のアクションを行えるルーターとノード間で実行されます。

- ルーターがノードに、特定のマルチキャスト・ストリームが必要かどうか尋ねます (IGMP 照会)。
- ノードがルーターに、特定のマルチキャスト・ストリームを探しているかどうか応答します (IGMP レポート)。

対話をし、マルチキャスト通信の基本原則およびマルチキャスト・グループ・メンバーシップを確立するために、IP ネットワーク上のノードおよび隣接するルーターによって IGMP 通信プロトコルが使用されます。

IGMP スヌープ

IGMP スヌープは、IGMP 通信に関連したパケット交換を追跡して、マルチキャスト・パケットのフィルタリングに適合するために、スイッチによって実行されるアクティビティーです。IGMP スヌープ機能を持つスイッチは、ノードおよびルーター間のこうした IGMP トランザクションを観察することによって、有用な情報を取り出します。この機能により、スイッチは必要に応じて、マルチキャスト・パケットをネットワーク・パス内の次のスイッチに正しく転送することができます。

スイッチは IGMP トラフィックをモニターし、必要があればマルチキャスト・パケットの送信のみを行います。スイッチは、通常、特定のマルチキャスト・グループを要求したすべてのポートのリストを持つ IGMP スヌープ・テーブルを構築します。IGMP スヌープ・テーブルを使用することによって、マルチキャスト・パケットのネットワーク内の転送を許可したり、それらのネットワーク内の転送を不許可にしたりできます。IGMP スヌープを行えないよう、スイッチを構成することができます。

マルチキャスト経路指定

マルチキャスト経路指定は、ネットワーク・デバイスがマルチキャスト・パケットをマルチキャスト・トラフィックの目的とするすべての宛先に配信する方法を決定する処理です。

ルーターがマルチキャスト経路指定を実装しているかどうかを判別するには、ルーターの資料をお読みください。

特殊なルーティング・アルゴリズムを使用してマルチキャスト・パケットを転送するネットワーク・エンティティを、*mrouter* と呼びます。ノードおよびその他のネットワーク・エレメントが *mrouter* を実装し、マルチキャスト・トラフィックのネットワーク内の転送を可能にしています。

注：一部のルーターは、マルチキャスト・パッケージ経路指定もサポートします。

スイッチがカスケードされる場合、*mrouter* を実装するために、必要に応じてパケットを転送するようネットワーク・スイッチをセットアップする必要が生じることがあります。クラスター内でマルチキャスト通信に問題が生じる場合は、ネットワーク・デバイスが IGMP およびスヌープ用に正しく設定されていることを確認してください。マルチキャスト・トラフィック用のネットワークのセットアップについて詳細は、ご使用のネットワーク・デバイスの資料を参照してください。

PowerHA SystemMirror リソースおよびリソース・グループ

ここでは、文書全体で使用され、PowerHA SystemMirror ユーザー・インターフェースでも使用されるリソース関連の概念および定義について説明します。

クラスター・リソースの識別、およびクラスター・リソースの可用性の保持

PowerHA SystemMirror ソフトウェアは、可用性の高い環境を提供します。

PowerHA SystemMirror ソフトウェアは、次の方法でこれを行います。

- アプリケーションの操作に必要なクラスター・リソースのセットを識別し、そのリソースを結合してリソース・グループにします。
- PowerHA SystemMirror が、クラスター操作の各段階（始動、フォールオーバー、およびフォールバック）で高可用性を維持するためにリソースをいかに管理するかを示すリソース・グループ・ポリシーおよび属性を定義します。

リソースを識別し、リソース・グループ・ポリシーを定義することにより、PowerHA SystemMirror ソフトウェアは数多くのクラスター構成が可能であり、個々の要件に合わせてクラスター環境を定義する際の柔軟性は非常に高くなります。

クラスター・リソースの識別

以下のクラスター・リソースには、ハードウェア・リソースとソフトウェア・リソースの両方を含めることができます。

- ディスク

- ボリューム・グループ
- 論理ボリューム
- ファイルシステム
- サービス IP ラベルまたはアドレス
- アプリケーション
- テープ・リソース

PowerHA SystemMirror を実行するプロセッサは、ユーザー定義のリソースの集合 (ディスク、ボリューム・グループ、ファイルシステム、IP アドレス、アプリケーション) を所有します。リソースの高可用性を保持する目的で、相互依存型のリソースの集合をリソース・グループの中に構成することができます。リソース・グループにより、関連のあるリソースを構成および管理しやすい単一の論理エンティティに結合することができます。PowerHA SystemMirror ソフトウェアは、リソース・グループを 1 つの単位として扱うので、これらの相互に依存するリソースを 1 つのノードにまとめておき、その上でこれらのリソースの高可用性を保持します。

関連概念:

43 ページの『ボリューム・グループ』

ボリューム・グループとは、連続したアドレス指定可能なディスク領域として AIX が扱う物理ボリュームの集合のことです。

43 ページの『論理ボリューム』

論理ボリュームは、AIX が単一のストレージ・ユニットとして使用可能にする論理区画の集合です。つまり、論理ボリュームは物理ディスクの論理ビューです。論理区画はミラーリングをインプリメントするために 1 つ、2 つ、または 3 つの物理区画にマップされる場合があります。

43 ページの『ファイルシステム』

ファイルシステムは、単一の論理ボリュームに書き込まれます。通常は、データの管理を容易、かつ高速化するためにファイルの集合をファイルシステムとして編成します。

44 ページの『アプリケーション』

また、可用性の高いシステムの目的は、ユーザーにクリティカルなサービスへのアクセスを保証することにあります。PowerHA SystemMirror 環境では、普通、アプリケーションは変更を加えなくても実行できます。予想外のシャットダウンの後でも正常に再始動できるアプリケーションは、PowerHA SystemMirror でも使用できるアプリケーションの候補です。

45 ページの『サービス IP ラベルおよび IP アドレス』

サービス IP ラベルは、クライアント・ノードとサーバー・ノードの間で通信を確立するために使用されます。データベース・アプリケーションなどのサービスは、サービス IP ラベルを使用して確立された接続を使用して提供されます。

45 ページの『テープ・リソース』

SCSI またはファイバー・チャネル・テープ・ドライブを非コンカレント・リソース・グループのクラスター・リソースとして構成し、そのデバイスをクラスター内の 2 つのノードに対して高可用性のデバイスにできます。

クラスター・リソースのタイプ

このセクションでは、PowerHA SystemMirror で構成したり、PowerHA SystemMirror で高可用性を維持するためにリソース・グループに組み込んだりできるリソースについて概説します。

ボリューム・グループ

ボリューム・グループとは、連続したアドレス指定可能なディスク領域として AIX が扱う物理ボリュームの集合のことです。

ボリューム・グループは AIX オペレーティング・システムに対して構成され、PowerHA SystemMirror でリソース・グループに組み込むことができます。PowerHA SystemMirror 環境において、共有ボリューム・グループとは、クラスター・ノードで共用される外部ディスクに完全に含まれるボリューム・グループのことです。共有ディスクは、物理的にクラスター・ノードに付加され、すべてのクラスター・ノードに論理的に構成されたディスクです。

論理ボリューム

論理ボリュームは、AIX が単一のストレージ・ユニットとして使用可能にする論理区画の集合です。つまり、論理ボリュームは物理ディスクの論理ビューです。論理区画はミラーリングをインプリメントするために 1 つ、2 つ、または 3 つの物理区画にマップされる場合があります。

PowerHA SystemMirror 環境で論理ボリュームを使用すれば、ジャーナル・ファイルシステム (非コンカレント・アクセス) またはロー・デバイス (コンカレント・アクセス) のサポートが可能になります。コンカレント・アクセスはファイルシステムをサポートしません。コンカレント・アクセス環境のデータベースおよびアプリケーションは、ロー論理ボリューム (例えば /dev/rshared1v) にアクセスする必要があります。

共有論理ボリュームでは、PowerHA SystemMirror クラスター内に固有の名前が必要です。

注: 共有ボリューム・グループにはアクティブなページング・スペースを含めることはできません。

ファイルシステム

ファイルシステムは、単一の論理ボリュームに書き込まれます。通常は、データの管理を容易、かつ高速化するためにファイルの集合をファイルシステムとして編成します。

共有ファイルシステム

PowerHA SystemMirror システムでは、共有ファイルシステムとは、共有論理ボリュームに存在しているジャーナル・ファイルシステムを指します。非コンカレント・アクセスの場合は、クラスター・ノードによって共用される外部ディスク上に共有ファイルシステムを配置するように計画できます。データをこれらの外部共有ディスク上のファイルシステムに常駐させるのは、可用性を向上させるためです。コンカレント・アクセスの場合は、ジャーナル・ファイルシステムを使用できません。その代わりに、ロー論理ボリュームを使用します。

ジャーナル・ファイルシステムと拡張ジャーナル・ファイルシステム

拡張ジャーナル・ファイルシステム (JFS2) は、ジャーナル・ファイルシステム (JFS) よりも大容量のファイルを保管する機能を提供します。JFS2 は、64 ビット・カーネルのデフォルト・ファイルシステムです。32 ビット環境の推奨ファイルシステムである JFS と、64 ビット機能を持つ JFS2 のどちらをインプリメントするかを選択できます。JFS2 は、ファイルシステム内に持つことができるファイルの数を動的に増減できるので、JFS に比べてかなりの柔軟性があります。さらに JFS2 では、ボリューム・グループ内のすべてのファイルシステムのログに別々の論理ボリュームを割り当てるのではなく、ファイルシステム・ログをデータと同じ論理ボリュームに組み込むことができます。

アプリケーション

また、可用性の高いシステムの目的は、ユーザーにクリティカルなサービスへのアクセスを保証することにあります。PowerHA SystemMirror 環境では、普通、アプリケーションは変更を加えなくても実行できます。予想外のシャットダウンの後でも正常に再始動できるアプリケーションは、PowerHA SystemMirror でも使用できるアプリケーションの候補です。

例えば、すべての商用 DBMS 製品は、何らかのトランザクション・ジャーナル内にディスクの状態に関するチェックポイントを提供します。サーバーに障害が発生した場合、フォールオーバー・サーバーは DBMS を再始動し、データベースの整合性を取り直してから処理を再開します。

AIX Fast Connect を使用して、オペレーティング・システムとして AIX を稼働していないシステムとリソースを共有できます。Fast Connect を SystemMirror リソースとして構成する場合、PowerHA SystemMirror は、Fast Connect について高可用性を維持し、ノードとネットワークのインターフェース障害から回復します。また、PowerHA SystemMirror は、クラスター検証時に Fast Connect の構成を検証します。アプリケーションを管理するには、そのアプリケーションをアプリケーション・コントローラー・リソースとして PowerHA SystemMirror に定義します。アプリケーション・コントローラーには、アプリケーション始動スクリプトおよび停止スクリプトが組み込まれています。PowerHA SystemMirror がこれらのスクリプトを使用するのは、特定ノード上でアプリケーションをオンラインまたはオフラインにして、そのアプリケーションの可用性を高いまま保持する必要がある場合です。

注: 始動および停止スクリプトは、PowerHA SystemMirror がアプリケーションを制御できる唯一の部分です。指定されたスクリプトは、アプリケーションをすべての面で始動および停止させるように正しく機能することがきわめて重要です。スクリプトがアプリケーションを適切に制御できないと、そのアプリケーション回復の他の部分に影響する可能性があります。例えば、使用した停止スクリプトがアプリケーションを完全には停止することができず、あるプロセスがディスクにアクセスし続けている場合、PowerHA SystemMirror は障害が発生したノードのボリューム・グループをオフラインにすることも、そのボリューム・グループをバックアップ・ノードに回復することもできません。

使用するアプリケーション始動および停止スクリプトを入念にテストしてから、アプリケーション・コントローラーを PowerHA SystemMirror リソース・グループに追加するようにしてください。

アプリケーション・コントローラーが含まれるリソース・グループには、そのアプリケーションが依存するすべてのリソース (サービス IP アドレス、ボリューム・グループ、およびファイルシステムなど) も含まれます。このようなりソース・グループが作成されると、PowerHA SystemMirror はそのリソース・グループ全体を管理するので、その中にあるすべての相互に依存するリソースは単一のエンティティとして扱われます。PowerHA SystemMirror はアプリケーション回復を調整し、すべての相互に依存するリソースが他のリソースの前に確実に活動化されるような順序でリソースを管理します。

さらに、PowerHA SystemMirror には、アプリケーション・モニター機能が含まれます。この機能を使用すれば、プロセスの予期しない終了を検出、またはアプリケーションの終了を定期的にポーリングするためのモニターを定義し、問題の検出時には自動的にアクションをとるようにすることができます。

複数のアプリケーション・モニターを構成し、それらを 1 つ以上のアプリケーション・コントローラーに関連付けることができます。アプリケーションごとに複数のモニターをサポートすることにより、PowerHA SystemMirror は、より複雑な構成をサポートできます。例えば、使用中の Oracle 並列サーバーの各インスタンスについて 1 つのモニターを構成できます。または、ユーザー定義モニターを構成してデータベースの状態をチェックし、プロセス終了モニターが直ちにデータベース・プロセスの終了を検出できるようにできます。

アプリケーション・モニターのモードも指定することができます。アプリケーションがどのように実行中であるか (実行モード)、またはアプリケーションが正常に始動したかどうか (アプリケーション始動モード) を追跡することもできます。モニターを使用してアプリケーションの始動を監視するのは、複雑なクラスター環境では特に有効です。

サービス IP ラベルおよび IP アドレス

サービス IP ラベルは、クライアント・ノードとサーバー・ノードの間で通信を確立するために使用されません。データベース・アプリケーションなどのサービスは、サービス IP ラベルを使用して確立された接続を使用して提供されます。

サービス IP ラベルは、リソースとしてリソース・グループに入れることができます。これにより、PowerHA SystemMirror は、1 つのノード内で、または IP アドレス・テークオーバーが構成されている場合は (障害発生時に別のノードにそのリソースを転送することにより) クラスター・ノード間で、そのリソースの正常性をモニターしたり、そのリソースの高可用性を保持したりできるようになります。

注: サービス IP ラベルをさまざまなタイプのリソース・グループのリソースとして構成する場合は、一定のサブネット要件が適用されます。

関連概念:

35 ページの『サービス IP ラベルおよびアドレス』

サービス IP ラベルは、サービス IP アドレスにマップするラベルです。これは、クライアント・ノードとサーバー・ノードの間で通信を確立するために使用されます。

関連情報:

PowerHA SystemMirror プランニング・ガイド

テープ・リソース

SCSI またはファイバー・チャネル・テープ・ドライブを非コンカレント・リソース・グループのクラスター・リソースとして構成し、そのデバイスをクラスター内の 2 つのノードに対して高可用性のデバイスにできます。

以下の PowerHA SystemMirror 機能を使用すれば、共用テープ・ドライブを簡単に管理できます。

- SMIT 構成ツールを使用したテープ・ドライブの構成
- テープ・ドライブが正しく構成されているかどうかの検査
- リソース・グループの開始および停止操作中のテープ・ドライブの自動管理
- ノード障害およびノード回復時のテープ・ドライブの再割り振り
- クラスター・シャットダウン時のテープ・ドライブの再割り振り制御
- クラスター・リソースの動的再構成時のテープ・ドライブの再割り振り制御

クラスター・リソース・グループ

PowerHA SystemMirror ソフトウェアでリソースの高可用性を実現するには、各リソースがリソース・グループに含まれていなければなりません。リソース・グループにより、関連のあるリソースを管理しやすい単一の論理エンティティに結合することができます。

サポートされるリソース・グループの最大数は 64 です。

参加ノード・リスト

参加ノード・リストは、特定のリソース・グループをホストできるノードのリストを定義します。

リソース・グループを構成するときにノード・リストを定義します。参加ノード・リストに含めるノードは、クラスター内のノードの一部でもすべてでもかまいません。

通常、このリストには、同じデータおよびディスクを共用するすべてのノードが含まれています。

デフォルト・ノード優先順位

デフォルト・ノード優先順位は、特定のリソース・グループのノード・リストのノードの位置によって判別されます。

ノード・リストの最初のノードは、最も高いノード優先順位を持っています。このノードは、リソース・グループのホーム・ノードとも呼ばれます。他のノードの前にリストされているノードは、そのノードより高いノード優先順位を持ちます。

(より低い優先順位のノードによって現在制御されている) リソース・グループに関して、より高い優先順位のノードがクラスターに結合または再統合されると、リソース・グループのフォールバック・ポリシーによって、そのノードはリソース・グループの制御を取得します。つまり、リソース・グループは、優先順位の低いノードから優先順位の高いノードへ移動します。

指定した時間に、リソース・グループは、参加ノード・リストによってデフォルト・ノード優先順位を指定することができます。ただし、ユーザーが選択するさまざまなリソース・グループ・ポリシーはデフォルト・ノード優先順位をオーバーライドし、実際のノード優先順位を「作成」します。そして、そのノード優先順位に従って特定のリソース・グループがクラスターに移動します。

動的ノード優先順位

動的ノード優先順位ポリシーを設定すると、「CPU 負荷最小 (lowest CPU load)」などの RSCT リソース変数を使用して、非コンカレント・リソース・グループのテークオーバー・ノードを選択できます。動的優先順位ポリシーを有効にすると、テークオーバー・ノード・リストの順序は、選択された RSCT リソース変数の基準で評価する、イベント発生時のクラスターの状態によって決定されます。グループごとにポリシーの設定を変えることも、複数のグループに対して同じポリシーを設定することもできます。

ホーム・ノード

ホーム・ノード (このリソース・グループに対して最も優先順位の高いノード) は、非コンカレント・リソース・グループの参加ノード・リストに示される最初のノードです。

ホーム・ノードは、通常はリソース・グループを所有するノードです。非コンカレント・リソース・グループには、ホーム・ノードがある場合とない場合があります。非コンカレント・リソース・グループにホーム・ノードがあるかないかは、リソース・グループの始動動作、フォールオーバー動作、およびフォールバック動作によって決まります。

クラスターのさまざまな変更によって、グループは常にホーム・ノード上で始動するとは限りません。リソース・グループのホーム・ノードと、リソース・グループを現在所有しているノードを区別することが重要です。

コンカレント・リソース・グループは複数のノードで所有されるため、「ホーム・ノード」という用語はコンカレント・リソース・グループには使用されません。

始動、フォールオーバー、およびフォールバック

PowerHA SystemMirror は、クラスターの状態が変化したときは、リソース・グループをノードから別のノードに移動させることにより、クラスター・リソースの可用性を保証します。

PowerHA SystemMirror は、クラスター始動時に特定ノードまたは複数のノード上のリソース・グループを活動化したり、クラスターの状態が変化した場合はリソース・グループを別のノードに移動することによって、リソース・グループを管理します。クラスターのライフ・サイクルには、PowerHA SystemMirror が特定リソース・グループを管理する方法に影響する以下のステージがあります。

クラスター始動時

ノードが稼働し、リソース・グループは、選択されたリソース・グループの始動ポリシーに従ってノード間に配布されます。

ノード障害

このノードでアクティブなリソース・グループは、他のノードにフォールオーバーされます。

ノード回復

ノードはクラスターに再統合され、ユーザーの選択するリソース・グループ・ポリシーに応じてリソース・グループが再度獲得されます。

リソースの障害および回復

リソース・グループは、リソースが使用可能になったとき、別のノードにフォールオーバーされ、再取得されます。

クラスター・シャットダウン

クラスターをシャットダウンする方法はいろいろあります。それらの方法の 1 つに、リソース・グループを別のノードに移動させる方法があります。

これらのクラスターの各段階で、PowerHA SystemMirror のリソース・グループの動作は、次の内容によって定義されます。

- どのノード (複数も可) が、クラスター始動時にリソース・グループをアクティブにするか
- クラスター始動時に、ノードに取得できるリソース・グループの数
- リソース・グループを所有していたノードが失敗したとき、および、PowerHA SystemMirror がリソース・グループを別のノードに移動する必要があるときに、どのノードがリソース・グループに渡されるか
- リソース・グループを、クラスターと結合したところのノードにフォールバックするか、それともそれを所有しているノード上にとどまらせるか

選択したリソース・グループ・ポリシーによって、どのクラスター・ノードがリソース・グループを本来制御するか、どのクラスター・ノードが元のノードが制御を放棄したときにリソース・グループの制御を引き継ぐかが決定されます。

これらのポリシーの組み合わせで、1 つまたは複数のノードがリソース・グループを制御する際の制御の度合いをいろいろ指定できます。

つまり、リソース・グループの所有権に PowerHA SystemMirror の重点が置かれていることから、さまざまなクラスター構成が可能であり、アプリケーションのニーズに合わせてクラスター環境をきわめて柔軟に定義できます。始動、フォールオーバー、およびフォールバック・ポリシーの組み合わせによって、以前のリリースで使用可能なすべての管理ポリシーを、「事前定義した」グループ・タイプの動作を変更するオプションのセットを指定しないで要約することができます。

リソース・グループの動作を定義するとき、リソース・グループがクラスター内で 1 つ以上のノードにより引き継がれることを覚えておいてください。

始動、フォールオーバー、およびフォールバックは、さまざまなクラスター・ステージでリソース・グループがどのように振る舞うかを表す特定の動作です。フォールオーバーとフォールバックとの違いに注意してください。これらの語は、さまざまなリソース・グループ・ポリシーの説明において頻繁に目にするようになります。

始動

始動とは、リソース・グループが現在置かれているノード (単一または複数) 上、あるいはリソース・グループのホーム・ノード上で、そのリソース・グループを活動化することです。リソース・グループの始動は、クラスターの始動時またはノード上のグループの初期取得時に起こります。

フォールオーバー

フォールオーバーとは、現在のノードに障害が発生した後に、現在リソース・グループを所有するノードから別のアクティブ・ノードにリソース・グループを移動することです。新しい所有者は再統合ノードでも結合ノードでもありません。

フォールオーバーは、非コンカレント・リソース・グループに対してのみ有効です。

フォールバック

フォールバックとは、リソース・グループが現在置かれているノード (このリソース・グループのホーム・ノードではない) から、クラスターに結合または再統合されるノードにリソース・グループを移動することです。

例えば、そのリソース・グループに関して他のノードよりも優先順位の高いノードがクラスターに結合または再統合されると、そのノードがリソース・グループの制御を取得します。つまり、リソース・グループは、優先順位の低いノードから優先順位の高いノードへフォールバックします。

フォールバック動作は、非コンカレント・リソース・グループに対してのみ定義できます。

リソース・グループのポリシーおよび属性

特定の始動、フォールオーバー、およびフォールバック・ポリシーを使用するリソース・グループを構成します。このセクションでは、グループの属性およびシナリオを説明し、どのリソース・グループがご使用のクラスター要件に合っているかを判別する手掛かりとなります。

リソース・グループ・ポリシーは、以下の方法で使用できます。

- オフピーク時にノードを再統合し、確実にオンラインに戻るように、リソース・グループを構成します。
- 特定のアプリケーションを含むリソース・グループが、特定のノードでの始動時に、参照され取得される唯一のグループであるように指定します。ノード配布ポリシーを指定しても同様に行うことができます。このことは、複数の非コンカレント・リソース・グループがノード上で取得できる場合に関連していますが、特定のリソース・グループは、可用性を維持することがより重要であるアプリケーションを所有しています。
- 始動、フォールオーバー、およびフォールバック時に、特定のリソース・グループが同じノード上で同時にオンライン状態を保つように指定したり、異なるノード上で別々にオンライン状態を保つように指定したりします。

リソース・グループの始動、フォールオーバー、およびフォールバック

個々のリソース・グループには、ポリシーがいくつかあります。

これらのポリシーは以下のとおりです。

表 4. リソース・グループ・ポリシー

リソース・グループ	ポリシー
始動	<ul style="list-style-type: none"> • Online on Home Node Only (ホーム・ノードでのみオンライン)。リソース・グループが、リソース・グループ始動時にそのホーム・ノード (最高優先順位) でのみオンラインになります。このためには、最高優先順位のノードが使用可能である必要があります (リソース・グループのノード・リスト内の最初のノード)。 • Online on First Available Node (最初の使用可能ノードでオンライン)。リソース・グループは、使用可能になった最初の参加ノードでオンラインになります。 • Online on All Available Nodes (使用可能なすべてのノードでオンライン)。リソース・グループは、すべてのノードでオンラインになります。 • Online Using Distribution Policy (配布ポリシーを使用してオンライン)。それぞれのノードで 1 つのリソース・グループのみがオンラインになります。
フォールオーバー	<ul style="list-style-type: none"> • Fallover to Next Priority Node in the List (リスト内の次の優先順位ノードへフォールオーバー)。リソース・グループは、リソース・グループのノード・リストに指定されたデフォルトのノード優先順位に従います。 • Fallover Using Dynamic Node Priority (動的ノード優先順位を使用したフォールオーバー)。このオプションを選択する前に、3 つの事前定義動的ノード優先順位ポリシーのいずれかを選択します。これらは、使用可能なメモリーが最も多いノードなど、RSCT 変数を基準にしています。サイトが構成済みの場合、このオプションのフォールオーバーはサポートされません。 • Bring Offline (on Error Node Only) (オフラインにする (エラー・ノードのみ))。このオプションを選択すると、エラー状況時にノード上でリソース・グループをオフラインにします。
フォールバック	<ul style="list-style-type: none"> • Fallback to Higher Priority Node in the List (リスト中のより高い優先順位のノードにフォールバック)。より高い優先順位ノードがクラスターと結合するとき、リソース・グループがフォールバックします。このオプションを使用する場合は、遅延フォールバック・タイマーを使用します。遅延フォールバック・ポリシーが構成されていない場合は、より高い優先順位のノードがクラスターに結合されると即時にそのリソース・グループがフォールバックされます。 • Never Fallback (フォールバックしない)。リソース・グループがクラスターに加わる場合、リソース・グループは優先順位が高いノードにはフォールバックしません。

関連情報:

Administering PowerHA SystemMirror

PowerHA SystemMirror プランニング・ガイド

整定時間、動的ノード優先順位、およびフォールバック・タイマー

リソース・グループでは、起動時、フォールオーバー時、およびフォールバック時のリソース・グループの動作を指示する追加パラメーターを構成することができます。

リソース・グループの構成可能な追加パラメーターは、以下のとおりです。

整定時間

リソース・グループの始動動作は、現在オフラインのリソース・グループに整定時間を指定することによって構成できます。整定時間が構成されないと、そのリソース・グループは、クラスターに結合するノードの中で、最初に使用可能でより優先順位の高いノード上で始動します。整定時間が構成されると、PowerHA SystemMirror はその整定時間の間待機し、より優先順位の高いノードがクラスターに結合されるのを待ってから、リソース・グループを活動化します。整定時間を指定すると、複数のノードが同時に結合するときに、リソース・グループをより優先順位の高いノード上で取得できます。整定時間は、クラスター全体の属性です。構成された場合、「Online on First Available Node (最初の使用可能ノードでオンライン)」始動動作を選択したクラスター内のすべてのリソース・グループの始動動作に影響を与えます。

配布ポリシー

ノード・ベースの配布ポリシーを使用するリソース・グループの始動動作を構成することができます。このポリシーにより、ノードは始動時にリソース・グループを必ず 1 つだけ獲得します。詳しくは次のセクションを参照してください。

動的ノード優先順位

リソース・グループのフォールオーバー動作は、3 つの動的ノード優先順位ポリシーのいずれかを使用して構成できます。これらのポリシーは、最大の空きメモリーまたは最小の CPU 使用量といった RSCT 変数を基準にしています。リソース・グループを回復するために、PowerHA SystemMirror はフォールオーバー時のポリシーに最も適したノードを選択します。

注: サイトが構成済みの場合、「動的ノード優先順位」オプションを使用するフォールオーバーはサポートされません。

遅延フォールバック・タイマー

リソース・グループのフォールバック動作は、遅延フォールバック・タイマーの指定と割り当てにより、定義済みの頻度 (日次、週次、月次、および毎年) のいずれか、または指定日時に発生するように構成できます。これは、例えば、オフピークの業務時間にリソース・グループのフォールバックが起きようスケジュールを立てるのに有効です。

配布ポリシー

クラスターの始動時に、リソース・グループに対してノード・ベースの配布ポリシーを使用できます。

複数のリソース・グループに対してこのポリシーを選択すると、PowerHA SystemMirror は、始動時に各ノードがこれらのリソース・グループから 1 つのみを取得するようにします。これによって、CPU 集中アプリケーションをさまざまなノードに分散できます。

関連情報:

Administering PowerHA SystemMirror

PowerHA SystemMirror プランニング・ガイド

ネットワークおよびリソース・グループ

すべてのリソース・グループで、エイリアスを使用するネットワーク上で構成されたサービス IP ラベルがサポートされます。

サービス IP ラベルは、任意の非コンカレント・リソース・グループに組み込むことができます。そのリソース・グループでは、「Online on All Available Nodes (使用可能なすべてのノードでオンライン)」を除き、許可されている任意の始動ポリシーを使用できます。

リソース・グループ依存関係

PowerHA SystemMirror では、リソース・グループの順序付けおよびリソースのカスタマイズされたシリアル処理をサポートし、異なる複数のリソース・グループに常駐するアプリケーション間に依存関係が存在するクラスター構成にも対応しています。

シリアル処理をカスタマイズすると、特定のリソース・グループを別のリソース・グループより前に処理するように指定できます。PowerHA SystemMirror では、クラスター・イベント発生時に適切な処理が確実に行われるようにするために、リソース・グループ (およびそこに属するアプリケーション) 間で親/子依存関係を簡単に構成する方法を提供しています。

ロケーション依存関係ポリシーを使用すれば、クラスター始動時だけでなく、フォールオーバーおよびフォールバック時にも、期待どおりの方法で配布されるようにリソース・グループを構成できます。また、指定

したグループが別々のノード上でも、同じノード上でもオンライン状態になるように、依存関係を構成できます。PowerHA SystemMirror は、可能な場合は並列処理を、また必要に応じてシリアル処理を使用し、依存関係のあるリソース・グループを適切な順序で処理します。処理をカスタマイズする必要はありません。リソース・グループ間に異なるタイプの依存関係を構成できます。

- 親/子依存関係
- ロケーション依存関係

リソース・グループ間に構成する依存関係には、以下のものがあります。

- SMIT インターフェースを使用して明示的に指定されます。
- ローカル・ノードだけでなくクラスター全体でも設定されます。
- クラスターでの受け入れを保証します。

リソース・グループの親/子依存関係

リソース・グループに親/子依存関係を構成すると、クラスターの構成と制御が容易になります。1 つのアプリケーションが別のアプリケーションの正常な始動に依存し、両方のアプリケーションが PowerHA SystemMirror によって高可用性を保つ必要があるような多層アプリケーションを持つクラスターを制御することもできます。

以下の例は、親/子依存関係の動作を示しています。

- リソース・グループ A が、リソース・グループ B に依存している場合、ノードの始動時に、リソース・グループ A がクラスター内のノードで取得される前に、リソース・グループ B はオンラインになる必要があります。フォールオーバー時、この順序は逆転します。リソース・グループ A がオフラインになってから、リソース・グループ B がオフラインになる必要があります。
- さらに、リソース・グループ A が、リソース・グループ B に依存している場合、ノードの始動時または再統合時に、リソース・グループ B がオンラインになってからリソース・グループ A は、オンラインになります。リソース・グループ B をオフラインにすると、リソース・グループ A もオフラインになります。これは、リソース・グループ A が B に依存しているからです。

リソース・グループ間の依存関係は、多層のアプリケーションでクラスターを構築する、予測可能で信頼性のある方法を提供します。依存リソース・グループを使用できる標準クラスター環境について詳しくは、『多層アプリケーションを伴うクラスター構成』を参照してください。以下の用語は、リソース・グループ間の親/子依存関係について説明しています。

- 親リソース・グループ は、依存するリソース・グループ (子) が開始できる前に、オンライン状態にする必要があります。
- 子リソース・グループ は、親リソース・グループに依存します。子リソース・グループは、親リソース・グループがアクティブにされた後でのみ、クラスターの任意のノードでアクティブになります。通常、子リソース・グループは、親リソース・グループが提供するアプリケーション・サービスに依存します。

リソース・グループのリリース時 (例えば、フォールオーバー中やクラスター・サービスの停止中)、PowerHA SystemMirror は、親リソース・グループがオフラインになる前に子リソース・グループをオフラインにします。次の図は、リソース・グループ間の親/子依存関係を示しています。

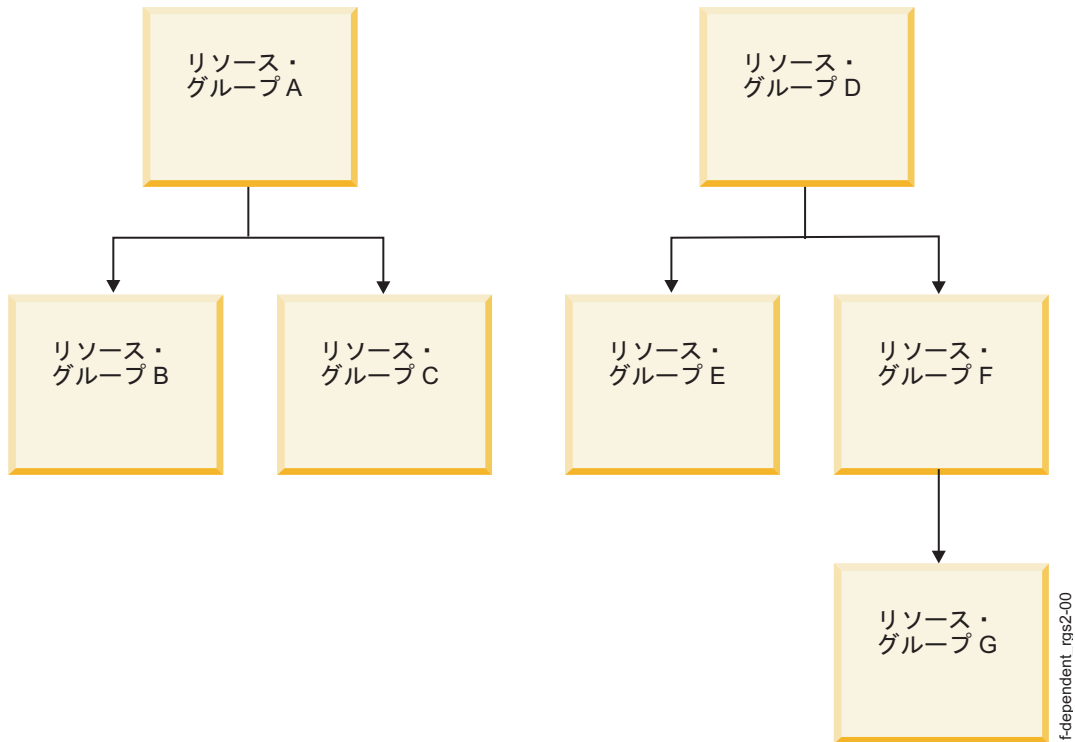


図 7. リソース・グループ間の 2 レベルおよび 3 レベルの依存関係の例

この例では、次のガイドラインと制限の下で構成された関係を示しています。

- 親リソース・グループが、子 (依存) リソース・グループがノード上で活動化する前に、クラスター内のノードでオンラインになる必要がある依存関係のタイプを構成できます。
- リソース・グループは、指定された依存関係の目的に応じて、親および子リソース・グループとして機能することができます。
- リソース・グループの依存関係には、3 つのレベルを指定することができます。
- リソース・グループ間の循環依存関係を指定することはできません。

これらのガイドラインと制限は、以下のように、リソース・グループ間の親/子依存関係にも適用されません。

- クラスター・サービスが実行している間、リソース・グループ間の依存関係を追加、変更、または削除できます。
- 2 つのリソース・グループ間の依存関係を削除するとき、これらのリソース・グループ間のリンクのみが PowerHA SystemMirror 構成データベースから除去されます。リソース・グループは削除されません。
- 親リソース・グループのフォールオーバー中に、アプリケーションを含む子リソース・グループは、一時的にオフラインになり、使用可能なノードではオンラインになります。また、子リソース・グループに属するアプリケーションは、停止およびリスタートされます。

関連概念:

97 ページの『多層アプリケーションを使用したクラスター構成』

親/子依存関係のリソース・グループを利用できる典型的なクラスター構成は、WebSphere® のようなアプリケーションが DB2® のような別のアプリケーションに依存している環境です。

リソース・グループ・ロケーション依存関係

個々のリソース・グループおよび親/子依存関係に対するさまざまなポリシーに加えて、PowerHA SystemMirror では、リソース・グループの相互依存関係全体を取り扱うポリシーを提供します。PowerHA SystemMirror はこれらの関係を識別し、適切な順序でリソース・グループを処理します。

リソース・グループを以下のように構成できます。

- 指定した 2 つ以上のリソース・グループが、常に同じノード上でオンラインになります。同じノードで、始動、フォールオーバー、およびフォールバックされます。
- 指定した 2 つ以上のリソース・グループが、常に異なるノード上でオンラインになります。異なるノードで、始動、フォールオーバー、およびフォールバックされます。リソース・グループに優先順位を割り当て、フォールオーバーおよびフォールバックの発生時に最も重要なものが最初に処理されるようにします。
- 指定した 2 つ以上のリソース・グループ (複製されるリソースがある) が、常に同じサイトでオンラインになります。

指定されたロケーション依存関係を使用して個々のリソース・グループが構成されると、これらのグループは、クラスター・マネージャーによって 1 つの単位として処理されるセットを形成します。リソース・グループを **clRGmove** コマンドを使用して明示的に移動する場合、次のルールが適用されます。

- リソース・グループが同じノード上でオンラインの依存関係のセットに加わると、同じノード・セットに属する他のすべてのリソース・グループが現在オンラインになっているノード上でのみオンラインになります。これは、クラスター・マネージャーの規則と同じです。
- リソース・グループが同じサイト上でオンラインの依存関係のセットに加わると、同じサイト・セットに属する他のリソース・グループが現在オンラインになっているサイト上でのみオンラインになります。これは、クラスター・マネージャーの規則と同じです。
- リソース・グループが異なるノード上でオンラインの依存関係のセットに加わると、異なるノードの依存関係セット内の任意の他のリソース・グループをホストしないノード上でのみオンラインになります。(これは、クラスター・マネージャーの規則と同じです。) ただし、このセットに属するリソース・グループを移動すると、優先順位の値は等しいものとして扱われます。一方、PowerHA SystemMirror がこれらのグループをオンラインにする場合には、優先順位が考慮されます。

サンプル・ロケーション依存関係モデル

架空の会社、XYZ 出版社では、Web コンテンツの開発に使用するさまざまなプラットフォームに優先順位を付けるビジネス継続性モデルを採用しています。ロケーション・ポリシーは、一部のリソース・グループを厳密に別々のノード上に配置し、一部のリソース・グループを同じノード上にまとめて配置しておくために使用されます。

次の図は、3 つの異なるアプリケーションに使用されるノードとそれぞれに関連するデータベースを持つ、ロケーション依存関係モデルの例を示しています。各アプリケーションのデータベースは、常にアプリケーションと同じノード上で実行されなければなりません。例えば、実動データベースは実動アプリケーションを含んでいなければなりません。

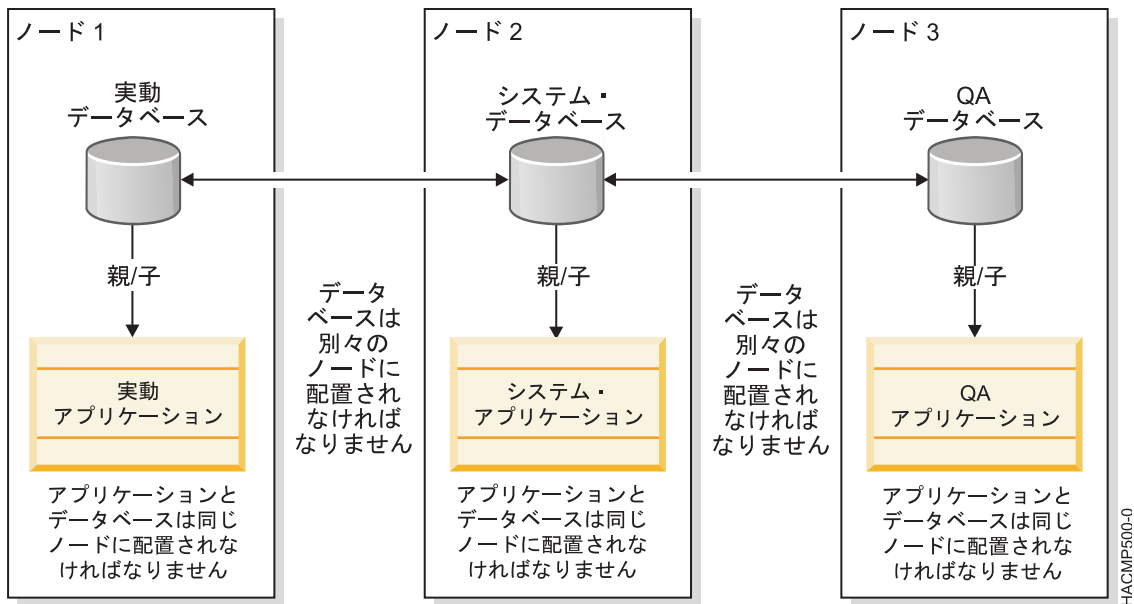


図 8. 別々のデータベースに使用されているノード 1、2、および 3 の例

関連情報:

Administering PowerHA SystemMirror

PowerHA SystemMirror プランニング・ガイド

サイトおよびリソース・グループ

大部分の PowerHA SystemMirror 構成ではサイトを指定せず、デフォルトの IGNORE 管理ポリシーを使用します。災害時回復のため PowerHA SystemMirror Enterprise Edition コンポーネントをインストールした場合は、クラスター・ノードを地理的に離れたサイトに分散させ、サイト間管理ポリシーのいずれかを選択します。

リソース・グループ内で複製したいリソースを指定します。複製されたリソース・グループの 1 次インスタンスに対して始動、フォールオーバー、およびフォールバックのポリシーを定義します。1 次インスタンスは、リソース・グループがオンラインになっている場所です。リソース・グループの 1 次インスタンスのあるノードでは、すべてのグループのリソースをアクティブにします。2 次インスタンス (複製) は、バックアップと異なるサイト上にあるノードでアクティブにされます。リソース・グループの始動、フォールオーバー、およびフォールバックのポリシーを組み合わせたサイト間管理ポリシーによって、1 次インスタンスを最初に検索するサイト、およびサイト間のフォールオーバーおよびフォールバックを処理する方法が決まります。

リソース・グループのサイト内管理ポリシーを構成するために、以下のオプションがあります。

- 1 次サイトを優先
- 一方のサイトでオンライン
- 両方のサイトでオンライン

クラスターにサイトを定義する場合、複製する各リソース・グループに対して始動、フォールオーバー、およびフォールバックのポリシーを定義する際に、リソース・グループを 1 次サイトのノードに割り当て、

次に別の (2 次) サイトのノードに割り当てることができます。リソース・グループの 1 次インスタンスが 1 次サイト上で稼働し、2 次インスタンスが 2 次サイト上で稼働します。

コンカレント・リソース・グループがある場合は、すべてのノード上で稼働するよう定義します。この場合、「両方のサイトでオンライン」のサイト間管理ポリシーを選択できます。これで、両方のサイトのインスタンスがアクティブになります (この場合 2 次インスタンスはありません)。他のサイト間管理ポリシーを選択して、1 つのサイトのすべてのノード上でコンカレント・リソース・グループをオンラインにし、別のサイトにバックアップ・インスタンスを置くこともできます。

注: 定義済みのサイトを使用しており、非コンカレントのリソース・グループがある場合、1 つのサイトで `unmanage` オプションを使用してクラスター・サービスを停止すると、このリソース・グループはそのサイトのすべてのノードで UNMANAGED 状態になります。

また、`clRGmove` ユーティリティを使用すれば、サイトの境界を越えてリソース・グループの 1 次インスタンスを移動することもできます。PowerHA SystemMirror は、必要に応じて、ピア 2 次インスタンスを再配布します (または、構成要件によって移動が許可されない場合に、警告を出します)。

リソース・グループの 2 次インスタンスがオフラインの間にそのリソース・グループの 1 次インスタンスを移動すると、2 次インスタンスはオフラインのままです。PowerHA SystemMirror が、そのリソース・グループの 2 次インスタンスを自動的にオンラインにすることはありません。

関連情報:

管理ガイド

計画ガイド

ログ・アナライザー

ログ・アナライザー機能は、PowerHA SystemMirror、AIX、およびその他の各システム・コンポーネントのログ・ファイルから、ディスク障害やインターフェース障害などのさまざまなタイプのエラーに関する詳細情報をスキャンしたり抽出したりできるようにします。

ログ・アナライザー機能は、次のタスクを実行します。

- ログ・ファイルの分析と、エラー・ストリングまたはタイム・スタンプに基づくエラー報告書の提供。
- AIX エラー・ログからのコア・ダンプ・ファイルの分析。
- `snap` ユーティリティおよび `clsnap` ユーティリティを使用して収集されたログ・ファイルの分析。
- 提供されたエラー・ストリングに基づくユーザー指定 `snap` ファイルの分析および報告書の生成。

ログ・アナライザー機能は、問題判別のプロセス段階で、エラー情報を手動で見つけてアSEMBルする必要なしに、迅速かつ効率的に続行するのに役立ちます。

PowerHA SystemMirror でサポートされるハードウェア

さまざまなタイプの IBM ハードウェアを PowerHA SystemMirror と組み合わせて使用すれば、基本レベルの高可用性環境を実現できます。

PowerHA SystemMirror に対してサポートされているハードウェアの要約については、「PowerHA for AIX Version Compatibility Matrix」(<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD101347>) を参照してください。

関連情報:

PowerHA SystemMirror インストール・ガイド

PowerHA SystemMirror クラスタ・ソフトウェア

このセクションでは、高可用性の環境を実現する PowerHA SystemMirror ソフトウェアについて説明します。

PowerHA SystemMirror ノードのソフトウェア・コンポーネント

PowerHA SystemMirror ノードのソフトウェア・コンポーネントには多くの層があります。

次の図は、PowerHA SystemMirror クラスタ内のノード上のソフトウェアの各層を示しています。



図 9. PowerHA SystemMirror クラスタ・ノードのモデル

各層の説明リストは次のとおりです。

アプリケーション層

PowerHA SystemMirror for AIX によって提供されるサービスを介して高い可用性を維持する任意のアプリケーションです。

PowerHA SystemMirror for AIX 層

クラスター内の変化を認識し、AIX 機能の使用を調整することによって、クリティカルなデータおよびアプリケーションに可用性の高い環境を作り出すソフトウェアです。PowerHA SystemMirror は、より下位にある層を補完します。すなわち、アプリケーションが関連の通信を使用できるようにするより多数のサービスを提供すること、およびデータ・ストレージ・サービスの可用性を高めることによって補完します。

RSCT 層

IBM Reliable Scalable Cluster Technology (RSCT) サービスは、AIX にパッケージされています。RSCT 層は、分散環境またはクラスター環境において、高信頼性メッセージングを使用することにより、クラスター化アクティビティの同期化と調整を行う機能を提供します。RSCT には、Resource Monitoring and Control (RMC)、構成リソース・マネージャー、およびグループ・サービス・コンポーネントがあります。

Cluster Aware AIX 層

AIX オペレーティング・システムのこの拡張機能では、ノード・メンバーシップとネットワーク・インターフェースの正常性、およびイベント通知をモニターする機能を提供します。

AIX 層

PowerHA SystemMirror クラスターを根本でサポートするソフトウェアです。

論理ボリューム・マネージャー (LVM) サブシステム層

論理レベルでデータ・ストレージを管理します。

TCP/IP サブシステム層

PowerHA SystemMirror クラスターに通信サポートを提供します。

関連概念:

59 ページの『IBM Reliable Scalable Cluster Technology の可用性サービス』

IBM Reliable Scalable Cluster Technology (RSCT) のハイ・アベイラビリティ・サービスは、スケーラビリティを向上させ、分散した各サブシステムにソフトウェア障害を通知し、ソフトウェア・スタックのすべてのサブシステム内の回復および同期を調整します。RSCT は、ノードおよびネットワークの障害を PowerHA クラスター・マネージャーに報告します。

クラスター・マネージャー

クラスターの状態の変化はクラスター・イベントと呼ばれます。各ノードで、クラスター・マネージャーは、ローカル・ハードウェアおよびソフトウェア・サブシステムのイベント (アプリケーション障害イベントなど) をモニターします。

このようなイベントに応じて、クラスター・マネージャーは 1 つ以上のイベント・スクリプト (アプリケーション再始動スクリプトなど) を実行します。すべてのノードのクラスター・マネージャーは、イベントに応じて必要となるアクションを調整するため、メッセージを交換します。

クラスター・マネージャーは、各クラスター・ノードで実行されるデーモンです。クラスター・マネージャーの主なタスクは、予期しないイベント (ソフトウェアおよびハードウェア障害からの回復など) やユーザーが開始したイベント (結合ノード・イベントなど) に対応することです。RSCT サブシステムは、ノードおよびネットワーク関連イベントについてクラスター・マネージャーに通知します。

他の PowerHA SystemMirror デーモンとクラスター・マネージャーの接続

クラスター・マネージャーは、ノードやインターフェースのクラスター状態の変化に関連する情報を収集します。クラスター情報プログラム (Clinfo) は、この情報をクラスター・マネージャーから取得し、Clinfo と通信するクライアントにクラスターの状態の変化を通知します。このクラスター状態情報は、PowerHA

SystemMirror Management Information Base (MIB) に保管されます。

関連概念:

105 ページの『PowerHA SystemMirror 構成プロセスおよび機能』
以下のトピックでは、PowerHA SystemMirror クラスター構成プロセス、および PowerHA SystemMirror ソフトウェアに付属の管理ツールについて概説します。

59 ページの『IBM Reliable Scalable Cluster Technology の可用性サービス』
IBM Reliable Scalable Cluster Technology (RSCT) のハイ・アベイラビリティ・サービスは、スケラビリティを向上させ、分散した各サブシステムにソフトウェア障害を通知し、ソフトウェア・スタックのすべてのサブシステム内の回復および同期を調整します。RSCT は、ノードおよびネットワークの障害を PowerHA クラスター・マネージャーに報告します。

PowerHA SystemMirror ソフトウェア・コンポーネント

PowerHA SystemMirror ソフトウェアには多くのコンポーネントがあります。

クラスター・セキュア通信サブシステム

PowerHA SystemMirror では共通の通信インフラストラクチャーによってシステム間通信のセキュリティが強化されました。クラスター・ユーティリティは、ノード間の通信において、各ノードで稼働するクラスター通信デーモンを使用します。

共通の通信パスは 1 つのみであるため、すべての通信を確実に保護することができます。ほとんどのコンポーネントは、クラスター通信デーモンを介して通信しますが、次のコンポーネントは、ノード間通信で別のメカニズムを使用します。

コンポーネント	通信メソッド
クラスター・マネージャー	RSCT
ハートビート・メッセージング	Cluster Aware AIX
クラスター情報プログラム (Clinfo)	SNMP

PowerHA SystemMirror では、さらに高いレベルのセキュリティが必要なユーザー向けに、クラスター・ノード間で送信されるメッセージのメッセージ認証およびメッセージ暗号化を提供します。

接続の認証

標準セキュリティ・モードは、アクセス・リストのソース IP アドレスをチェックし、ソース・ポートの値が 571 から 1023 であるかをチェックし、リモート・コマンド実行の最小特権の原理を使用します。標準のセキュリティは、デフォルトのセキュリティ・モードです。追加のセキュリティとしては、PowerHA SystemMirror 内部ノード通信でノード間の接続に VPN をセットアップすることができます。

メッセージの認証および暗号化

メッセージの認証およびメッセージの暗号化によって、クラスター・ノード間で送信される PowerHA SystemMirror メッセージの追加セキュリティが提供されます。メッセージの認証は、メッセージの作成元および保全性を確実にします。メッセージの暗号化は、データの外観を送信時に変更し、メッセージを認証するノードに受信されたときに、元の形式に変換されます。セキュリティ・オプションおよび SMIT インターフェースを使用した分散暗号化鍵のオプションを構成することができます。

IBM Reliable Scalable Cluster Technology の可用性サービス

IBM Reliable Scalable Cluster Technology (RSCT) のハイ・アベイラビリティ・サービスは、スケーラビリティを向上させ、分散した各サブシステムにソフトウェア障害を通知し、ソフトウェア・スタックのすべてのサブシステム内の回復および同期を調整します。RSCT は、ノードおよびネットワークの障害を PowerHA クラスタ・マネージャーに報告します。

PowerHA SystemMirror および RSCT ソフトウェア・スタックは、それぞれのクラスタ・ノード上で動作します。PowerHA SystemMirror クラスタ・マネージャーは、起こりうる障害の指示を以下の送信元から取得します。

- ネットワーク・デバイスの状態をモニターする RSCT
- ボリューム・グループおよびディスクの状態をモニターする AIX LVM
- アプリケーションの状態をモニターするアプリケーション

PowerHA SystemMirror クラスタ・マネージャーは、コンポーネントに障害が発生すると、クラスタ回復を実行します。各ノードで実行中の RSCT は、そのピアとハートビートを交換し、クラスタ内の他のノードの可用性をモニターします。ハートビートが停止した場合、クラスタ内の他のシステムが回復処理を開始します。クラスタ内の他のシステムは、クリティカル・アプリケーションを起動し、そのデータが破壊または失われていないことを確認するために必要な措置を実行します。RSCT サービスには、次のコンポーネントがあります。

- Resource Monitoring and Control (以前のバージョンの PowerHA SystemMirror では、イベント管理サブシステムを使用)ハイ・アベイラビリティ・サービス・セットを提供する分散サブシステム。これは、システム・リソースの状態に関する情報をクライアント・プログラムに関係のあるリソース状態に関する情報と突き合わせることによって、イベントを作成します。これにより、クライアント・プログラムは、イベント通知を使用してシステム障害からの回復を開始できるようになります。
- グループ・サービス。一連のノードで実行中のアプリケーションの状態変化を調整およびモニターするシステム規模のハイ・アベイラビリティ機能です。グループ・サービスは、高可用性アプリケーションのデザインや実装、および複数のアプリケーションの一貫した回復に役立ちます。これは、統合されたフレームワークの中でこの 2 つのタスクを実施します。

次の図は、PowerHA SystemMirror アーキテクチャーを構成する主なコンポーネントを示したものです。

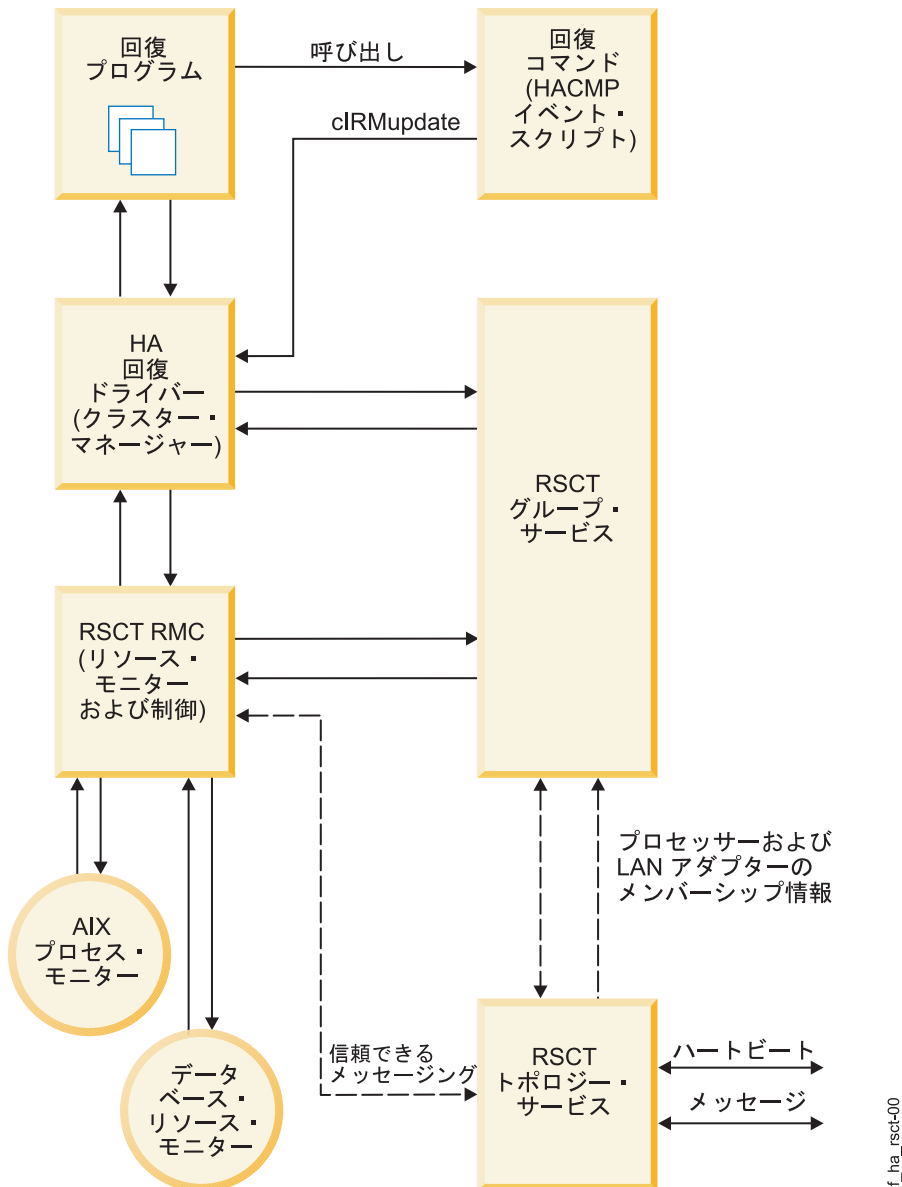


図 10. PowerHA SystemMirror アーキテクチャーのコンポーネント

クラスター・マネージャーおよび SNMP モニター・プログラム

PowerHA SystemMirror クラスターは動的であり、時間の経過と共に各種の推移が可能です。

例えば、ノードがクラスターと結合または分離したり、物理ネットワーク・インターフェース・カード上のサービス IP ラベルが別の IP ラベルに置き換えられることがあります。特に、可用性の高いクライアントやアプリケーションがクラスター・ノードの提供するサービスを使用する必要がある場合には、こうした変化の 1 つ 1 つがクラスターの構成に影響を与えます。

SNMP サポート

クラスター・マネージャーは、Simple Network Management Protocol (SNMP) サポートをクライアント・アプリケーションに提供します。SNMP は、TCP/IP ベースのネットワークのモニターと管理のため

の業界標準仕様です。SNMP には、プロトコル、データベース仕様書、およびデータ・オブジェクトの集合が含まれています。このデータ・オブジェクトの集合は Management Information Base (MIB) を形成します。SNMP が提供する標準 MIB には、IP アドレス、アクティブ TCP 接続の数などの情報が含まれています。標準の SNMP エージェントは **snmpd** デーモンです。

SNMP を拡張し、離散的環境またはアプリケーションに関する情報が入ったエンタープライズ固有の MIB を組み込むことができます。PowerHA SystemMirror では、クラスター・マネージャーがその MIB に定義されているオブジェクトの情報を保持し、この情報を専用ネットワーク・モニターまたはネットワーク管理ステーションに渡します。

PowerHA SystemMirror MIB

クラスター・マネージャーは、特別な PowerHA SystemMirror MIB にクラスターの状態の情報を保持します。クラスター・マネージャーがクラスター・ノードで始動されると、SNMP デーモン **snmpd** に登録され、継続的にクラスター情報を収集します。クラスター・マネージャーは、イベントとその結果のクラスターの状態を追跡し、PowerHA SystemMirror MIB 内のクラスターのトポロジを最新の状態に保ちます。

関連情報:

クライアント・アプリケーションのプログラミング

クラスター情報プログラム

クラスター情報プログラム (Clinfo) の **clinfo** デーモンは、SNMP ベースのモニターです。

クライアント・マシンまたはクラスター・ノード上で実行中の Clinfo は、更新されたクラスター情報を MIB に照会します。Clinfo を使用して、PowerHA SystemMirror クラスター、ノード、およびネットワークの状態に関する情報を、クライアントおよびアプリケーションが利用できるようにすることができます。クライアントは、ナイーブおよびインテリジェントという 2 つのカテゴリーに分けることができます。

- ナイーブ・クライアントは、クラスター複合体を 1 つのエンティティと見なします。クラスター・ノードに障害が発生したときに、IP アドレス・テークオーバー (IPAT) が使用可能になっていない場合、クライアントを再始動 (または少なくともそのノードに再接続) する必要があります。
- これに対してインテリジェント・クライアントはクラスター対応です。つまり、ノード障害に適切に対処し、代替ノードに接続します。また、その障害をユーザーから隠すこともあります。このようなインテリジェント・クライアントはクラスターの状態についての知識が必要です。

PowerHA SystemMirror ソフトウェアは、可用性の高いサーバー、データ、およびアプリケーションの利点をクライアントに拡張します。これはクラスター・マネージャーおよび Clinfo API 機能を使用してクライアントにクラスター状態の変化を通知することによって実現されます。

クラスターの変化への応答

Clinfo は、クラスター・イベント、ネットワーク・イベント、またはノード・イベントが発生すると必ず **/usr/es/sbin/cluster/etc/clinfo.rc** スクリプトを呼び出します。デフォルトでは、**clinfo.rc** スクリプトはシステムの ARP キャッシュをフラッシュしてネットワーク IP アドレスの変更を反映します。また、他のアドレスが ping 要求に応答するまでキャッシュを更新しません。PowerHA SystemMirror ハードウェア・アドレス・スワップ機能が使用可能になっている場合は、ハードウェア・アドレス・スワップ機能によって IP アドレスとハードウェア・アドレスの間の関係が維持されるため、一般的には ARP キャッシュをフラッシュする必要はありません。

交換イーサネット・ネットワークでは、新しい MAC アドレスがスイッチと通信できるように ARP キャッシュをフラッシュする必要があります。

さらにアクションが必要な場合は、`clinfo.rc` スクリプトにロジックを追加できます。

Clinfo API

Clinfo API は、クラスター状況情報にアクセスするために C と C++ の両言語のインターフェースをアプリケーション開発者に提供します。PowerHA SystemMirror ソフトウェアには、2 つのバージョンの Clinfo API が含まれています。1 つは単一スレッド・アプリケーション用です。もう 1 つはマルチスレッド・アプリケーション用です。Clinfo とそれに関連した API を使用すると、開発者は、クラスター内の変化を認識しその変化に応じたアプリケーションを作成できるようになります。

関連概念:

18 ページの『PowerHA SystemMirror クラスター・ノード、ネットワーク、およびハートビートの概念』

このセクションでは、文書全体にわたって使用され、PowerHA SystemMirror ユーザー・インターフェースの中でも使用される、主要なクラスター・トポロジに関する概念および定義について説明します。

65 ページの『アプリケーションの可用性の確保』

このセクションでは、PowerHA SystemMirror ソフトウェアがクラスター・コンポーネントの可用性を確保することで、どのようにアプリケーションの可用性を確保するかについて説明します。PowerHA SystemMirror は、すべてのキー・システム・コンポーネントについて単一障害点を除去し、ほとんどのルーチン・クラスター保守タスクについてスケジュールされたダウン時間の必要を除去します。

関連情報:

クライアント・アプリケーションのプログラミング

高可用性の NFS サーバー

PowerHA SystemMirror 製品サブシステムには、高可用性の NFS サーバー機能が組み込まれています。

高可用性 NFS サーバーによって、1 次 NFS サーバーに障害が発生した場合でも、バックアップ・プロセッサが現行の NFS 活動を回復することができます。この特別な NFS サーバー機能には、ネットワーク・ファイルシステム (NFS) 上での可用性の高い変更およびロック機能が含まれます。以下の操作が可能です。

- ロックおよび重複キャッシュを保持する高信頼性 NFS サーバー機能の使用 (NFS バージョン 2 およびバージョン 3 を使用する場合は 2 ノード・クラスターのみ)
- NFS クロスマウントを行うネットワークの指定
- NFS エクスポートおよびクロスマウントのディレクトリー・レベルでの定義
- NFS をエクスポートするディレクトリーおよびファイルシステムに関するエクスポート・オプションの指定
- NFS を使用する 2 つのノードの構成

PowerHA SystemMirror クラスターには、最大 16 のノードが含められます。NFS バージョン 2 およびバージョン 3 を使用するクラスターは、最大 2 ノードまでしか構成できないのに対し、NFS バージョン 4 を使用するクラスターは最大 16 ノードまで構成できます。

共用外部ディスク・アクセス

PowerHA SystemMirror ソフトウェアは、2 つの方式の共用外部ディスク・アクセス (非コンカレントとコンカレント) をサポートします。

非コンカレント共用外部ディスク・アクセス:

非コンカレント環境では、一時点では 1 つのノードだけが共用外部ディスクにアクセスできます。

このノードに障害が発生した場合は、ピア・ノードの 1 つがそのディスクを獲得し、リソースとして定義されているファイルシステムをマウントし、アプリケーションを再始動して、クリティカル・サービスを復元します。ファイルシステムの数とサイズにより異なりますが、一般には 30 から 300 秒かかります。

非コンカレント構成では以下のディスクおよびディスク・アレイを使用できます。

- SCSI ディスク
- SCSI ディスク・アレイ
- シリアル・ディスク
- ファイバー・チャンネルと直接接続のディスク
- ファイバー・チャンネルと SAN 接続のディスク

障害が発生したディスクが単一障害点となるのを防ぐために、AIX LVM 機能を使用して共用ボリューム・グループ内の各論理ボリュームをミラーリングする必要があります。IBM エンタープライズ・ストレージ・システムまたはサポートされるその他の RAID アレイを使用する場合は、LVM ミラーリングを使用しないでください。RAID デバイスによって独自のデータ冗長性が提供されます。

単一マシン・モードで実行できるソフトウェアのほとんどは、修正することなく PowerHA SystemMirror ソフトウェアによって管理できます。

非コンカレント・アクセスでは一般に、可用性の高いソリューションを提供するために、サーバー・プログラム (例えば、データベース管理システム) またはアプリケーションにコードの変更を加える必要はありません。エンド・ユーザーにとって、ノード障害は非常に速いマシン・リブートのように見えます。クラスターに残ったノードの 1 つが障害ノードのリソース・グループの所有権を取得して、高可用性のアプリケーションを再始動します。AIX のネイティブ・ファイルシステムであるジャーナル・ファイルシステムは、ファイルシステムの健全性を保証します。サーバー・プログラムはトランザクション・データの健全性を保証します。

ユーザーは、クラスターに残ったノードの 1 つにログオンしてアプリケーションを再始動するだけです。ログオンとアプリケーション再始動の手続きの実行を、PowerHA SystemMirror ソフトウェアに任せることもできます。PowerHA SystemMirror の構成によっては、ユーザーが何らかのアクションを取らずに済む場合もあります。その場合は、フォールオーバー中に遅延が発生するだけです。

コンカレント共用外部ディスク・アクセス:

コンカレント・アクセス機能は、PowerHA SystemMirror クラスターが提供する利点を拡張します。コンカレント・アクセスでは、複数の (最大 16 の) ノードに接続されたディスク・サブシステム上のボリューム・グループに同時アクセスすることが可能です。

コンカレント・アクセスを使用すれば、クラスターは、耐障害性に匹敵するほぼ連続したデータ可用性を、より低コストで実現できます。また、コンカレント・アクセスにより、パフォーマンスが向上し、アプリケーションの開発が容易になり、水平方向の拡張が可能になります。コンカレント・アクセスでは複数のノードからデータに同時アクセスするので、複数のノードが同じデータ・ブロックに対して相反する変更を行うことを防ぐための追加ツールが必要になる場合があります。PowerHA SystemMirror ソフトウェアは、アプリケーションをコンカレント・アクセス環境で実行するための準備をする Clinfo プログラムを提供します。Clinfo API は、アプリケーションが「クラスター対応」になるような API を提供します。Clinfo ツールについては、この章の前のほうで説明しています。

コンカレント共用外部ディスク・アクセスには、次のような利点があります。

- 透過的回復による可用性の向上。コンカレント・アクセスでは、フォールオーバーの時間が大幅に短縮され、場合によってはほんの数秒で済むこともあります。これは、ピア・システムがすでに共用ディスクに物理的にアクセスしていて、独自のアプリケーション・インスタンスを実行するためです。

コンカレント・アクセス環境では、フォールオーバーに伴って、不完了トランザクションは障害の発生したプロセッサから基本的にバックアウトされます。クラッシュによる部分トランザクションの回復は、復帰ノードで実行中のサーバー・ソフトウェアが引き受けます。すべてのノードがデータにコンカレント・アクセスするため、クライアント/サーバー・アプリケーションは、失敗した要求を、復帰ノードで即時に再試行し、復帰ノードで着信トランザクションの処理が続けられます。

- 複数プロセッサの利用によるスループットの向上。アプリケーションはもはや、単一のプロセッサのスループットに制約されません。アプリケーションの複数のインスタンスが複数のプロセッサ上で同時に実行できます。さらに大きな処理能力が必要になったときは、クラスターにシステムを追加してスループットを向上させることができます。
- 単一データベース・イメージによるアプリケーション開発および保守の容易化。非コンカレント環境では、アプリケーションおよびそのデータの区分化が、パフォーマンスを向上させる唯一の方法です。コードおよびデータを細分化すると、アプリケーションの開発および保守が複雑になります。

データベースを分割するには、プロセッサ間にデータおよび作業負荷を均等に分布させる必要があります、そのため高度の熟練が必要です。コンカレント・アクセス環境では、コードおよびデータの区分化は必要ありません。同じアプリケーションの複数のインスタンスを別々のプロセッサ上で実行し、共用外部ディスク上のデータベースに同時にアクセスさせることによって、スループットを向上させることができます。コンカレント構成では以下のディスクおよびディスク・アレイを使用できます。

- SCSI ディスク
- SCSI ディスク・アレイ
- シリアル・ディスク
- ファイバー・チャンネルと直接接続のディスク
- ファイバー・チャンネルと SAN 接続のディスク

コンカレント・アクセス論理ボリュームを作成するときは、独自のミラーリングを提供する RAID ディスク・サブシステムの場合を除き、ディスクが単一障害点になるのを避けるために LVM ミラーリングを使用してください。コンカレント・アクセスでは、ジャーナル・ファイルシステムの使用はサポートされません。したがって、データベース・マネージャーは、共用ボリューム・グループのロー論理ボリュームまたは `hdisks` に直接書き込む必要があります。アプリケーションでは、共用データへのすべての要求を調整するための方法を使用する必要があります。大半の商用 UNIX データベースでは、PowerHA SystemMirror ソフトウェアと互換性を保つロック・モデルを提供しています。特定のアプリケーションがコンカレント・アクセス処理をサポートしているかどうかは、データベースのベンダーにお問い合わせください。

関連概念:

65 ページの『拡張コンカレント・モード』

AIX には、新しい形式のコンカレント・モード (拡張コンカレント・モード) が用意されています。拡張コンカレント・モードでは、コンカレント論理ボリューム・マネージャー (CLVM) のインスタンスは、AIX の Reliable Scalable Cluster Technology (RSCT) 機能のグループ・サービス・コンポーネントを介してノード間の変更を調整します。PowerHA SystemMirror は自動的に、共用ボリューム・グループを拡張コンカレント・モードとして作成し、既存の共用ボリューム・グループを拡張コンカレント・モードに変換します。

コンカレント・リソース・マネージャー

PowerHA SystemMirror のコンカレント・リソース・マネージャーは、高可用性クラスターの共用ディスクへのコンカレント・アクセスを提供します。これにより、ビジネス・ニーズに合うように調整されたアクションをテークオーバー中に実行することができます。

コンカレント・リソース・マネージャーにより、あらゆる種類のディスク上の共用ボリューム・グループに対応する拡張コンカレント・サポートが追加されます。また、サポートされる RAID ディスク・サブシステム用のコンカレント共用アクセス管理も追加されます。

拡張コンカレント・モード:

AIX には、新しい形式のコンカレント・モード (拡張コンカレント・モード) が用意されています。拡張コンカレント・モードでは、コンカレント論理ボリューム・マネージャー (CLVM) のインスタンスは、AIX の Reliable Scalable Cluster Technology (RSCT) 機能のグループ・サービス・コンポーネントを介してノード間の変更を調整します。PowerHA SystemMirror は自動的に、共用ボリューム・グループを拡張コンカレント・モードとして作成し、既存の共用ボリューム・グループを拡張コンカレント・モードに変換します。

グループ・サービス・プロトコルは、クラスター・ノード間の通信リンク上を流れます。PowerHA SystemMirror によって複数ノードへの接続がサポートされているディスクは、拡張コンカレント・モード・ボリューム・グループに含めることができます。

補足クラスター・ソフトウェア

広い範囲に及ぶ追加ツールは、可用性の高いクラスターを AIX 環境に効率よく構築、管理、展開する上で役立ちます。

これに含まれるものは、次のとおりです。

- General Parallel File System (GPFS™) for AIX。複数ディスク・ドライブや複数ノードにわたるファイルにユーザーが共有アクセスすることが可能なクラスター全体のファイル・システムです。
- Workload Manager for AIX は、アプリケーション間のリソース・バランシングを提供します。
- PowerHA SystemMirror クラスター内で Oracle、DB2、WebSphere、およびその他のアプリケーションを構成する Smart Assist ソフトウェア。

関連情報:

PowerHA SystemMirror Smart Assist アプリケーション開発ガイド

Smart Assist for PowerHA SystemMirror

アプリケーションの可用性の確保

このセクションでは、PowerHA SystemMirror ソフトウェアがクラスター・コンポーネントの可用性を確保することで、どのようにアプリケーションの可用性を確保するかについて説明します。PowerHA SystemMirror は、すべてのキー・システム・コンポーネントについて単一障害点を除去し、ほとんどのルーチン・クラスター保守タスクについてスケジュールされたダウン時間の必要を除去します。

概説: アプリケーションの可用性

可用性の高いクラスターのキーとなる面は、提供する重要なサービスを中断する可能性のある変更を検出し応答できる能力です。PowerHA SystemMirror ソフトウェアは、キー・システム・コンポーネント (例え

ば、ネットワーク・インターフェース・カード) が使用できなくなっても、クラスターがインストールに必要なアプリケーション・サービスを提供し続けられるようにします。

コンポーネントが利用不可能になったとき、PowerHA SystemMirror ソフトウェアは、喪失を検出し、そのコンポーネントからのワークロードをクラスター内の別のコンポーネントにシフトすることができます。可用性の高いクラスターを計画する場合、キー・コンポーネントが確実に単一障害点でなくなるようにしようとします。

Smart Assist for PowerHA SystemMirror を使用して、DB2、SAP、Oracle、および WebSphere などの高可用性アプリケーションを管理できます。これらのアプリケーションを単一の PowerHA SystemMirror エンティティとして表示することが可能になり、このエンティティは PowerHA SystemMirror ソフトウェアで 1 つのアプリケーション名で示されます。

PowerHA SystemMirror ソフトウェアによって、クラスターは、動的再構成と呼ばれるプロセスを使用してルーチン保守タスクが実行される間、アプリケーション・サービスを提供し続けることができます。動的再構成では、実行中のクラスター内のコンポーネントを変更することができます。例えば、クラスター・サービスを停止し再始動する必要なしに、ノードまたはネットワーク・インターフェースを追加または除去することなどです。変更された構成が、動的にアクティブな構成になります。また、障害のあるディスクを動的に取り替えることもできます。以下のセクションでは、PowerHA SystemMirror ソフトウェアを使用して以下のことを行う方法について概念的に説明します。

- クラスター内の単一障害点の除去
- 動的再構成、リソース・グループ管理、およびクラスター管理 (C-SPOC) ユーティリティによる、PowerHA SystemMirror クラスター内のスケジュールされたダウン時間の最小化
- 素早いリカバリー機能、およびリソース・グループの遅延フォールバック・タイマー・ポリシーの指定による、スケジュールされないダウン時間の最小化
- ディスク・テークオーバーを実行するのにかかる時間の最小化
- クラスター・イベントの解釈およびエミュレート

注: キー・コンポーネントが障害を起こし、クラスターがアプリケーションの可用性を提供し続ける間、クラスター・アクティビティのモニターが必要となる場合があります。

関連概念:

105 ページの『PowerHA SystemMirror 構成プロセスおよび機能』

以下のトピックでは、PowerHA SystemMirror クラスター構成プロセス、および PowerHA SystemMirror ソフトウェアに付属の管理ツールについて概説します。

関連情報:

Smart Assist for PowerHA SystemMirror

PowerHA SystemMirror クラスター内の単一障害点の除去

PowerHA SystemMirror ソフトウェアでは、単一障害点 (SPOF) を除去することで、高可用性でありながら非常にスケーラブルなクラスターを構築できます。重要なクラスター機能が単一のコンポーネントによって提供されるときに、単一障害点が生じます。

そのコンポーネントに障害が発生すると、クラスターにはその機能を提供する他の方法がなく、重要なサービスが利用不可能になります。例えば、重要なアプリケーション用のデータがすべて、ミラーリングされていない単一ディスク上にある場合、そのディスクに障害が発生すると、そのディスクはシステム全体にとっての単一障害点となります。そのディスク上のデータが復元されるまで、クライアント・ノードはそのアプリケーションにアクセスできません。

PowerHA SystemMirror クラスター内の潜在的な単一障害点

高可用性であるためには、クラスターに単一障害点があってはなりません。

PowerHA SystemMirror には、以下のクラスター・コンポーネントに対する回復オプションが用意されています。

- ノード
- アプリケーション
- ネットワークおよびネットワーク・インターフェース
- ディスクとディスク・アダプター

すべての単一障害点の除去が目標であっても、妥協が必要となる場合があります。単一障害点を除去することにはコストが関連しているのが普通です。例えば、冗長なハードウェアはコストを増大させます。単一障害点を除去するコストを、そのコンポーネントが故障した場合に失われるサービスのコストと比較する必要があります。PowerHA SystemMirror ソフトウェアの目的は、将来の処理要求を満たすために拡張できる費用対効果の高い高可用性コンピューティング環境を提供することです。

単一障害点としてのノードの除去

ノードは、計画された変位 (ノードのシャットダウン、またはノード上のクラスター・サービスの停止)、または障害のどちらかのために、クラスターを去ります。

ノード障害は、隣接するノードをモニターするノードが、定義された時間についてハートビート・トラフィックの受信を停止するときに開始します。他のクラスター・ノードが、障害がノード障害であることに同意した場合、その障害ノードはクラスターから除去され、そのリソースは、そのように構成されたノードによって引き継がれます。例えば、アクティブ・ノードが、障害ノード上に構成された共用ディスクを制御できます。または、アクティブ・ノードは、(サービス IP アドレスを取得して) 障害ノードになりすまし、自身のプロセスを維持しながら、障害ノードのプロセスを実行することができます。このように、クライアント・アプリケーションは、共用ディスクとプロセッサ・サービスについて、残ったノードに切り替えできます。PowerHA SystemMirror ソフトウェアには、ノード障害を処理するために以下の機能が用意されています。

- ディスク・テークオーバー
- IP エイリアスによる IP アドレス・テークオーバー
- IP 置換による IP アドレス・テークオーバー (ハードウェア・アドレス・テークオーバーあり、または無しで)

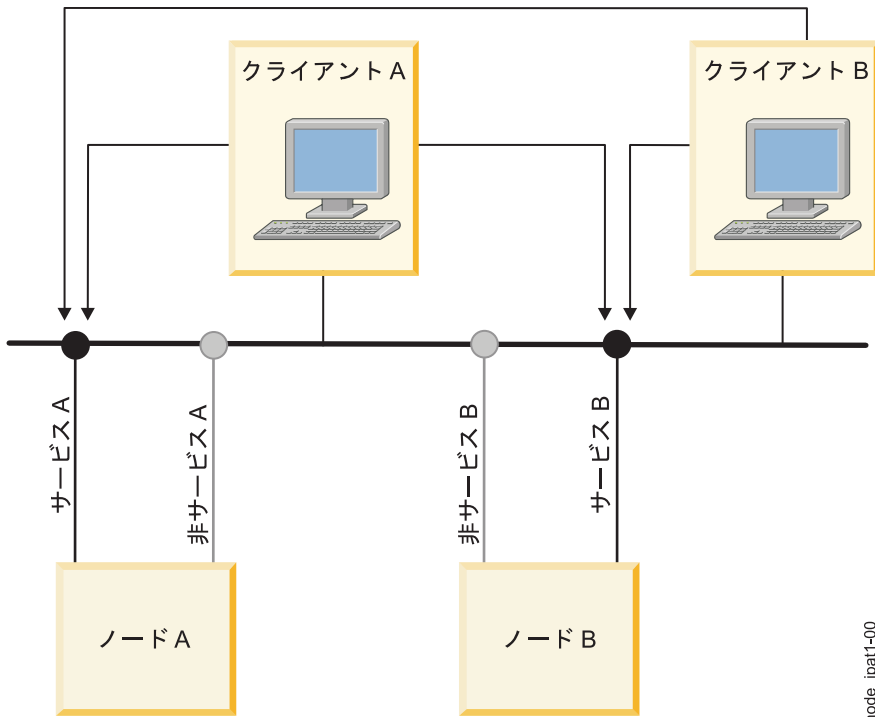
ディスク・テークオーバー:

PowerHA SystemMirror 環境では、共用ディスクは物理的に複数のノードに接続されています。

コンカレント環境でのディスク・テークオーバー

コンカレント・アクセス構成では、共用ディスクはアクティブに同時に複数のノードに接続されています。したがって、ディスク・テークオーバーは、ノードがクラスターを去るときには必要ありません。

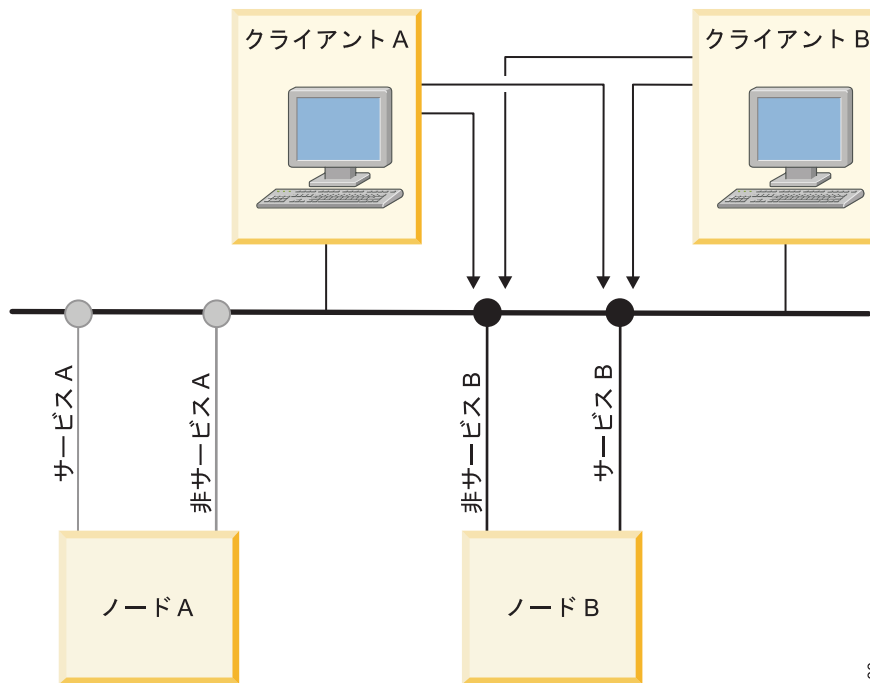
以下の図は、コンカレント環境におけるディスク・テークオーバーを図示しています。



c_2node_ipati-00

各ノードは、個別のネットワーク・サービスを提供します。

図 11. ディスク・テークオーバーの前のコンカレント・アクセス構成



ノード B は、ノード A の非サービス・インターフェース上に
あるノード A の IP ラベルをテークオーバーし、このインター
フェース上のクライアントにノード A のアプリケーション・
サービスを提供します。

c_2node_ipat2-00

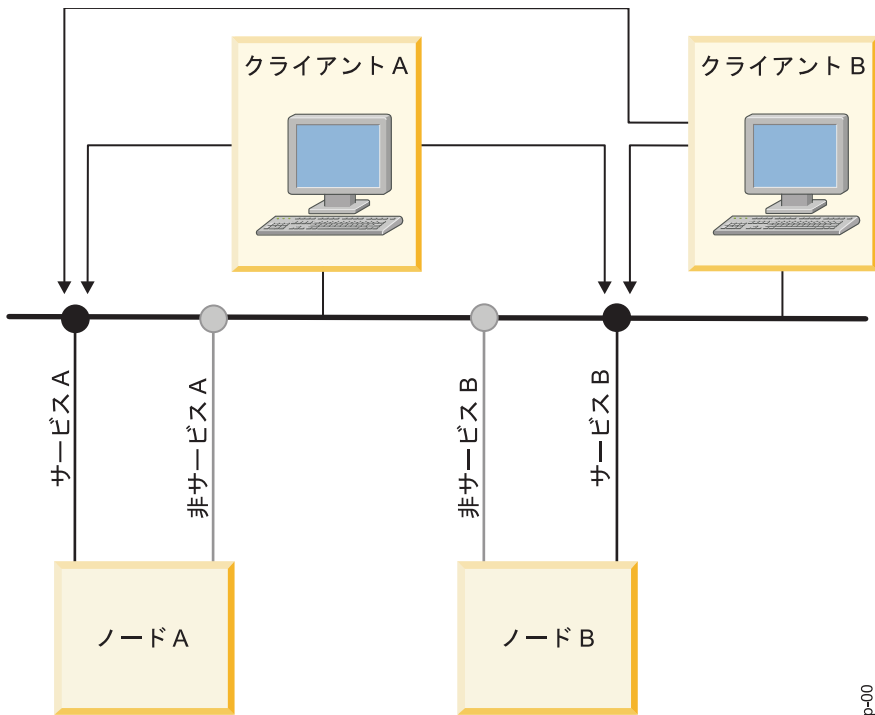
図 12. ディスク・テークオーバーの後のコンカレント・アクセス構成

IP アドレス・テークオーバー:

IP アドレス・テークオーバー (IPAT) は、クラスターを離れたノードのネットワーク・アドレスをノードが取得できるようにするネットワーク機能です。

クライアントに提供中のサービスが特定の IP アドレスに結合されているとき、つまり、サービスがクライアントに提供される際に使用されるサービス IP ラベルが、クラスター・リソース・グループ内にリソースとして含まれているときに、PowerHA SystemMirror では IP アドレス・テークオーバーが必要になります。IPAT を実行する代わりに、残りのノードが単にディスクおよびアプリケーションのテークオーバーを行うと、クライアントは、指定のサーバー IP アドレスでアプリケーションを使用し続けることはできなくなります。PowerHA SystemMirror はエイリアスを介して IPAT を使用します。

以下の図は、IP アドレス・テークオーバーを図示しています。



各ノードは、個別のネットワーク・サービスを提供します。

図 13. IP 置換による IP アドレス・テークオーバーの前の構成

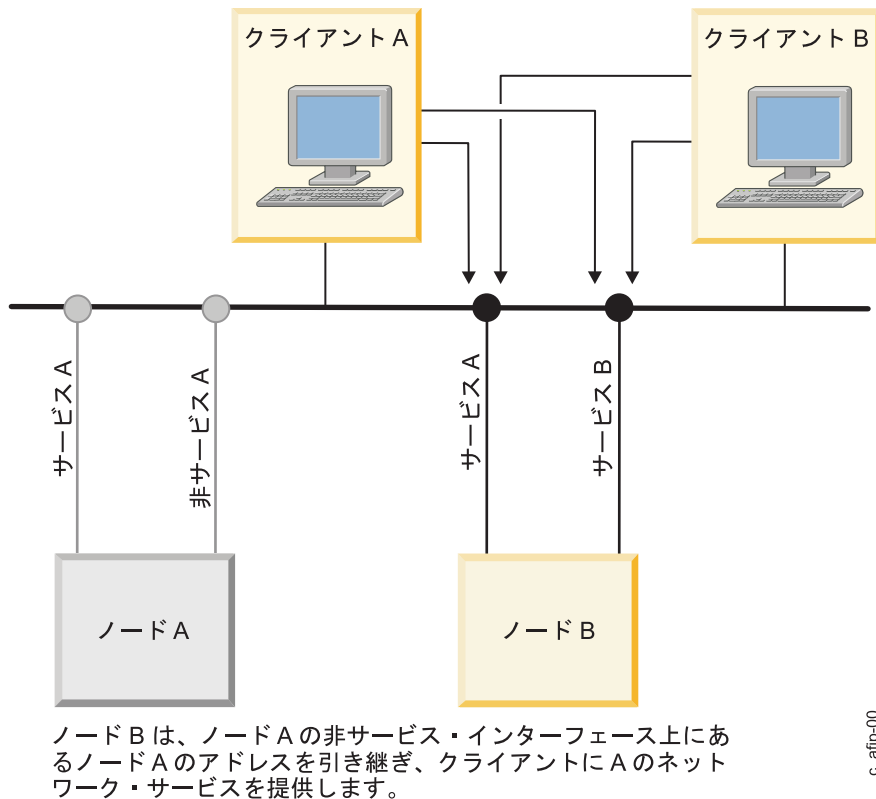


図 14. IP 置換による IP アドレス・テークオーバーの後の構成

関連概念:

36 ページの『IP エイリアスによる IP アドレス・テークオーバー』
 AIX オペレーティング・システムの IP エイリアス・ネットワーク機能を使用して、特定タイプのネットワーク上で IP アドレス・テークオーバーを構成できます。

単一障害点としてのアプリケーションの除去

PowerHA SystemMirror クラスターを作成する主要な理由は、主幹業務アプリケーション用に可用性の高い環境を提供することです。

例えば、PowerHA SystemMirror クラスターは、クライアント・アプリケーションにサービスするデータベース・サーバー・プログラムを実行できます。クライアントは、共用外部ディスクに格納されたデータベースにアクセスすることによって、要求に応えるサーバー・プログラムにクエリーを送信します。

PowerHA SystemMirror クラスターでは、これらの重要なアプリケーションが単一障害点になる可能性があります。これらのアプリケーションの可用性を確実にするために、クラスターを去ってノードのリソースをテークオーバーするよう構成されたノードは、これらのアプリケーションがクライアント・プロセスで引き続き利用できるよう、再始動する必要があります。以下を使用してアプリケーションの可用性を高くすることができます。

- アプリケーション・コントローラー
- クラスター制御
- アプリケーション・モニター
- アプリケーション可用性分析ツール

アプリケーションを PowerHA SystemMirror の制御下に置くために、サーバーのユーザー定義名を、アプリケーションを開始および停止するよう書かれたユーザー作成スクリプトの名前と対応させるアプリケーション・コントローラー・クラスター・リソースを作成します。アプリケーション・コントローラーを定義することによって、PowerHA SystemMirror は、フォールオーバーが発生したときに、テークオーバー・ノード上でアプリケーションの別のインスタンスを開始できます。

特定のアプリケーションは、アプリケーション・コントローラーなしで可用性を高めることができます。リソース・グループの一部としてのアプリケーションの面を構成することによって、このようなアプリケーションをクラスター制御下に置くことができます。例えば、Fast Connect のサービスを、すべてリソースとしてクラスター・リソース・グループに追加し、ノードまたはネットワーク・インターフェース障害の場合にそれらの可用性を高めることができます。

注: アプリケーション・テークオーバーは、通常、IP アドレス・テークオーバーと関連付けられます。アプリケーションを再始動しているノードが障害ノード上の IP サービス・アドレスも取得した場合、クライアントは、同じサーバー IP アドレスにただ再接続するだけです。IP アドレスがテークオーバーされなかった場合、クライアントは、新しいサーバーに接続して、アプリケーションへのアクセスを継続する必要があります。

さらに、AIX システム・リソース・コントローラー (SRC) を使用して、アプリケーション・デーモンの存在または不在をモニターし、それに応じて応答することができます。

アプリケーション・モニター

アプリケーション・モニターを構成して、プロセス障害または別のアプリケーション障害をチェックし、アプリケーションを再始動する処置を自動的に取ることもできます。

複数のアプリケーション・モニターを構成し、それらを 1 つ以上のアプリケーション・コントローラーに関連付けることができます。アプリケーションごとに複数のモニターをサポートすることにより、PowerHA SystemMirror は、より複雑な構成をサポートできます。例えば、使用中の Oracle 並列サーバーの各インスタンスについて 1 つのモニターを構成できます。または、ユーザー定義モニターを構成してデータベースの状態をチェックし、プロセス終了モニターが直ちにデータベース・プロセスの終了を検出できるようにできます。

アプリケーション可用性分析ツール

アプリケーション可用性分析ツールは、アプリケーションが利用可能になった時間の正確な量を測定します。PowerHA SystemMirror ソフトウェアは、このツールでモニターすることを選択したアプリケーションについて包括的な情報を収集し、タイム・スタンプを取り、ログに記録します。SMIT を使用すると、期間を選択することができ、ツールは、その期間内の特定のアプリケーションについてのアップ時間とダウン時間の統計を表示します。

単一障害点としての通信インターフェースの除去

PowerHA SystemMirror ソフトウェアは、サービス IP ラベルが構成されているネットワーク・インターフェースの障害を取り扱います。

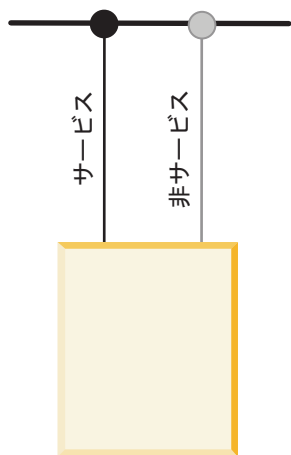
そのような障害の 2 つのタイプは次のとおりです。

- ノード上で構成された 2 つのネットワーク・インターフェースのうち、サービス IP ラベルのあるネットワーク・インターフェースは失敗しますが、追加バックアップ・ネットワーク・インターフェース・カードは同じノード上で利用できます。この場合、クラスター・マネージャーは、そのノード上でこれ

ら 2 つのインターフェースの役割をスワップします。このようなネットワーク・インターフェース障害は、あるノード上でシステムがネットワーク・インターフェースを再構成する間の短い遅延を除いて、ユーザーに透過です。

- ノード上で構成された 2 つのネットワーク・インターフェースのうち、追加、またはバックアップ・ネットワーク・インターフェースは失敗しますが、そのノード上に構成されたサービス IP ラベルのあるネットワーク・インターフェースはそのまま利用できます。この場合、クラスター・マネージャーは、バックアップ・ネットワーク・インターフェース障害を検出し、イベントをログに記録し、システム・コンソールにメッセージを送信します。追加処理が必要な場合は、このイベントについて処理をカスタマイズできます。

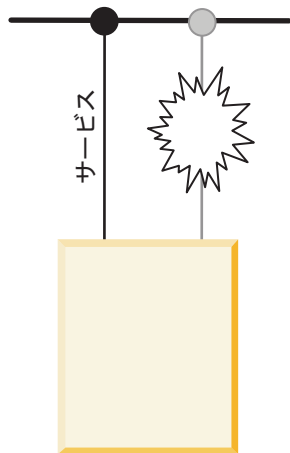
以下の図は、同じノードで発生するネットワーク・インターフェース・スワッピングを図示しています。



ここでは、サービス・インターフェースによってネットワークへの接続が提供されています。非サービス・インターフェースは、アプリケーションからは隠され、クラスター・マネージャーにのみ認識されるようにします。

c_adapswrap1-00

図 15. ネットワーク・アダプター・スワップの前の構成



ここでは、サービス・インターフェースで障害が発生し、クラスター・マネージャーが、それまで非サービスだったインターフェースを新しいサービス・インターフェースとして指定しています。

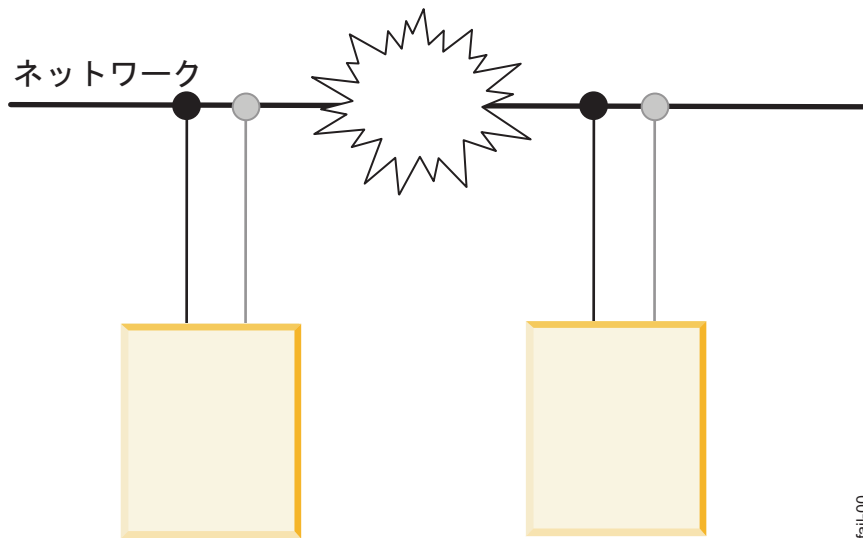
c_adepswap2-00

図 16. ネットワーク・アダプター・スワップの後の構成

単一障害点としてのネットワークの除去

ネットワーク障害は、クラスター内のすべてのノードについて PowerHA SystemMirror ネットワークが失敗したときに発生します。このタイプの障害が発生するのは、指定の PowerHA SystemMirror ネットワークについて構成されたネットワーク・インターフェース・カードを使用して、クラスター・ノードのどれもがお互いにアクセスできないときです。

次の図にネットワーク障害を図示します。



ここでは、ノード同士を接続するネットワークで障害が発生しています。該当するノードは、このネットワークを使用して通信できなくなりました。

c_2node_nettai-00

図 17. ネットワーク障害

ネットワーク障害に対する PowerHA SystemMirror ソフトウェアの最初の防御のラインは、クラスター内のノードを複数のネットワークによって接続させることです。1 つのネットワークが失敗すると、PowerHA SystemMirror ソフトウェアは、クラスター・トラフィックおよびノードのステータス・モニターに依然として利用できるネットワークを使用します。

追加の処置を指定して、ネットワーク障害を処理できます。例えば、代替りのネットワークを介した再ルーティングです。ネットワーク障害に対して保護する少なくとも 2 つのネットワークを備えることを強くお勧めします。

ローカル・ネットワーク障害イベントが発生するとき、クラスター・マネージャーは、そのネットワークに接続されたサービス IP ラベルを含むリソースのグループについて選択的リカバリー処置を取ります。クラスター・マネージャーは、特定のノード上のリソース・グループをすべて移動するのではなく、ローカル・ネットワーク障害イベントにより影響を受けるリソース・グループのみ移動を試みます。

ノード分離と区分クラスター:

ノード分離が発生するのは、クラスターの 2 つ以上の部分を接続するすべてのネットワークが失敗したときです。ノードの各グループ (1 つ以上) が他のグループから完全に分離されます。区分クラスターは、特定のグループのノードが他のグループのノードと通信できないクラスターです。

区分クラスターの次の図では、ノード A とノード C が区分の一方の側にあり、ノード B とノード D が区分の他方の側にあります。

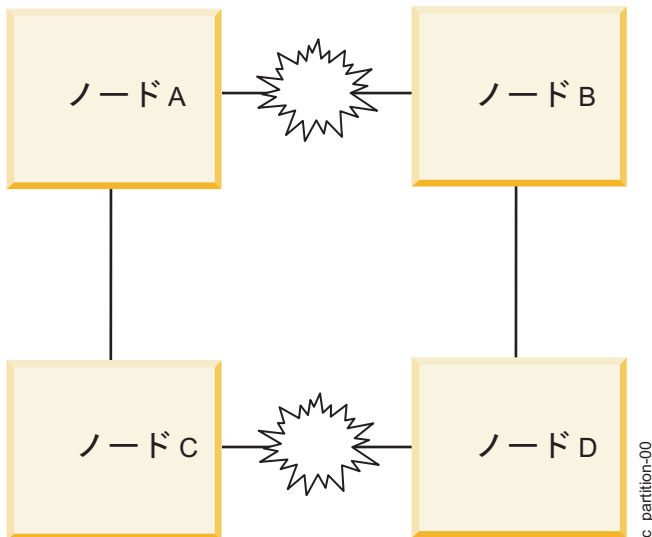


図 18. 区分クラスター

区分クラスターの問題は、区分の一方の側のノードが区分の他方の側のノードからのハートビートの不在に割り込み、これらのノードが失敗し、これらのノードについてノード障害イベントを生成するという意味です。いったんこの状況が発生すると、クラスターの各側のノードは (そのように構成されている場合)、依然としてアクティブであるために正当にこれらのリソースを所有するノードから、リソースをテークオーバーしようとしてみます。このようなテークオーバーが試みられた場合、ディスクがリセットされてデータが破損するなど、予想不能な結果がクラスターで生じる可能性があります。

Storage Area Network を使用して区分化を予防:

TCP/IP ネットワークが失われることでクラスターが区分化されるのを予防するために、Cluster Aware AIX ソフトウェアはクラスター・ノードにまたがる Storage Area Network を介してハートビートを提供します。

関連情報:

PowerHA SystemMirror プランニング・ガイド

単一障害点としてディスクおよびディスク・アダプターを除去

PowerHA SystemMirror ソフトウェアは、ディスクおよびディスク・アダプター障害を直接には処理しません。代わりに、AIX が、ディスク上の LVM ミラーリングを使用して、さらにディスク上の内部データ冗長性によって (この機能を備えたディスクの場合)、このような障害を処理します。

例えば、複数のストレージ・アダプター、ストレージ・エリア・ネットワークを使用したハートビート、および LVM ミラーリングでシステムを構成することによって、ディスク・サブシステム (アダプター、ケーブルリング、ディスク) 内の単一のコンポーネントが、ディスク上のデータを利用不可にすることなく失敗する場合があります。

IBM 2105 ESS またはその他の RAID ストレージ・サブシステムを使用している場合は、ディスク・アレイ自体が、データ冗長性を提供します。

AIX エラー通知機能

AIX エラー通知機能を使用すると、PowerHA SystemMirror ソフトウェアでは特にモニターされないイベントを検出することができます。例えば、ディスク・アダプターの障害を検出し、そのイベントに対する応答をプログラムすることができます。

ディスク・ドライブ、コントローラー、またはアダプター上の永続的なハードウェア・エラーはデータの障害弾力性に影響を及ぼします。エラー通知方法によってこれらのエラーをモニターすることによって、高い可用性を提供するクラスターの障害の影響を評価できます。エラー通知の単純なインプリメンテーションは、メール・メッセージをシステム管理者に送信して、問題をさらに調査してもらうことです。より複雑なインプリメンテーションについては、障害を分析し、処理を続けるか、処理を停止するか、または障害をノード障害まで拡大させてテークオーバー・ノードをクライアントに利用可能なボリューム・グループ・リソースにするか、という論理を組み入れることができます。

ディスク・サブシステムに影響を及ぼすすべてのエラーについて、エラー通知メソッドをインプリメントすることを強くお勧めします。そうすると、低下した障害回復力が必ず検出されるようになります。

AIX エラー通知メソッドは、PowerHA SystemMirror 内で自動的に使用され、ボリューム・グループ喪失エラーのような、特定のリカバリー可能な LVM エラーをモニターします。

自動エラー通知

SMIT で特定のオプションを使用すると、クラスター・リソースに対して自動的にエラー通知を構成できます。このオプションを選択すると、特定のデバイス用のクラスター内のすべてのノードについてエラー通知が自動的にオンになります。

ディスク・アダプター・エラーなど、特定の回復不能エラー・タイプは、自動エラー通知によってサポートされます。この機能は、メディア・エラー、リカバリーされたエラー、一時エラーはサポートしません。2つのエラー通知メソッドのうち 1 つは、自動エラー通知によってサポートされるすべてのエラー・タイプに割り当てられます。

さらに、リソース・グループにボリューム・グループを追加した場合、PowerHA SystemMirror はそのボリューム・グループ用の AIX エラー通知メソッドを作成します。ボリューム・グループがクォーラムを喪失した場合、PowerHA SystemMirror は、この方式を使用して、影響のあるリソース・グループを別のノードに選択的に移動します。PowerHA SystemMirror が生成したエラー通知メソッドは編集または変更しないでください。

PowerHA SystemMirror エラー通知機能は、MPIO ディスクをサポートしません。

予定されたダウン時間を PowerHA SystemMirror で最小限に抑える

PowerHA SystemMirror ソフトウェアを使用すれば、変更した構成をアクティブにするためにクラスター・サービスを停止して再始動することなく、アクティブなクラスター上で多くの日常的な保守タスクを動的に実行できます。

アプリケーションを停止せずにクラスター・サービスを開始する

アプリケーションを停止せずにノード上で PowerHA SystemMirror クラスター・サービスを開始できます。

関連情報:

クラスター・サービスの開始および停止

動的自動再構成

アクティブなクラスターを変更した後でクラスター構成を同期化するとき、動的自動再構成 (DARE) が起動されます。SMIT を使用したクラスター・スナップショットを適用しても、動的再構成イベントを起動できます。

例えば、実行中のクラスターにノードを追加するには、ノードをクラスターに接続し、既存のクラスター・ノードのいずれかの上のクラスター・トポロジーにノードを追加し、クラスターを同期化するだけです。新しいノードがすべてのクラスター・ノード上のクラスター・トポロジー定義に追加され、変更された構成が現時点でアクティブな構成になります。動的再構成イベントが完了した後、新しいノードでクラスター・サービスを開始できます。

PowerHA SystemMirror は、変更された構成を確認したあとで、それを現時点でアクティブな構成にし、行った変更の結果が有効な構成であることを確認します。

動的再構成の動作:

実行中のクラスターで動的自動再構成 (DARE) を使用する場合、PowerHA SystemMirror が開始すると、DARE によってシステム・デフォルト・オブジェクト・データ・モデル (ODM) に格納された PowerHA SystemMirror 特有のオブジェクト・クラスの専用コピーが作成されます。

これ以降、ODM は、PowerHA SystemMirror 構成データベースと呼ばれます。構成データベースのデータは、次の 2 つのディレクトリーに保管されます。

- アクティブ構成ディレクトリー (ACD)。このディレクトリーは専用ディレクトリーであり、実行中のノード上のあらゆる PowerHA SystemMirror デモン、スクリプト、およびユーティリティーによる参照用に PowerHA SystemMirror 構成データベースのデータを保管します。
- デフォルト構成ディレクトリー (DCD)。このディレクトリーはシステム・デフォルト・ディレクトリーであり、PowerHA SystemMirror 構成データベースとデータを保管します。

注: DARE の運用についてここで完全に説明します。PowerHA SystemMirror にこれらの操作を確実に実行させるために、手動で操作を行う必要はありません。PowerHA SystemMirror は、クラスター内のすべての動的再構成操作を正しく管理します。

DCD は `/etc/objrepos` という名前のディレクトリーです。このディレクトリーには、カスタマイズ・デバイス・データベース (CuDv)、事前定義デバイス・データベース (PdDv)、特定の PowerHA SystemMirror オブジェクト・クラスなどのデフォルトのシステム・オブジェクト・クラスが含まれています。ACD は `/usr/es/sbin/cluster/etc/objrepos/active` です。

注: クラスターを構成するときは、ACD 内のデータではなく、DCD に格納された PowerHA SystemMirror 構成データベース・データを変更します。SMIT、および他の PowerHA SystemMirror 構成ユーティリティーはすべて、DCD 内の PowerHA SystemMirror 構成データベース・データを変更します。さらに、`cllsif` コマンドのような、PowerHA SystemMirror 構成データベース・データを表示するすべてのユーザー・コマンドは DCD からデータを読み取ります。

次の図では、PowerHA SystemMirror デモン、スクリプト、およびユーティリティーがすべて、構成情報にアクセスするときに、どのようにして ACD を参照するかを説明しています。

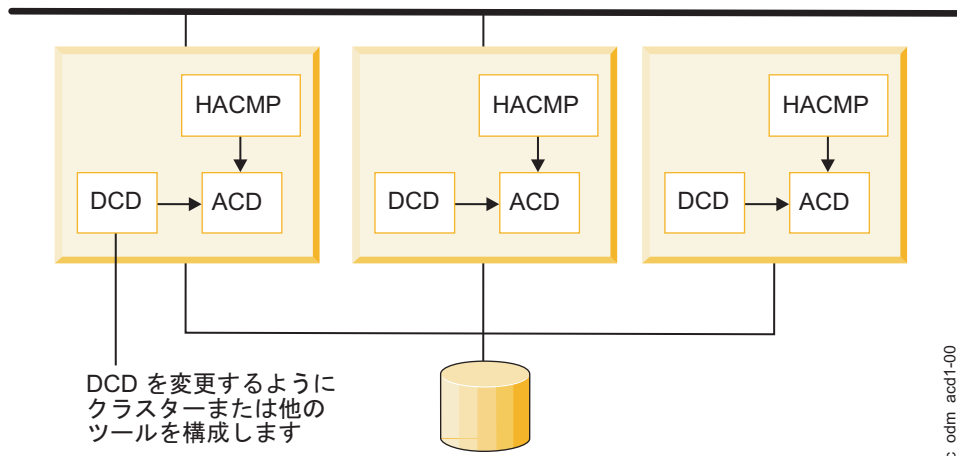


図 19. クラスター開始時の PowerHA SystemMirror から ACD への関係。

クラスターを動的に再構成:

PowerHA SystemMirror ソフトウェアは、構成データを格納する特定の PowerHA SystemMirror 構成データベース・リポジトリの位置に依存します。

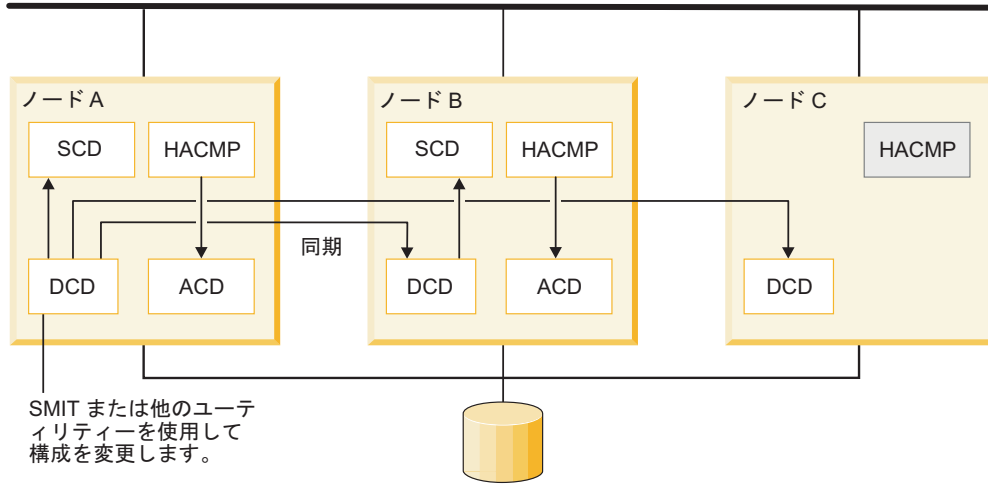
こうしたリポジトリの有無が、クラスターの構成時や操作時の処理を決定するために使用されることもあります。ODMPATH 環境変数を使用すれば、PowerHA SystemMirror 構成データベース・コマンドおよびサブルーチンは、照会対象オブジェクトがデフォルト位置に存在しない場合に、(ODMDIR 環境変数に保持されている) デフォルト位置以外の位置を照会できるようになります。この変数は設定できますが、`/etc/objrepos` ディレクトリーを含むように設定してはなりません。そのように設定してしまうと、PowerHA SystemMirror 構成情報の健全性が失われます。

アクティブ・クラスターの構成を変更する場合は、SMIT を使用してデフォルト構成ディレクトリー (DCD) 内の PowerHA SystemMirror 特有の構成データベース・クラスに格納されているクラスター定義を変更します。アクティブ・クラスターのクラスター構成を変更するときは、すべての変更と同じ SMIT のパスを使用しますが、変更は即時には適用されません。したがって、一度の操作で複数の変更を行うことができます。すべてのクラスター・ノードで構成内容を同期化させると、クラスター全体に及ぶ動的再構成イベントが始動します。PowerHA SystemMirror は、動的再構成イベントを処理する場合、各クラスター上の DCD 内に格納されている PowerHA SystemMirror 構成データベース・オブジェクト・クラスを更新し、調整されたクラスター全体に及ぶ移行の中で、ACD 内に格納されている PowerHA SystemMirror 構成データベース・データを、新しい PowerHA SystemMirror 構成データベース・データで置換します。また、クラスター・デーモンがこの新しい構成データを参照するようリフレッシュします。

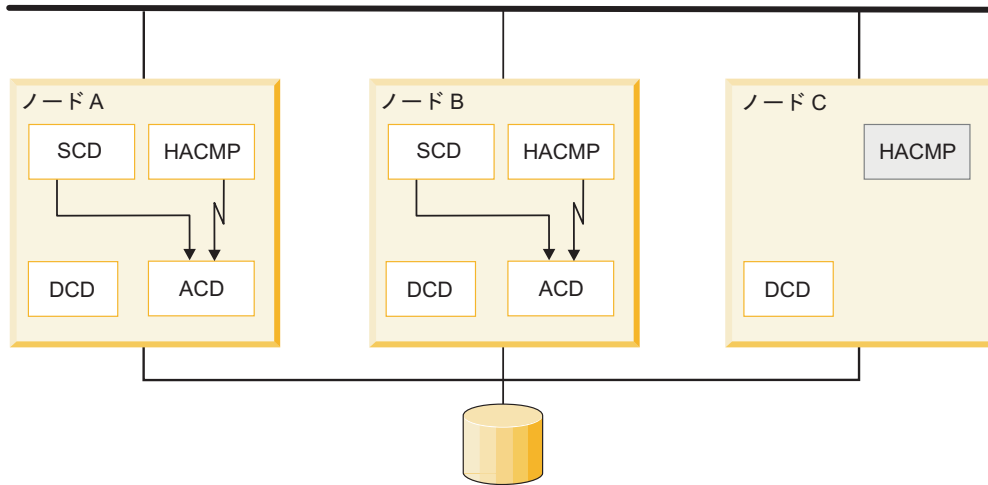
この処理のあと、クラスター・ハートビートが少しの間中断され、クラスターが不安定な状態になります。変更された構成がアクティブ構成になります。新規に追加されたノード上でクラスター・サービスが始動すると、そのクラスターに自動的に組み込まれます。

次の図では、動的再構成を使用してアクティブ・クラスターにノードを追加する際に行われる処理について説明しています。

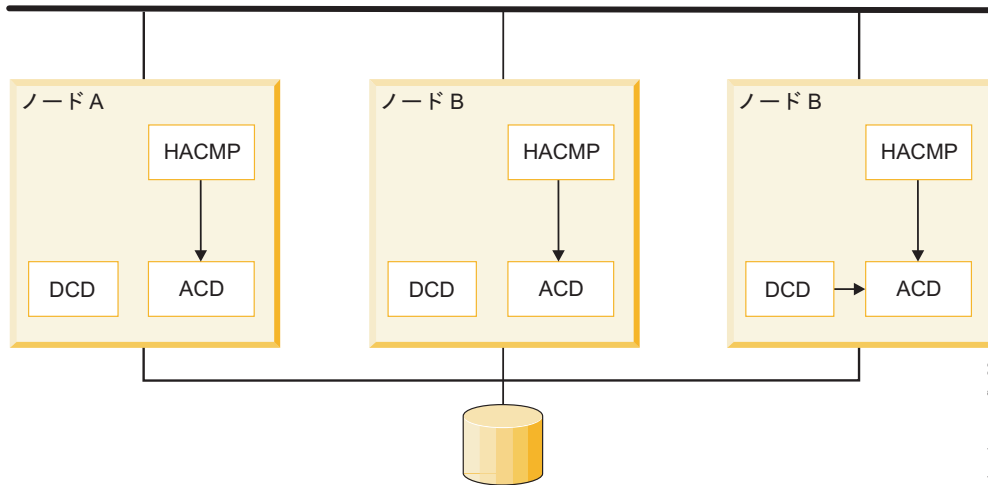
同期 1: DCD が同期化されます



同期 2: デーモンがリフレッシュされます



同期 3: 動的再構成が完了しました。NodeC でクラスター・サービスが開始されます



c_odm_dynreconfig-00

図 20. 動的再構成処理

追加されるノードは、実行中のクラスターに接続されますが、クラスター・サービスはこのノード上では非アクティブです。この構成はノード A に再定義されます。この構成に対する変更内容が同期化されると、ノード A 上の DCD 内に格納されている PowerHA SystemMirror 構成データベース・データが、他のクラスター・ノード上の DCD にコピーされ、動的再構成イベントが起動します。PowerHA SystemMirror は、DCD 内の新規 PowerHA SystemMirror 構成データベース・データを各ノード上のステージング構成ディレクトリー (SCD) と呼ばれる一時的な場所にコピーします。SCD のロケーションは /usr/es/sbin/cluster/etc/objrepos/stage です。この一時的な場所を使用することにより、PowerHA SystemMirror では、動的再構成の進行中に構成に追加変更を加えることができます。ACD 内の現在の PowerHA SystemMirror 構成データベース・データ上に SCD 内の新しい PowerHA SystemMirror 構成データベース・データをコピーする前に、PowerHA SystemMirror は新しい構成を確認します。

注: 動的再構成の進行中に 2 番目の再構成を開始できますが、それを同期化することはできません。クラスター・ノード上にステージング構成ディレクトリーが存在すると、ロックの役目を果たすため、新規動的再構成の開始ができなくなります。

リソース・グループ管理

リソース・グループ管理 (clRGmove) ユーティリティーを使用して、リソース・グループを他のクラスター・ノードおよびサイトに移動したり、クラスター・サービスを停止せずにそれらをオンラインまたはオフラインにしたりできます。

これにより、リソース・グループおよびそのアプリケーションの管理に柔軟性を持たせることができます。また、このユーティリティーを使用して、任意のリソース・グループのノードを解放し、特定のクラスター・ノード上のシステム保守を実行することができます。PowerHA SystemMirror リソース・グループ管理ユーティリティー (clRGmove) の機能が大幅に改良され、クラスター管理のためのリソース・グループの移動がさらに簡単になりました。また、手動でリソース・グループを移動した場合の影響を、より容易に把握できるようになりました。例えば、リソース・グループが、その移動先のノードにとどまるかどうかについて予測することができます。PowerHA SystemMirror は、「どのような場合でも、移動されたリソース・グループは再度移動されるまでノード上にとどまる」という原則に従います。PowerHA SystemMirror は、リソース・グループの回復が必要になったときにそのグループを移動します。

リソース・グループが現在使用可能な最も優先順位の高いノードに現時点でホストされているかどうかを確認したい場合、PowerHA SystemMirror はノードとサイト用のインテリジェントなピック・リストの選択項目を提示します。例えば、現在、あるノードでグループがホストされているときに、PowerHA SystemMirror がそのノードよりも優先順位が高い別のノードを検出すると、宛先ノードの SMIT ピック・リストにより、より高い優先順位を持つノードが示されます。このように、該当のグループを該当のノードに移動することをいつでも選択できます。グループを他のノードに移動する場合は、次のルールが適用されます。

- 「Never Fallback (フォールバックしない)」というフォールバック・ポリシーを持つリソース・グループの場合、グループを移動しても、今後のクラスター・イベント時にそのグループの動作が影響を受けることはありません。このことは、「Online on Either Site (片方のサイトでオンライン)」サイト・フォールバック・ポリシーを持つリソース・グループについても当てはまります。
- 「フォールバックしない」または「1 次サイトを優先」以外のフォールバック・ポリシーを持つリソース・グループの場合は、グループを移動すると、再度移動されない限り、宛先ノードが活動中の最高優先順位ノードになります。ノードの移動後は、それが活動中の最高優先順位ノードになります。

この動作についての重要な考慮事項の 1 つに、「最高優先順位ノードへのフォールバック」ポリシーまたは「1 次サイトを優先」ポリシーを持つリソース・グループに関するものがあります。このようなりソース・グループをその最高優先順位ノード以外のノード、または 1 次サイトに移動すると、リソース・グループの移動先のノードは一時的に「優先度の高い」ノードになりますが、(構成に従った) 最高優先順位の

ノードにはなりません。このようなグループは、再度移動するまでは、移動先のノードにとどまります。また、そのグループはこれらのノードまたはサイトにフォールバックします。

関連概念:

105 ページの『PowerHA SystemMirror 構成プロセスおよび機能』

以下のトピックでは、PowerHA SystemMirror クラスター構成プロセス、および PowerHA SystemMirror ソフトウェアに付属の管理ツールについて概説します。

関連情報:

クラスター内のリソース・グループの管理

リソース・グループの計画

ユーザー要求リソース・グループ管理と自動リソース・グループ管理:

一般に、PowerHA SystemMirror はアプリケーションの可用性を高く保つために、リソース・グループとそれに含まれるアプリケーションを自動的に管理 (および場合によっては移動) します。

例えば、リソース・グループを回復する必要があるときに、PowerHA SystemMirror はフォールオーバーまたはフォールバック操作時に、そのグループを別のノード上で自動的に回復しようとする場合があります。PowerHA SystemMirror は、グループを移動する際、ユーザーが指定したリソース・グループ・ポリシーおよびその他の設定に従います (例えば、障害が発生したリソース・グループを別のノードで自動的に回復するのではなく、グループの障害を単に通知するよう PowerHA SystemMirror に指示することができます)。

PowerHA SystemMirror にリソース・グループ管理を実行するよう要求すると、*clRGmove* ユーティリティを使用します。このユーティリティは、*rg_move* イベントを呼び出してリソース・グループを移動します。

注: ログ・ファイルをトラブルシューティングする場合、PowerHA SystemMirror によって自動的に起動されることのある *rg_move* イベントと、PowerHA SystemMirror にリソース・グループを管理するよう要求したときに発生する *rg_move* イベントを区別することが重要です。クラスター内のリソース・グループについて実行される操作の原因を識別するには、SMIT 内のコマンド出力と、*hacmp.out* ファイル内の情報を探します。

リソース・グループ管理操作:

リソース・グループ管理で実行できるタスクがいくつかあります。

リソース・グループ管理を使用して次のことを行います。

- あるサイトのノードから別のサイトのノードにリソース・グループを移動する。
- あるノードから別のノードにリソース・グループを移動する。

作業クラスターでは、非コンカレント・リソース・グループを、現在それが配置されているノードから任意の宛先ノードに一時的に移動します。移動するリソース・グループは、それらの構成方法と整合した方法で動作し続けます。すなわち、それらのリソース・グループに指定された起動、フォールオーバー、およびフォールバックの各ポリシーに従います。SMIT ユーザー・インターフェースを使用することにより、リソース・グループを別のノードに移動することを決定した場合に、そのリソース・グループの動作を明確に指定し、予測することができます。

SMIT を使用して、リソース・グループを別のノードに移動する場合、手動で再度移動するまで、そのリソース・グループは新規の宛先ノードにとどまります。PowerHA SystemMirror では、場合によっては、フォールオーバー時にリソース・グループを移動する必要があります。

- リソース・グループを、もともと優先順位が一番高かったノードに再度移動する。

リソース・グループは、フォールバック・ポリシーを持っていることも持っていないこともあります。

「Fallback to Highest Priority Node (最高優先順位ノードへのフォールバック)」のフォールバック・ポリシーを持つリソース・グループを移動すると、そのグループは、「新規」ノードが現時点で優先度の高い一時ロケーションであると想定し、そのノードにフォールバックします。この動作を変更するために、当初優先順位が一番高かったノードにいつでもそのグループを再移動できます。

同様に、フォールバック・ポリシー「Never Fallback (フォールバックしない)」を持つリソース・グループを移動すると、このグループは移動元のノードに戻されることはなく、別のノードに再度移動されるまで新規の宛先ノードにとどまります。このように、グループが常に、それに指定された「Never Fallback (フォールバックしない)」ポリシーに確実に準拠するようにすることができます。

- クラスター内の 1 つまたはすべてのノードでリソース・グループをオンラインまたはオフラインにする。

関連情報:

Administering PowerHA SystemMirror

Cluster single point of control

Cluster single point of control (C-SPOC) ユーティリティを使用すれば、単一のクラスター・ノードからクラスター全体を変更できます。

それぞれのクラスター・ノードで管理タスクを実行する必要はなく、1 つのノードで SMIT インターフェースを使用して C-SPOC コマンドを一度実行すれば、その変更がすべてのクラスター・ノードに伝搬されます。

関連概念:

110 ページの『PowerHA SystemMirror システム管理 (C-SPOC)』

クラスターの管理を容易にするため、PowerHA SystemMirror では、あるノードからコマンドを実行してから、変更内容を検証して他のすべてのノードと同期化する方法を提供しています。PowerHA SystemMirror システム管理ツール、Cluster Single Point of Control (C-SPOC) を使用すれば、主幹業務のジョブを停止することなく、ユーザー、ファイル、およびハードウェアを自動的に追加することができます。

動的アダプター・スワップ

動的アダプター・スワップ機能によって、アクティブなネットワーク・インターフェース・カード (NIC) の IP アドレスを、同じノードとネットワーク上のユーザー指定のアクティブで利用可能な「バックアップ」ネットワーク・インターフェース・カードの IP アドレスとスワップすることができます。

スワップを実行するためにクラスター・サービスを停止する必要はありません。

この機能を使用すると、ノードをシャットダウンせずに、動作が不安定なネットワーク・インターフェース・カードから別の NIC に IP アドレスを切り替えることができます。この機能は、ホット・プラグ可能な NIC をノード上で交換する場合にも使用できます。ホット・プラグ可能な NIC は、ノードの電源を切らずに物理的に取り外して交換できます。交換したい (ホット・プラグ可能な) NIC をノードから除去すると、PowerHA SystemMirror はその NIC をバックアップとして使用不能な状態にします。

SMIT を使用してアダプター・スワップを構成できます。サービス IP アドレスが現行の NIC からユーザー指定の NIC に移行されます。これにより、このサービス IP アドレスは、使用可能な「バックアップ」

アドレスになります。ノードに新しいカードが装着されると、その NIC は使用可能な「バックアップ」として再びクラスターに組み込まれます。このあとで、IP アドレスをバックアップ NIC から元の NIC に交換することができます。

注: このタイプの動的アダプター・スワップは、同じノード内でのみ実行できます。この機能では、あるノード上の IP アドレスを別のノード上の IP アドレスと交換することはできません。あるサービス IP アドレスを別のノードに移すには、リソース・グループ管理ユーティリティを使用して、そのリソース・グループを移動させます。

自動検証および同期化

自動検証および同期化によって、クラスターにノードを追加する際のダウン時間が最小化されます。

この処理は、クラスター・サービスを開始する前に実行し、クラスターに結合中のノードが適切に同期化されることを確認します。この処理では、アクティブまたは非アクティブな構成に加わるノードをチェックします。自動検証および同期化によって修正される典型的な構成不整合の例を以下に挙げます。

- IP アドレスを RSCT が予期するネットワーク・インターフェース上で構成する
- 共有ボリューム・グループが自動的に varyon されるようには設定されない
- ファイルシステムが自動的にマウントされるようには設定されない

これ以外の構成エラーが検出される場合、ノード上でクラスター・サービスは開始されません。詳細なエラー・メッセージを参照して不整合を解消してください。

関連情報:

PowerHA SystemMirror クラスターの検査および同期化

予定外のダウン時間を最小限に抑える

PowerHA SystemMirror のもう 1 つの重要な目的は、計画外のクラスター・コンポーネント障害に対処するための予定外のダウン時間を最小限に抑えることです。

PowerHA SystemMirror ソフトウェアには、予定外のダウン時間を最小限に抑えるために以下の機能が用意されています。

- 大規模クラスターでのフォールオーバーを高速化する高速回復。
- 指定の時刻にカスタム・リソース・グループがフォールバックできるようにする遅延フォールバック・タイマー。
- サービス IP ラベルの回復時の処理を高速化する IP エイリアスによる IPAT。
- エラー状態にあるリソース・グループの自動回復 (クラスター・ノードの起動ごと)。詳しくは、次のセクションを参照してください。

ノード始動時のリソース・グループの回復

クラスター・マネージャーは、結合ノードで、現在エラー状態にあるリソース・グループをオンライン (アクティブ) 状態にしようとします。これはさらにアプリケーションをオンラインにする機会を増大させます。ノードが始動したときに、クラスター内のいずれかのノード上でリソース・グループがエラー状態にある場合、このノードはリソース・グループを得ようとします。

ノードは、リソース・グループ用のノード・リストに含まれていなければならないことに注意してください。

ノード始動時のリソース・グループ・リカバリーは、コンカレント・リソース・グループと非コンカレント・リソース・グループとは異なります。

- 開始ノードが、エラー状態にある非コンカレント・リソース・グループをアクティブにできない場合、使用可能なノードがあれば、そのリソース・グループは引き続きノード・リスト内の別のノードにフォールオーバーします。フォールオーバー処置は、ノード・リスト内のすべての利用できるノードが試されるまで継続します。
- 開始ノードが、ノード上でエラー状態にあるコンカレント・リソース・グループをアクティブにできない場合、そのコンカレント・リソース・グループはそのノード上でエラー状態のままです。リソース・グループは他のノード上で依然としてオンラインである可能性があることに注意してください。

高速リカバリー

PowerHA SystemMirror の高速リカバリー機能は、大規模なクラスターにおけるフォールオーバーをスピードアップします。

高速回復機能では、ファイルシステム整合性検査およびファイルシステム回復方法を選択できます。

- ファイルシステムが整合性チェックとリカバリー方法を使用するように構成すると、各ファイルシステム上で `fsck` ではなく、`logredo` を実行することで時間を節約します。後続の `mount` が失敗すると、`fsck` が実行されます。

ファイルシステムが障害で損傷を受けたものの、依然としてマウントできる場合は、`logredo` は、データ・アクセスの間の損傷の修正やエラーの生成に成功しない場合があります。

- また、取得、解放、およびフォールオーバーという作業が、逐次的に行われるのではなく、すべてのリソース・グループおよびファイルシステムで並行して行われることになるため、時間の節約になります。

ネストされた共有ファイルシステムがある場合には、これらのコマンドが並行して実行されるようにシステムを設定しないでください。そのようなファイルシステムは、順番に回復する必要があります(クラスター検証ユーティリティは、ファイルシステムおよび高速リカバリーの不整合を報告しません)。`varyonvg` コマンドおよび `varyoffvg` コマンドは、リカバリー方式の設定とは無関係に、必ずボリューム・グループを並行して実行します。

リソース・グループの遅延フォールバック・タイマー

遅延フォールバック・タイマーによって、リソース・グループは、より高い優先順位のノードに指定時刻にフォールバックできます。

遅延フォールバック・タイマーが構成されている場合、ホームでないノード上に現在常駐しているリソース・グループは、指定の頻度(日次、週次、月次、または毎年)でまたは指定日時に、より優先順位の高いノードにフォールバックします。

関連情報:

PowerHA SystemMirror プランニング・ガイド

テークオーバー時間を最小限に抑える

クラスター障害の場合、改良されたディスク・テークオーバー機構のために、拡張コンカレント・ボリューム・グループが、PowerHA SystemMirror の以前のリリースより素早くテークオーバーされます。

クラスター障害の場合、改良されたディスク・テークオーバー機構のために、拡張コンカレント・ボリューム・グループが、PowerHA SystemMirror の以前のリリースより素早くテークオーバーされます。非コンカレント・リソース・グループにおいて拡張コンカレント・モード・ボリューム・グループが含まれている

場合 (または既存のボリューム・グループが拡張コンカレント・ボリューム・グループに変換されている場合)、PowerHA SystemMirror は、拡張コンカレント・ボリューム・グループを自動的に検出し、ノード障害時にボリューム・グループ・テークオーバーのより高速なオプションが起動されるようにします。

この機能は、多数のディスクで構成されるボリューム・グループのフォールオーバーの際に特に役立ちます。

高速ディスク・テークオーバーの際には、PowerHA SystemMirror は、ディスク予約を解除するため、または遅延更新の実行によって LVM 情報を更新および同期化するために必要となる、余分な処理をスキップします。このため、拡張コンカレント・ボリューム・グループに使用される PowerHA SystemMirror のディスク・テークオーバー機構は、非コンカレント・リソース・グループに含まれている標準のボリューム・グループに使用されるディスク・テークオーバーと比べて高速です。

災害時回復を最大限に行う

PowerHA SystemMirror を企業の総合的災害復旧計画の必要不可欠な部分とすることができます。

災害復旧活動用として別々のサイトにデータのバックアップ・コピーを配布する方法には、次のようにいくつかのものが考えられます。

- PowerHA SystemMirror Enterprise Edition for Geographic Logical Volume Manager (GLVM)
- PowerHA SystemMirror Enterprise Edition for Metro Mirror (ESS および DS システムでの同期 PPRC)
- サイト間 LVM ミラーリング

サイト間 LVM ミラーリング

例えば、ストレージ・エリア・ネットワーク (SAN) を使用して、リモート LVM ミラーのために 2 つの異なるサイトにディスクをセットアップすることができます。サイト間 LVM ミラーリングは、災害時リカバリー用の各サイトのディスク・サブシステムの間でデータを複製します。

SAN は、ファイバー・チャネルでサポートされる距離内のストレージ・デバイスとプロセッサ (サーバー) の間の直接的な接続を確立できる高速ネットワークです。したがって、共通の SAN を介して、別々のサイトに存在する 2 つ以上のサーバー (ノード) が同じ物理ディスクにアクセスでき、それらは距離的に離れていてもかまいません。ディスクは、AIX 論理ボリューム・マネージャーを使用してボリューム・グループに結合でき、このボリューム・グループを異なるサイトにあるノードにインポートできます。このボリューム・グループ内の論理ボリュームには最大 3 つのミラーを備えることができます。そのため、各サイトに少なくとも 1 つのミラーをセットアップできます。この論理ボリュームに格納された情報は可用性が高く、特定の障害の場合、別のサイトのリモート・ミラーが依然として最新の情報を持っているので、操作を別のサイトで継続できます。PowerHA SystemMirror は、ディスクまたはノード障害、および後続の再統合のあとで、自動的にミラーを同期化します。

PowerHA SystemMirror は、ディスクの 1 つが PVREMOVED または PVMISSING 状態にある場合でも、自動ミラー同期化を処理します。自動同期化はすべての場合に可能とは限りませんが、ディスクまたはサイト障害、および後続の再統合の後で、C-SPOC を使用して、残存ミラーから失効したミラーにデータを手動で同期化できます。

クラスター・イベント

このセクションでは、PowerHA SystemMirror ソフトウェアがクラスター内の変化に応じて高可用性を維持する方法について説明します。

PowerHA SystemMirror クラスター・ソフトウェアは、高可用性アプリケーションを構成するすべてのコンポーネント (ディスク、ネットワーク・インターフェース、ノード、およびアプリケーション自体を含む) をモニターします。モニターする際に、クラスター・マネージャーはリソースに応じたモニター方法を使用します。

- RSCT サブシステムは、ネットワークとノードをモニターします。
- AIX LVM サブシステムは、ボリューム・グループのクォーラムの損失に関するエラー通知を生成します。
- クラスター・マネージャー自体はアプリケーション・モニターをディスパッチします。

PowerHA SystemMirror クラスター環境はイベント・ドリブンです。イベントとは、クラスター・マネージャーが認識し、処理するクラスター内の状況の変化です。クラスター・イベントは、ネットワーク・インターフェース・カード、ネットワーク、またはノードに影響する変化によって引き起こされる場合と、時間制限を超えたクラスター再構成プロセスによって引き起こされる場合があります。クラスター・マネージャーは、クラスター状況の変化を検出すると、そのイベントおよびサブイベントに対応するスクリプトを実行します。

注: クラスター・イベントの論理についてここで完全に説明します。PowerHA SystemMirror がクラスター・イベントを正しく実行するように、手動による関与は不要です。

以下では、クラスター・マネージャーが認識するイベントの例を示します。

- `node_up` イベントと `node_up_complete` イベント (クラスターと結合するノード)
- `node_down` イベントと `node_down_complete` イベント (クラスターから離れるノード)
- ローカルまたはグローバルの `network_down` イベント (ネットワーク障害が発生しました)
- `network_up` イベント (ネットワークが接続しました)
- `swap_adapter` イベント (ネットワーク・アダプターが失敗し、新しいものが取って代わりました)
- 動的再構成イベント
- `site_up` と `site_up_complete` (クラスターと結合するサイト)
- `site_down` と `site_down_complete` (クラスターから離れるサイト)

クラスター・イベントが発生すると、クラスター・マネージャーはそのイベントに対応するイベント・スクリプトを実行します。イベント・スクリプトの処理中に一連のサブイベント・スクリプトが実行されることがあります。PowerHA SystemMirror ソフトウェアは、各イベントおよびサブイベント用のスクリプトを提供します。デフォルトのスクリプトは `/usr/es/sbin/cluster/events` ディレクトリーにあります。

デフォルトでは、クラスター・マネージャーは、特定イベントに対して、PowerHA SystemMirror ソフトウェア付属の、そのイベントに対応するイベント・スクリプトを呼び出します。

必要であれば、追加処理情報を指定して自分のサイトのイベント処理をカスタマイズできます。

クラスター・イベントの処理

PowerHA SystemMirror ソフトウェアが扱う主要なクラスター・イベントは、フォールオーバーと再統合の 2 つです。

- フォールオーバーとは、クラスター・コンポーネントで障害が発生したとき、またはノードがクラスターを離れたときに、PowerHA SystemMirror ソフトウェアによって行われるアクションのことです。
- 再統合とは、以前にクラスターを離れたコンポーネントがクラスターに戻るときに、クラスター内で行われるアクションのことです。

イベント・スクリプトは両方のタイプの処置を制御します。イベント・スクリプトの処理中は、クラスター対応のアプリケーション・プログラムはクラスターの状態を不安定と見なします。

フォールオーバー

フォールオーバーは、ホーム・ノードがクラスターから切り離されたためにリソース・グループがホーム・ノードから別のノードに移動したときに行われます。

ノードがクラスターから離れるのは、計画された変更処理（ノードのシャットダウンまたはノード上のクラスター・サービスの停止）によるか、または障害によるかのどちらかです。計画された変更処理の場合、クラスター・マネージャーは、既存のノードが保有しているリソースの解放と、クラスター内でアクティブのままになっているノードによるこれらのリソースの取得を制御します。必要に応じて、ユーザーは（例えば、システム保守を行うために）リソースの解放と取得をオーバーライドすることができます。（ユーザー定義リソース・グループに対する遅延フォールバック・タイマーを設定して）ノードを組み込むことにより、リソースの取得を延期することもできます。

ノード障害は、隣接するノードを監視するノードが、定義された一定期間キープアライブ・トラフィックの受信を停止するときに開始します。この障害がノード障害であることをほかのクラスター・ノードが認めると、障害ノードはクラスターから除去され、そのノードのリソースはテークオーバー用に構成されているアクティブ・ノードによって引き継がれます。

その他のコンポーネント（ネットワーク・インターフェース・カードなど）に障害が発生した場合、クラスター・マネージャーは、イベント・スクリプトを実行して、バックアップ・ネットワーク・インターフェース・カードが存在する場合は、このバックアップ NIC にネットワーク・トラフィックを切り替えます。

再統合

再統合、すなわちフォールバックが行われるのは、リソース・グループが、クラスターに結合したばかりのノードに移動したときです。実行中のクラスターにノードが結合すると、そのクラスターは一時的に不安定になります。メンバー・ノードは結合プロセスの始めを同調させてから、イベント・スクリプトを実行して、結合ノードが引き継ぐように構成されているリソースを解放します。次に、結合ノードがイベント・スクリプトを実行してこれらのリソースを引き継ぎます。最後に、結合ノードはクラスターのメンバーになります。この時点でクラスターは再び安定になります。

イベント処理のカスタマイズ

PowerHA SystemMirror ソフトウェアにはイベント・カスタマイズ機能があり、ユーザーはこの機能を使用してイベント処理を調整することができます。

クラスター・マネージャーには特定のシリーズのイベントおよびサブイベントを認識する能力があるため、非常に柔軟なカスタマイズを行うことができます。イベント処理をカスタマイズすることにより、どのリソースに障害が発生したのかを確実に判別することができます。

「**Change/Show Pre-Defined PowerHA SystemMirror Events** (事前定義済み HACMP イベントの変更/表示)」SMIT パネルまたは SMIT 高速パス `clcsclev.select` を使用して、ピック・リストに表示されるイベントのリストに対して複数のイベント前処理とイベント後処理を定義できます。1 つのイベントに関するカスタマイズには、次のリストに示すように、イベントの処理前および処理後のシステム管理者への通知と、イベント処理前および処理後のユーザー定義のコマンドまたはスクリプトが含まれます。

- 処理すべきイベントについてのシステム管理者への通知
- イベント前処理スクリプトまたはコマンド
- PowerHA SystemMirror for AIX イベント・スクリプト

- イベント後処理スクリプトまたはコマンド
- イベント処理が完了したことについてのシステム管理者への通知

次のようなタイプのカスタマイズには、この機能を使用します。

- イベント前処理とイベント後処理
- イベント通知
- イベント・リカバリーと再試行

注: PowerHA SystemMirror では、PowerHA SystemMirror 構成データベースに格納されているイベント・カスタマイズ情報は、クラスター・リソースが同期化されると、すべてのクラスター・ノードにおいて同期化されます。したがって、イベント前処理、イベント後処理、通知、およびリカバリーのイベント・スクリプトは、実際に行う処理が異なっても構いませんが、どのノードにおいても同じ名前になっている必要があります。

クラスターの検証には、**cluster_notify** という名前の新しいイベントを使用して、クラスター構成を自動的にモニターする機能が組み込まれています。このイベントを使用して、PowerHA SystemMirror のリモート通知メソッド (数字または英数字のページ、またはテキスト・メッセージ) を構成し、クラスター構成内にエラーが検出された場合にメッセージを送信できます。また、このイベントの出力は、クラスター・サービスを実行している各クラスター・ノード上の **hacmp.out** にログとして記録されます。

イベント通知スクリプトによって、E メール通知を携帯電話に送信することもできます。ただし、リモート通知を使用する利点がいくつかあります。イベント通知の変更に対応する担当者は、各イベント通知スクリプト内に記述されたアドレスを手作業で変更する必要があります。個々のユーザーごとに、すべてのイベントとノードを含むリモート通知メソッドを定義できるので、応答者が変わっても 1 ユニットとして通知方法を切り替えることができます。

新しいイベントの定義

PowerHA SystemMirror では、既存のイベントをカスタマイズするだけでなく、新規イベントを定義することもできます。

イベント前処理とイベント後処理

ご使用の環境に合わせてイベント処理をカスタマイズするには、クラスター・マネージャーにより、そのイベントが実行される前後に実行されるコマンドまたはユーザー定義スクリプトを指定します。例えば、前処理の場合、特定ユーザーにメッセージを送信して、イベントの発生中は待機するよう通知する場合があります。後処理の場合では、特定のネットワークに障害が発生したときに、特定グループのユーザーに対してログインを使用不能にする場合があります。

イベント通知

あるコマンドまたはユーザー定義スクリプトを指定することにより、イベントが発生しようとしていることを、さらにイベントがたった今発生したことをその実行が正常であったか、あるいは異常であったかという情報と共に (メールなどで) 通知することができます。クラスター・イベントに応じてカスタマイズしたりリモート通知メソッドを発行するために、SMIT インターフェースによって **notification** メソッドを定義することもできます。

イベント・リカバリーと再試行

イベント・コマンドの失敗から回復しようとするコマンドを指定できます。再試行カウントがゼロより大きい場合に、リカバリー・コマンドの実行が成功すると、そのイベント・スクリプト・コマンドは再実行され

ます。リカバリー・コマンドの実行を試みる回数も指定できます。

イベント期間のカスタマイズ

PowerHA SystemMirror ソフトウェアは、クラスター・イベントの完了が指定したタイムアウト期間を上回るたびにシステム警告を出します。

SMIT インターフェースを使用すると、PowerHA SystemMirror がシステム警告を出すまでに、クラスター・イベントが完了できる期間をカスタマイズできます。

PowerHA SystemMirror クラスター構成

この章では、PowerHA SystemMirror ソフトウェアがサポートする各種のサンプル・クラスター構成を提供します。

このリストは、PowerHA SystemMirror ソフトウェアを使用して定義することが可能な構成をすべて網羅したものではありません。お客様の環境に最も適したクラスター構成を考える際の出発点としてご利用ください。

スタンバイ構成

スタンバイ構成は、従来の冗長ハードウェア構成です。この構成では、1 つ以上のスタンバイ・ノードがアイドル状態のまま、サーバー・ノードがクラスターを離れるまで待機します。

コンカレント・リソース・グループは、すべてのノードが同時にリソース・グループにアクセスするものであり、スタンバイ構成では使用できません。

例: ホーム・ノードのみでオンラインの始動ポリシーを使用したスタンバイ構成

この例は、「ホーム・ノードのみでオンライン」始動ポリシー、「リスト内の次の優先順位ノードにフォールオーバー」フォールオーバー・ポリシー、および「リスト内の優先順位が上位のノードへフォールバック」フォールバック・ポリシーを持つリソース・グループを示しています。

次のスタンバイ構成のリソース・グループは以下のポリシーを持っています。

- 始動ポリシー: Online on Home Node Only (ホーム・ノードでのみオンライン)
- フォールオーバー・ポリシー: Fallover to Next Priority Node in the List (リスト内の次の優先順位ノードへフォールオーバー)
- フォールバック・ポリシー: Fallback to Higher Priority Node in the List (リスト中のより高い優先順位のノードにフォールバック)

図中では、番号が小さくなるに従って、優先順位が高くなっています。

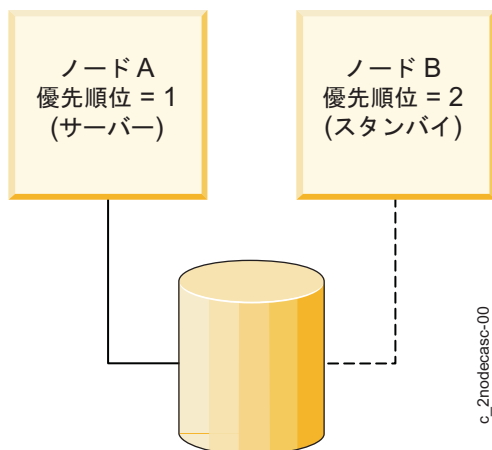


図 21. IP ラベルがホーム・ノードに返される 1 対 1 スタンバイ構成

このセットアップでは、クラスター・リソースは単一のリソース・グループの一部として定義されています。次に、2 つのノードからなるノード・リストが定義されています。最初のノードであるノード A には、1 のテークオーバー (所有権) 優先順位が割り当てられています。2 番目のノードであるノード B には、2 のテークオーバー優先順位が割り当てられています。

クラスターの始動時には、ノード A (1 の優先順位をもつ) がリソース・グループの所有権を取得しています。ノード A は「サーバー」ノードです。ノード B (2 の優先順位をもつ) はアイドル状態にあり、ノード A に障害が発生した場合、またはクラスターを離れた場合に作動可能になります。ノード B は実際上「スタンバイ」ノードです。

サーバー・ノードがクラスターを離れた場合、スタンバイ・ノードはサーバーが所有しているリソース・グループの制御を引き継ぎ、高可用性アプリケーションを始動し、クライアントにサービスを提供します。スタンバイ・ノードは、より高いテークオーバー優先順位をもつノードがクラスターに再結合するまでアクティブのままです。再結合の時点で、スタンバイ・ノードは引き継いだリソース・グループを解放し、サーバー・ノードが再利用します。次にスタンバイ・ノードはアイドル状態に戻ります。

拡張されたスタンバイ構成

前述の例にあるスタンバイ構成をさらに大きなクラスターに拡張することは簡単にできます。この構成の長所は、ハードウェアをより有効に使用できることです。短所は、複数のサーバー・ノードがクラスターを離れた場合にクラスターの性能が大幅に低下することです。

次の図は、以下のポリシーを持つリソース・グループを使用する 3 ノードのスタンバイ構成を表しています。

- 始動ポリシー: Online on Home Node Only (ホーム・ノードでのみオンライン)
- フォールオーバー・ポリシー: Fallover to Next Priority Node in the List (リスト内の次の優先順位ノードへフォールオーバー)
- フォールバック・ポリシー: Fallback to Higher Priority Node in the List (リスト中のより高い優先順位のノードにフォールバック)

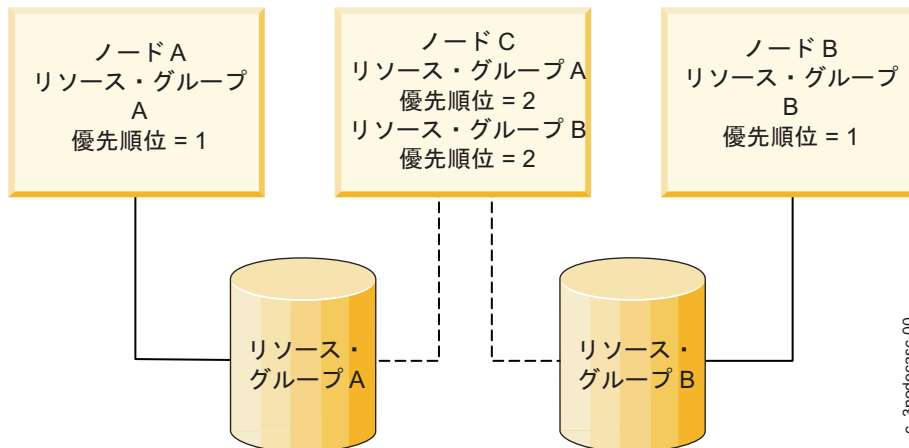


図 22. 3 つのリソース・グループを含む 2 対 1 スタンバイ構成

この構成には、2 つの別個のリソース・グループ (A と B) があり、さらに各リソース・グループにはそれぞれノード・リストが 1 つ存在します。リソース・グループ A のノード・リストは、ノード A とノード C からなります。ノード A のテークオーバー優先順位は 1 であり、ノード C のテークオーバー優先順位は 2 です。リソース・グループ B のノード・リストは、ノード B とノード C からなります。ノード B のテークオーバー優先順位は 1 であり、ノード C のテークオーバー優先順位は 2 です (非コンカレント構成では、1 つのノードが同時に複数のリソース・グループを所有することはできません)。

各リソース・グループのノード・リストの先頭にあるノードは異なっているため、クラスターの作業負荷はこれら 2 つのリソース・グループの間で分割されます。しかし、両方のリソース・グループのノード・リスト内のスタンバイ・ノードは同じです。いずれかのサーバー・ノードがクラスターを離れた場合、スタンバイ・ノードはそのサーバー・ノードのリソース・グループの制御を引き継ぎ、残留ノードとして機能します。

この例では、スタンバイ・ノードには 3 つのネットワーク・インターフェース (図には示されていない) および各サーバー・ノードの外部ディスクとの個別の物理接続があります。したがって、スタンバイ・ノードは、必要ならば両方のサーバー・ノードを並行して引き継ぐことができます。しかし、このスタンバイ・ノードは両方のサーバー・ノードとして機能するため、クラスターの性能はどうしても低下します。

例: 配布ポリシーを使用してオンラインの始動ポリシーを使用したスタンバイ構成

この例では、スタンバイ構成で使用する必要のあるリソース・グループ・ポリシーについて説明します。

次のスタンバイ構成のリソース・グループは以下のポリシーを持っています。

- 始動ポリシー: Online Using Distribution Policy (配布ポリシーを使用してオンライン) (ネットワーク・ベース、またはノード・ベース)
- フォールオーバー・ポリシー: Next Priority Node in the List (リスト中の次の優先順位のノード)
- フォールバック・ポリシー: Never Fallback (フォールバックしない)

この構成は、リソース・グループの所有者の優先順位が一定ではないスタンバイ構成とは異なります。リソース・グループは、ノード間を回ることができる IP アドレスに関連づけられます。このため、サーバーとスタンバイのロールは流動的であり、時と共に変わります。

次の図は、このセクションで説明している 1 対 1 スタンバイ構成を表しています。

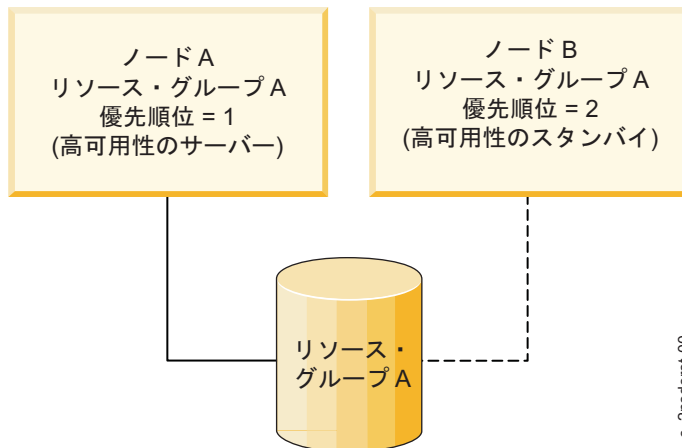


図 23. IP ラベルが循環し、リソース・グループが含まれる 1 対 1 スタンバイ構成

システム始動時には、リソース・グループは共用 IP アドレスを請求するノードに接続されています。このノードは、クラスターにとどまっている限り、このリソース・グループを所有します。このノードがクラスターを離れると、ピア・ノードが共用 IP アドレスを引き継ぎ、該当リソース・グループの所有権を請求します。今度は、ピア・ノードが、クラスターにとどまっている限り、このリソース・グループを「所有」します。

最初にリソース・グループを請求していたノードは、クラスターに再結合してもリソース・グループの所有権を取り戻しません。このノードは、共用 IP アドレスに現在バインドされているノードがクラスター内でアクティブである限り、アイドル状態のままです。ピア・ノードがクラスターを離れた場合にのみ、最初にリソース・グループを「所有」していたノードは再請求します。このように、リソースの所有権はノード間を回ります。

拡張されたスタンバイ構成

スタンバイ構成の最初の例と同様に、この例にある構成はより大きいクラスターに簡単に拡張できます。例えば、この例の 2 対 1 のスタンバイ構成では、クラスターはそれぞれ別々の共用 IP アドレスを含んだ 2 つのリソース・グループを持つことができます。

クラスター始動時には、最初の 2 つのノードがそれぞれ共用 IP アドレスを請求し、その共用 IP アドレスに関連づけられたリソース・グループの所有権を取得します。3 番目のノードはアイドル状態にあります。アクティブ・ノードの 1 つがクラスターを離れた場合、アイドル・ノードが共用 IP アドレスを請求し、リソース・グループの制御を取得します。

テークオーバー構成

テークオーバー構成では、すべてのクラスター・ノードが有用な役割を担っていて、クラスターのワークロードの一部を処理します。スタンバイ・ノードはありません。テークオーバー構成では、アイドル・プロセスがないため、スタンバイ構成の場合よりハードウェア・リソースが有効に使用されます。しかし、ノードの切り離しが発生すると、残りのノードの負荷が増大するため、性能が低下する場合があります。

ワンサイド・テークオーバー

この構成には作業を処理するアクティブ・ノードが 2 つありますが、クラスター・クライアントに高可用性サービスを提供するノードは 1 つだけです。つまり、クラスター内にリソースのセットが 2 つ (例えばクライアント要求を処理する 2 つのサーバー・アプリケーション) あっても、そのうちの 1 つしか高可用性にする必要はありません。

次の図は、2 ノード・ワンサイド・テークオーバー構成を説明したものです。図中では、番号が小さくなるに従って、優先順位が高くなっています。

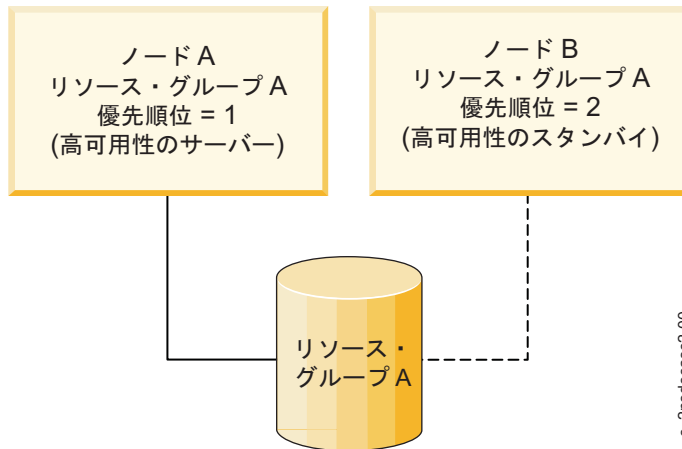


図 24. IP ラベルがホーム・ノードに返されるリソース・グループを持つワンサイド・テークオーバー構成

このセットのリソースは PowerHA SystemMirror リソース・グループとして定義され、両方のノードを含むノード・リストが 1 つあります。2 番目のセットのリソースは、リソース・グループとしては定義されておらず、したがって高可用性ではありません。

クラスター始動時には、ノード A (1 の優先順位をもつ) がリソース・グループ A の所有権を取得します。ノード A は、実際上、リソース・グループ A を「所有」します。ノード B (リソース・グループ A については 2 の優先順位をもつ) は、このリソース・グループとは無関係に、それ自体の作業負荷を処理します。

ノード A がクラスターを離れた場合、ノード B が共用リソースの制御を取得します。ノード A がクラスターに再結合すると、ノード B は共用リソースを解放します。

しかし、ノード B がクラスターから離脱した場合は、ノード A はノード B のリソースを 1 つも引き継ぎません。ノード B のリソースはノード A がチェーンに参加している高可用性リソース・グループの一部として定義されていないからです。

クラスター・クライアントにとって可用性を高くする必要のあるすべてのクリティカル・アプリケーションを単一ノードで実行できる場合は、この構成が適しています。

相互テークオーバー

非コンカレント・アクセス用相互テークオーバーの構成では、複数のノードがあり、それぞれがクラスター・クライアントに別々に高可用性サービスを提供します。例えば、各ノードが独自のデータベース・インスタンスを実行し、独自のディスクにアクセスするという構成が考えられます。

さらに、各ノードにはテークオーバー容量があります。1 つのノードがクラスターを離れた場合、残留ノードが退去ノードの所有していたリソース・グループを引き継ぎます。

高可用性が必要なクリティカル・アプリケーションをクラスター内の各ノードが実行する場合と、プロセッサが複数のノードの負荷を処理できる場合は、非コンカレント・アクセス用相互テークオーバーの構成が適しています。

次の図は、2 ノードの非コンカレント・アクセス用相互テークオーバー構成を説明したものです。図中では、番号が小さくなるに従って、優先順位が高くなっています。

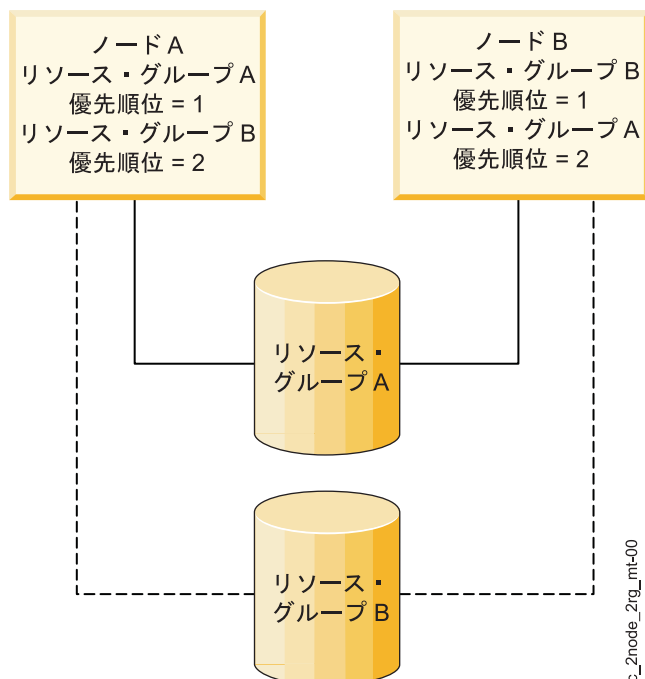


図 25. 非コンカレント・アクセス用相互テークオーバー構成

この構成の主な特徴は、クラスターの作業負荷がノード間で分割される、つまり区分化されることです。2つのリソース・グループが別々に存在し、それぞれのリソース・グループに別個のリソース・チェーンがあります。両方のリソース・チェーンに参加するノードは同じです。この構成が相互テークオーバーとなるのは、チェーン内の優先順位が異なるからです。

両方のリソース・グループのチェーンは、ノード A およびノード B からなります。リソース・グループ A の場合、ノード A には 1 のテークオーバー優先順位が割り当てられ、ノード B には 2 のテークオーバー優先順位が割り当てられています。リソース・グループ B では、テークオーバー優先順位が逆になっています。この場合、ノード B のテークオーバー優先順位は 1 であり、ノード A のテークオーバー優先順位は 2 です。

クラスター始動時には、ノード A はリソース・グループ A の所有権を取得し、ノード B はリソース・グループ B の所有権を取得します。

どちらかのノードがクラスターを離れた場合、そのピア・ノードが退去ノードのリソース・グループを制御します。このリソース・グループの「所有者」ノードがクラスターに再結合されると、テークオーバー・ノードが関連リソースを解放し、それよりも優先順位の高い再結合ノードがそのリソースを再取得します。

2 ノードの相互テークオーバー構成

この構成では、両方のノードが共用ディスクに同時にアクセスし、同じディスク・リソースを所有します。ノードの1つがクラスターを離れても、ピア・ノードでは既に共用ボリューム・グループがオンに変更されているため、共用ディスクのテークオーバーは起こりません。

次の図は、2 ノードのコンカレント・アクセス用相互テークオーバー構成を説明したものです。

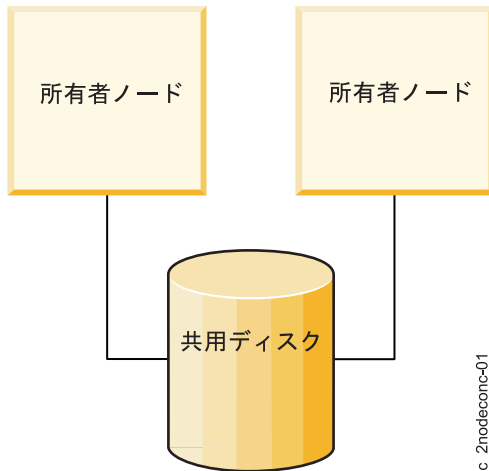


図 26. 2 ノードのコンカレント・アクセス用相互テークオーバー構成

この例では、両方のノードが共用ディスク上のデータベースにアクセスするサーバー・アプリケーションの1つのインスタンスを実行します。ディスク・リソースに対するアプリケーションの要求を調停するために、このアプリケーション専用ロック・モデルが使用されます。

同じサーバー・アプリケーションの複数のインスタンスを実行することにより、処理負荷を分散させることができます。負荷が増大するに従ってノードを追加すれば、負荷をさらに分散させることができます。

8 ノードの相互テークオーバー構成

この構成では、すべてのノードが共用ディスクに同時にアクセスし、同じディスク・リソースを所有します。

次の図は、8 ノードのコンカレント・アクセス用相互テークオーバー構成を説明したものです。

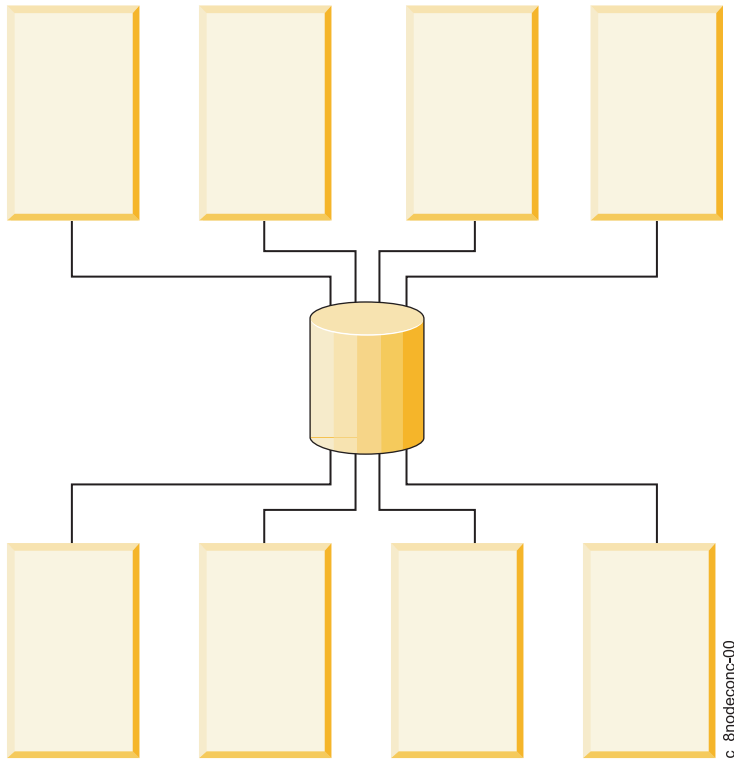


図 27. 8 ノードの相互テークオーバー

各ノードは別々のサーバー・アプリケーションを実行しています。クライアントは、特定の IP アドレスで特定のアプリケーションを照会します。したがって、各アプリケーション・コントローラーおよびそれに関連した IP アドレスを非コンカレント・リソース・グループの一部として定義する必要があります。また、このリソース・グループの潜在的所有者であるすべてのノードを、対応するノード・リストに含める必要があります。

PowerHA SystemMirror では、コンカレント・アクセス・リソース・グループは、最大 16 ノードまでのクラスターでサポートされます。

多層アプリケーションを使用したクラスター構成

親/子依存関係のリソース・グループを利用できる典型的なクラスター構成は、WebSphere のようなアプリケーションが DB2 のような別のアプリケーションに依存している環境です。

ビジネス要件を満たすには、2 つ以上のリソース・グループの間で、クラスター全体にわたる親/子依存関係を定義する必要があります。

次の図は、複数のアプリケーション間に存在する依存関係を利用するビジネス・シナリオを表しています。

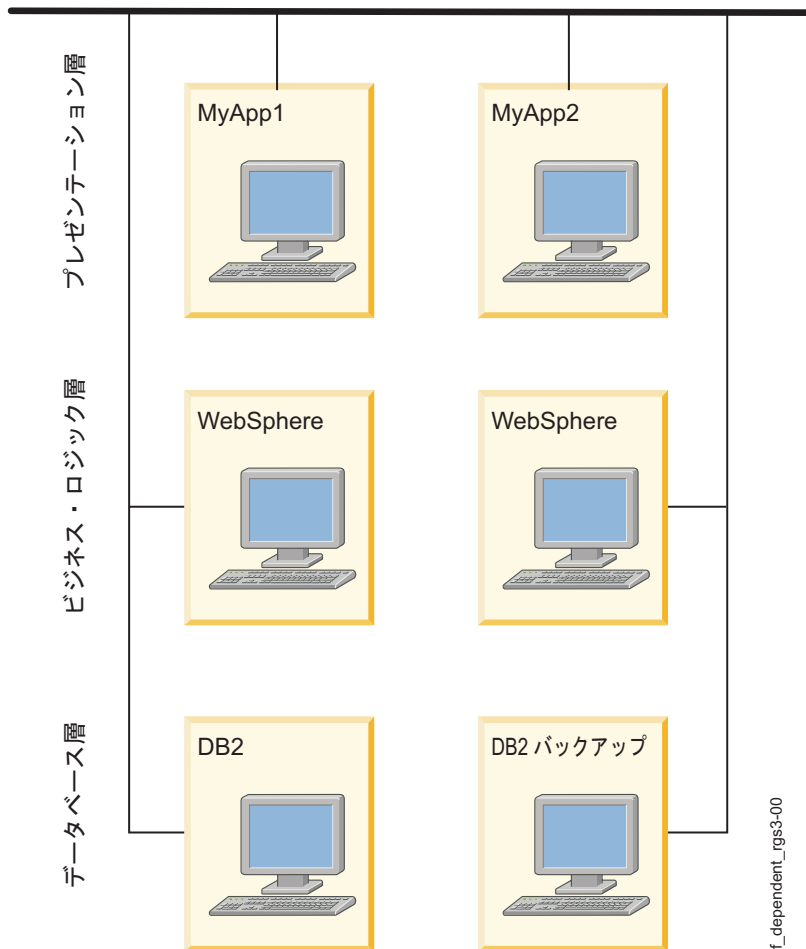


図 28. アプリケーション間の依存関係を持つ標準的な多層クラスター環境

多層アプリケーション

階層化されたアプリケーション、つまり多層アプリケーションを使用するビジネス構成では、依存リソース・グループを使用することもできます。例えば、バックエンド・データベースはアプリケーション・コントローラーよりも先にオンラインにする必要があります。この場合、このデータベースが動作を停止し、別のノードに移動した場合、アプリケーション・コントローラーを含むリソース・グループを停止し、クラスター内の任意のノードでバックアップする必要があります。

SAP のような環境では、データベースに障害が発生した場合に必ずアプリケーションを停止して再起動する必要があります。SAP のような環境は多数のアプリケーション・サービスを提供し、多くの場合、個々のアプリケーション・コンポーネントを特定の順序で制御する必要があります。

また、アプリケーション環境のサポートにシステム・サービスが必要である場合も、リソース・グループの間で相互依存関係を確立すると便利であることが分かっています。ログ・ファイルから余分なものを取り除いたり、バックアップを開始したりするための **cron** ジョブなどのサービスは、アプリケーションと共にノードからノードへ移動する必要がありますが、通常、アプリケーションが確立されるまで開始されません。これらのサービスは、アプリケーション・コントローラーの開始スクリプトや停止スクリプトに組み込むことができます。よりきめ細かな処理が必要とされる場合は、イベント前処理、およびイベント後処理を通じてこれらを制御することができます。親/子依存関係のリソース・グループを使うと、システム・サービスが、サービスの提供先となるアプリケーションに依存するように簡単に構成することができます。

関連概念:

50 ページの『リソース・グループ依存関係』

PowerHA SystemMirror では、リソース・グループの順序付けおよびリソースのカスタマイズされたシリアル処理をサポートし、異なる複数のリソース・グループに常駐するアプリケーション間に依存関係が存在するクラスター構成にも対応しています。

リソース・グループ・ロケーション依存関係を持つクラスター構成

特定のアプリケーションが始動時だけでなく、フォールオーバーおよびフォールバックのイベント発生時にも、同じノード上または異なるノード上にとどまるようにクラスターを構成できます。このためには、選択したリソース・グループをロケーション依存関係のセットの一部として構成します。

同じノードおよび異なるノードの依存関係を使用する出版社のモデル

架空の会社、XYZ 出版社では、Web コンテンツの開発に使用するさまざまなプラットフォームを分離するビジネス継続性モデルを採用しています。XYZ 出版社はロケーション依存関係ポリシーを使用して、一部のリソース・グループは厳密に別々のノード上に置き、他のリソース・グループは一緒に同じノード上に置いています。

実動データベース (PDB) および実動アプリケーション (PApp) は、保守がしやすいように、同じノード上で稼働しています (さらに、これらのリソース・グループ用のノードの優先順位は最高であり、メモリー容量も大きく、プロセッサも高速です)。また、アプリケーションはデータベースに依存するため、これら間に親/子依存関係を設定する意義があります。データベースは、アプリケーションが動作するためにオンラインである必要があります。同様の条件は、システム・データベース (SDB) とシステム・アプリケーション (Sapp)、QA データベース (QADB) と QA アプリケーション (QAapp) にも当てはまります。

実動データベースとアプリケーションの稼働状態を維持することは最高の優先順位を持つため、3 つのデータベース・リソース・グループが別々のノード上にとどまるようにクラスターを構成し (異なるノードでオンラインの依存関係のセットに設定)、PDB リソース・グループに高い優先順位を割り当てるのが合理的です。SDB の優先順位は中であり、QADB の優先順位は低です。

データベースと関連のアプリケーションは、同じノードでオンラインの依存関係のセットに属するようにそれぞれ構成します。

PowerHA SystemMirror では、始動、フォールオーバー、およびフォールバック・ポリシーの構成方法に応じて、これらのグループを若干異なる方法で取り扱います。参加ノード・リストをデータベースおよびアプリケーション・セットごとに変えると、特定のリソース・グループを優先度の高いノード上に保持しやすくなります。

以下の図は、3 つのノードと 6 つのリソース・グループからなる基本構成を表しています。

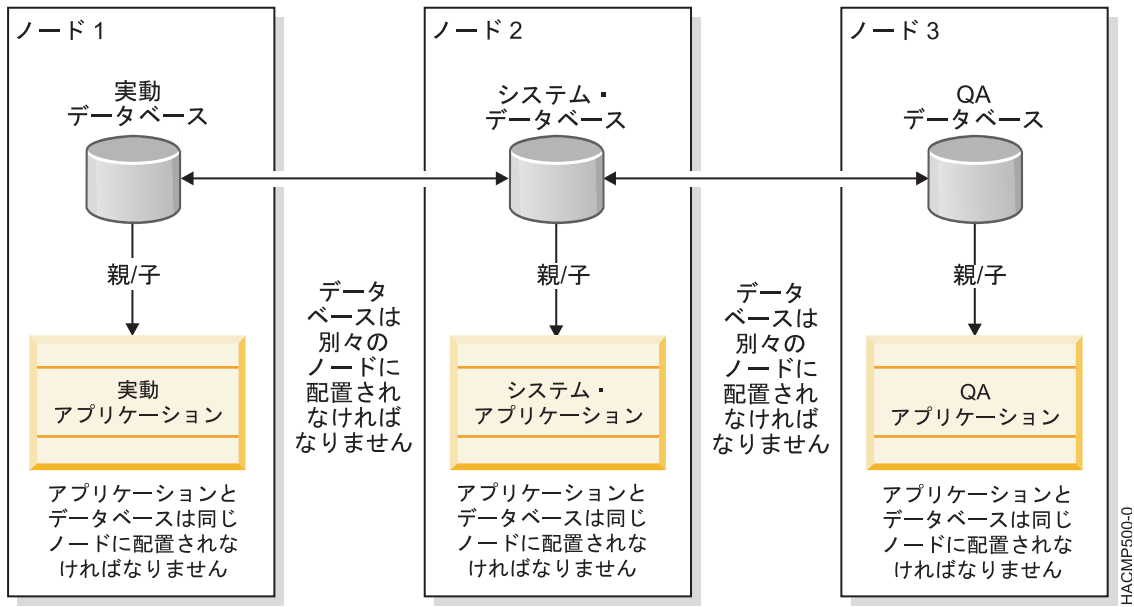


図 29. 親依存関係または子依存関係およびロケーション依存関係がある発行モデル

リソース・グループ・ポリシー

ここでは説明上の便宜のため、6 つのリソース・グループすべてが以下の動作のポリシーを持っています。

- 始動ポリシー: Online On First Available Node (最初の使用可能ノードでオンライン)
- フォールオーバー・ポリシー: Fallover to Next Priority Node (次に優先順位が高いノードにフォールオーバー)
- フォールバック・ポリシー: Never Fallback (フォールバックしない)

参加ノード	ロケーション依存関係	親/子依存関係
<ul style="list-style-type: none"> • PApp: 1、2、3 • PDB: 1、2、3 • SApp: 2、3 • SDB: 2、3 • QAApp: 3 • QADB: 3 	同じノードでオンラインの依存関係のグループ: <ul style="list-style-type: none"> • PApp と PDB • SApp と SDB • QAApp と QADB 異なるノードでオンラインの依存関係のセット: [PDB SDB QADB] 優先順位: PDB > SDB > QADB	<ul style="list-style-type: none"> • PApp (子) は PDB (親) に依存する • SApp (子) は SDB (親) に依存する • QAApp (子) は QADB (親) に依存する

関連情報:

クラスター・イベントでのリソース・グループの動作

災害時回復のためのサイト間 LVM ミラー構成

2 つの異なるサイトにあるディスクを、ストレージ・エリア・ネットワーク (SAN) を使用してリモート LVM ミラーリング用にセットアップすることができます。

SAN は、ファイバー・チャネルでサポートされる距離内のストレージ・デバイスとプロセッサ (サーバー) の間の直接的な接続を確立できる高速ネットワークです。したがって、共通の SAN を使用することにより、別々のサイトに存在する距離的に離れた 2 つ以上のサーバー (ノード) が同じ物理ディスクにアク

セスでき、それらは距離的に離れていてもかまいません。C-SPOC を使用して、このようなりモート・ディスクをボリューム・グループに結合することができます。

例えば、1 つのサイトに 1 つのミラーのように、ボリューム・グループ内の論理ボリュームは、最高 3 つのミラー、つまりコピーを持つことができます。したがって、この論理ボリュームに保管された情報では、障害の発生時にも高可用性が保持されます。例えば、1 つのサイトの (そのサイトのディスク・サブシステムも含めて) すべてのノードで障害が発生した場合でも、別のサイトにあるリモート・ミラーには最新の情報が保管されているので、そのサイトで操作を続けることができます。

この機能の主な目的は、SAN を通じた LVM ミラーリングにより、災害時回復のために各サイトにあるディスク・サブシステムの間でデータを複製する 2 サイト・クラスターをサポートすることにあります。

サイト間 LVM ミラーには、サイト/ディスク障害の発生後、サイトの再統合が行われた後で、PowerHA SystemMirror により、残されたディスクから、結合ディスクへのデータの同期が自動的に行われるという利点もあります。この同期はバックグラウンドで行われるので、再統合時間に大きな影響は与えません。

次の図は、SAN を使用したサイト間 LVM ミラー構成を表しています。

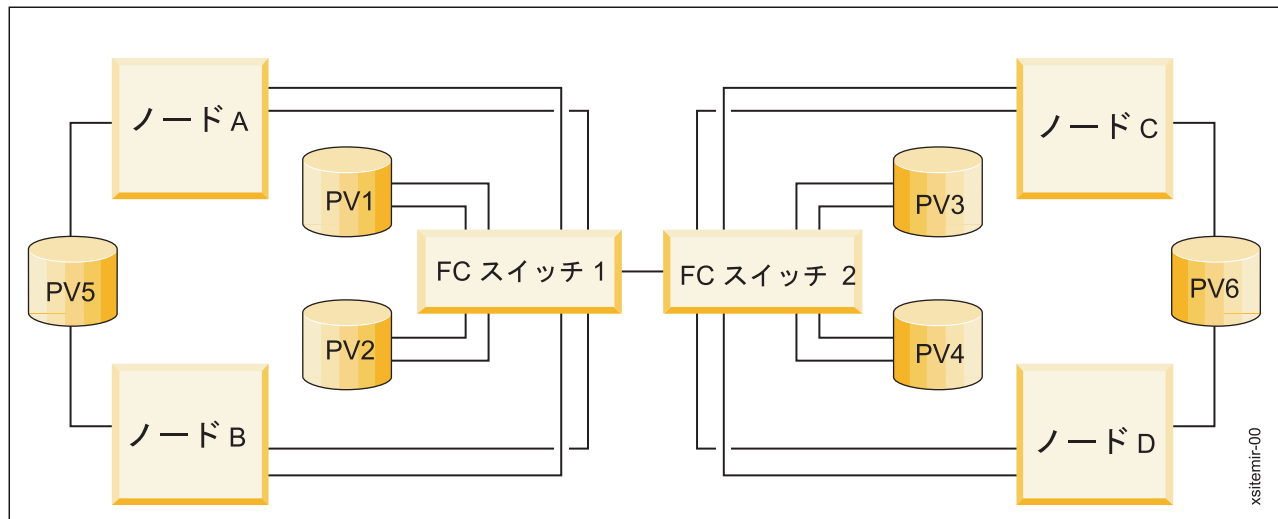


図 30. SAN を使用したサイト間 LVM ミラー構成

2 つのサイトそれぞれにある少なくとも 1 つのノードに接続されているディスクがミラーできます。この例では、PV4 は、ファイバー・チャンネル・スイッチ 1 およびファイバー・チャンネル・スイッチ 2 を使用してサイト 1 のノード A および B から見ることもでき、またファイバー・チャンネル・スイッチ 2 を使用してノード C から見ることもできます。サイト 1 に PV4 のミラーを持つことができます。1 つのサイトのみでノードに接続されているディスク (PV5 と PV 6) を、サイトをまたいでミラーすることはできません。

ディスク情報はローカル・サイトからリモート・サイトに複製されます。このデータ転送の速度は、チャンネルの物理的特性、距離、および LVM ミラー効率によって変わります。

動的 LPAR を使用したクラスター構成

AIX の拡張区画機能には、システムの CPU、メモリー、入出力スロット・リソースを動的に割り振る機能 (動的 LPAR) が用意されています。

LPAR と組み合わせて PowerHA SystemMirror を使用することにより、以下のことができます。

<物理アダプター名>_Large_Send

このアダプターの大量送信オプションを示すフラグ。

<物理アダプター名>_Large_receive

このアダプターの大量受信オプションを示すフラグ。注: このフラグが設定されていて、実際のアダプターがこのフラグをサポートしている場合、実際のアダプターが受信したパケットは、上位層に渡される前に集約されます。その結果、パフォーマンスが向上します。

<物理アダプター名>_DMA_Errors

リソースなしエラーによりハードウェアがドロップした受信パケット数。注: 通常、このエラーは、アダプター上の受信バッファがすべて使い果たされたことが原因で発生します。アダプターによっては、受信バッファのサイズが構成可能なパラメーターとなっている場合もあります。

<物理アダプター名>_Speed

このアダプターのメディア速度属性を示します。

- システム・リソースの動的な割り振りにより、通常のシステム・アップグレードを実行する。動的 LPAR と共に PowerHA SystemMirror を使用すると、アプリケーションの作業負荷をある論理区画から別の論理区画に自動的に移動することにより、アプリケーションを危険にさらすことなく、最初の論理区画がアップグレードできるようにうまく計画されたシステム・アップグレードが必要とされるダウン時間を短縮することができます。
- CPU およびメモリー・リソースを効果的に再配分して、作業負荷を管理する。動的 LPAR を PowerHA SystemMirror と結合することにより、カスタマイズされたアプリケーション開始スクリプトと停止スクリプトを使用して、その時点でアプリケーションの作業負荷がかかっている論理区画に動的に CPU およびメモリー・リソースを再配分し、単一フレームでのアプリケーションの移行をさらにサポートすることができます。これにより、アプリケーションのサポートに必要な処理能力やリソースを維持しながら、消費するリソースがより少ないタスクであるアップグレードに必要な最低限のリソースを提供することができます。

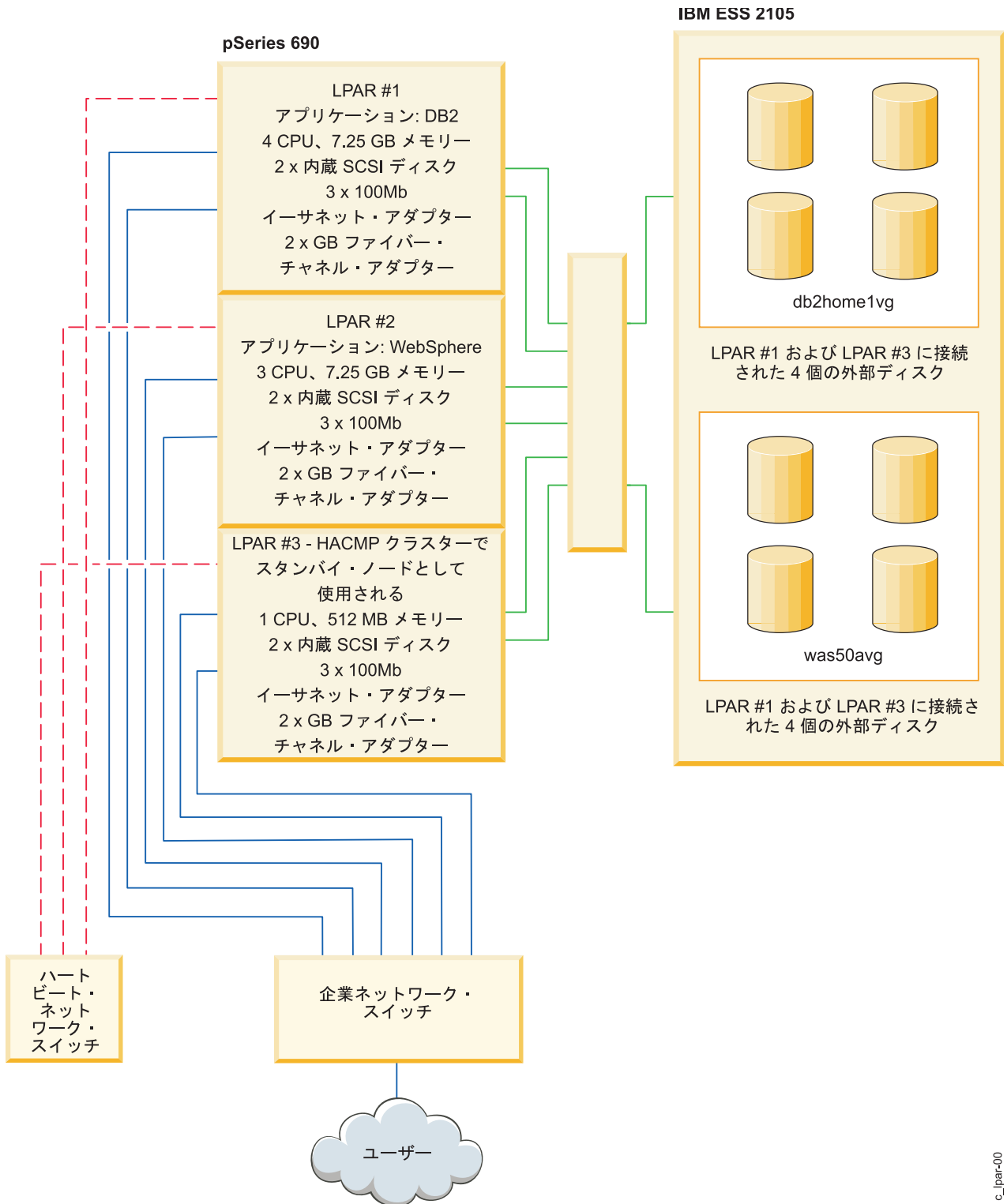
注: 同じ物理サーバー内ですべてのクラスター・ノードを LPAR として構成しないでください。このような構成は重大な単一障害点となる可能性があります。次の例は、3 つの LPAR を使用したクラスター構成を表しています。

- LPAR 1 はバックエンド・データベース (DB2 UDB) を実行します。
- LPAR 2 は WebSphere Application Server (WAS) を実行します。
- LPAR 3 は、DB2 LPAR および WAS LPAR の両方で、バックアップ (スタンバイ) として稼働しています。この LPAR には必要最低限の CPU およびメモリー・リソースしか含まれません。

例えば、計画的なアップグレードや、1 つめまたは 2 つめの LPAR におけるリソース障害などにより、DB2 または WAS アプリケーションのどちらかを 3 つめの LPAR に移動する時が来たら、PowerHA SystemMirror で、1 次 LPAR からスタンバイ LPAR へ、CPU とメモリーの動的再割り振りを自動的に行うためのカスタマイズされたアプリケーション開始スクリプトと停止スクリプトを使用することができます。この操作により、3 つめの LPAR は、ビジネス効率要件を満たすために必要な CPU およびメモリー・リソースを取得できます。PowerHA SystemMirror により、アプリケーションを含むリソース・グループが本来の LPAR に戻されると、CPU とメモリー・リソースもこのグループと共に自動的に移動します。注: 一般に、動的 LPAR では、CPU、メモリー、および入出力スロットの各リソースを動的に割り振

ることができます。PowerHA SystemMirror と動的 LPAR 入出力スロット・リソースには互換性はありません (ただし、PowerHA SystemMirror クラスターの外部で入出力スロット・リソースを動的に割り振ることはできます)。

このクラスター環境を図に表すと次のようになります。



c_lpar=00

図 31. 3 つの LPAR を持つクラスター

DLPAR とキャパシティー・オンデマンド

PowerHA SystemMirror は LPAR 間でアプリケーション・リソースを移動できます。また、リソース最適化高可用性 (ROHA) 機能を使用して、必要な動的リソース調整を実行することができます。ROHA は、Power Systems で使用できるフィーチャーを使用して、以下のタイプのハードウェア・リソースを動的に管理します。

- Capacity on Demand (CoD) 機能 (On/Off CoD および Enterprise Pool CoD を含む) は、フレーム (CEC) レベルでメモリーおよび CPU リソースを管理します。
- DLPAR 機能は、論理区画レベルでメモリーおよび CPU リソースを管理します。

CoD 機能は、非アクティブなプリインストールされたリソース (CPU およびメモリー) を活動化できます。ROHA は CoD 機能を使用して、環境のリソース要件が変わったときにこれらのリソースを活動化します。例えば、テークオーバー・プロセス中、環境にはスタンバイ・フレーム上に追加のリソースが必要です。この例では、追加リソースが動的に提供されるため、スタンバイ・フレームで永続ハードウェア・リソースを追加する必要はありません。

アクティブ・ノードは、十分な永続リソースのあるフレーム上の LPAR によってホストされます。スタンバイ・ノードは、最小の永続リソースを持つフレーム上の LPAR によってホストされ、リソースの動的追加を ROHA に頼ります。ROHA 機能を使用すれば、ご使用の環境内のピーク・ワークロードまたは予期しないワークロードに合わせて、素早く簡単に追加リソースを獲得することができます。

PowerHA SystemMirror 構成プロセスおよび機能

以下のトピックでは、PowerHA SystemMirror クラスター構成プロセス、および PowerHA SystemMirror ソフトウェアに付属の管理ツールについて概説します。

PowerHA SystemMirror に渡す情報

クラスターを構成する前に、ビルディング・ブロックが計画および構成されていること、ならびに PowerHA SystemMirror が各ノードにアクセスできるように初期通信パスが存在することを確認してください。このセクションでは、クラスターを構成するために実行する必要がある基本的なタスクについて説明します。

クラスターの物理構成に関する情報

クラスターの物理構成は、いくつかの計画タスクおよび構成タスクからなります。

そのタスクを以下に示します。

- クラスターの TCP/IP ネットワークのサポートを確認します。
- クラスターの共用ディスク・デバイスを構成します。
- クラスターの共用ボリューム・グループを構成します。
- PowerHA SystemMirror を使用するときの対象となる基幹業務のアプリケーションを検討します。また、アプリケーション・コントローラーと、アプリケーションごとに最適なりソース・グループ管理のタイプを検討します。
- PowerHA SystemMirror クライアントに関連する問題を調査します。
- 複数の回線または無停電電源装置、冗長物理ネットワーク・インターフェース・カード、ノードを接続する複数のネットワーク、およびディスク・ミラーリングを用いて、物理的冗長度を保証します。

関連情報:

PowerHA SystemMirror プランニング・ガイド

AIX 構成情報

クラスター・コンポーネントは、AIX レベルで正しく構成されていなければなりません。

そのためには、以下の事項を確認します。

- クラスター・ノードへの基本的通信が存在します。
- ボリューム・グループ、論理ボリューム、ミラーリング、およびファイルシステムが構成され、設定されていること。論理的冗長度を保証するためには、さまざまなタイプのリソース・グループを検討し、リソースをどのようにリソース・グループにグループ分けするかを計画します。

関連情報:

PowerHA SystemMirror インストール・ガイド

PowerHA SystemMirror によってディスカバーされる情報

わずか数ステップで基本的なクラスター・コンポーネントを定義することができます。PowerHA SystemMirror は、クラスター構成を支援するため、構成に必要な情報を各ノードから自動的に取得できます。

注: クラスター構成を、より簡単に素早くするために、クラスター構成アシスタントを使用することもできます。詳しくは、『2 ノード・クラスター構成アシスタント』を参照してください。

自動ディスカバリー・プロセスを機能させるには、PowerHA SystemMirror で以下の条件を満たす必要があります。

- 事前に、物理コンポーネントが構成されていて、必要な AIX 構成がすべて実行されている。
- 各ノードに作動通信パスが存在する。この情報は、標準構成パスの使用時にクラスターの TCP/IP トポロジーを自動構成するために使用されます。

これらのタスクが完了すると、PowerHA SystemMirror はクラスター内の事前定義物理コンポーネントを自動的にディスカバーし、デフォルトの動作を選択します。また、構成プロセスの途中で変更があると、PowerHA SystemMirror はクラスター情報のディスカバリーを実行します。

ディスカバリーの実行により、現在の AIX 構成情報がすべてのクラスター・ノードから取得されます。この情報はピック・リストに表示され、ユーザーが既存のコンポーネントを正確に選択できるようにします。

PowerHA SystemMirror 自動ディスカバリー・プロセスは、扱いやすく、高速で実行されます。クラスター管理者がこのプロセスを待機するようなことはありません。

クラスター構成オプション: 標準および拡張

このセクションでは、構成プロセスが大幅に単純化されています。構成プロセスについては「*Administration Guide*」に記載されていますが、このセクションでは、PowerHA SystemMirror クラスターを構成する 2 つの方法の簡単な概要を示します。

関連情報:

Administering PowerHA SystemMirror

標準構成パスを使用した PowerHA SystemMirror クラスターの構成

わずか数ステップでクラスターの基本コンポーネントを PowerHA SystemMirror 構成データベースに追加できます。標準クラスター構成パスにより、構成プロセスが単純化され、スピードアップされます。PowerHA SystemMirror が自動的にディスカバリーを起動して、情報を収集し、デフォルトの動作を選択するためです。

このパスを使用すると、次のようになります。

- クラスター情報の自動ディスカバリーがデフォルトで実行されます。PowerHA SystemMirror 構成プロセスを開始する前に、AIX でネットワーク・インターフェースまたはネットワーク・デバイスを構成する必要があります。PowerHA SystemMirror では、ユーザーは他のノードへの初期通信パスを確立します。この操作を行うと、PowerHA SystemMirror はこの情報を収集し、物理接続に基づいてクラスター・ノードおよびネットワークを自動的に構成します。ディスカバーされたネットワークはすべてクラスター構成に追加されます。
- IP エイリアスは、ネットワーク・インターフェースへの IP ラベルまたはアドレス結合のデフォルト機構として使用されます。
- 最も一般的なタイプのリソースを構成することができます。ただし、リソース・グループのフォールオーバーおよびフォールバック動作のカスタマイズは限定されています。

拡張構成パスを使用した PowerHA SystemMirror クラスターの構成

それほど一般的でないクラスター・エレメントを構成する場合、またはクラスター・ノードのそれぞれに対する接続が確立されていない場合、前のリリースの PowerHA SystemMirror ソフトウェアと類似した方法で、情報を手動で入力することができます。

PowerHA SystemMirror 拡張構成 SMIT パスを使用する場合、リモート・ノード上にコンポーネントが存在するのであれば、クラスター情報のディスカバリーを手動で開始する必要があります。すなわち、ディスカバリーはオプションです (標準の PowerHA SystemMirror 構成 SMIT パス使用時のように自動ではありません)。

拡張構成メニューのオプションを使用して、クラスターの基本的なコンポーネントを PowerHA SystemMirror 構成データベースに追加することができ、また多くの補助的なタイプのリソースも追加することができます。標準構成メニューに含まれていないすべてのコンポーネント、ポリシー、およびオプションについては、拡張構成パスを使用して、クラスターをカスタマイズしてください。

クラスター・セキュリティー

各ノードで実行されるクラスター通信デーモン **clcomd** 経由でノード間のすべての通信が送信されます。

clcomd デーモンは、ノード間の接続認証およびメッセージ認証または構成された暗号化を管理します。PowerHA SystemMirror クラスター通信デーモンは、信頼されている **/etc/cluster/rhosts** ファイルを使用し、**/rhosts** ファイルへの依存を除去します。デーモンは、メッセージ認証と暗号化のサポートを提供します。

インストール・ツール、構成ツール、および管理ツール

PowerHA SystemMirror には、クラスターをインストール、構成、および管理するためのツールが組み込まれています。そのツールについて、以下のセクションで説明します。

特定のアプリケーションを PowerHA SystemMirror と統合するための Smart Assist

特定のアプリケーション用の Smart Assist は、システム上の構成を検査し、PowerHA SystemMirror がモニターしなければならないリソース (サービス IP ラベル、ボリューム・グループ) を判別します。その後、Smart Assist は 1 つ以上のリソース・グループを構成し、アプリケーションとそのリソースの可用性を高めます。

Smart Assist は次のアクションを実行します。

- インストールされているアプリケーション、および必要に応じて、現在構成されているリソース (サービス IP アドレス、ファイルシステム、およびボリューム・グループなど) を発見する

- ユーザーから構成情報 (新規サービス IP アドレスなど) を取得または変更するための SMIT インターフェイスを提供する
- PowerHA SystemMirror に対するアプリケーションを定義し、そのアプリケーションのカスタム開始/停止スクリプトを提供する
- そのアプリケーション用のアプリケーション・モニターを提供する
- 次のものを格納するためのリソース・グループを構成する
- 一時リソース・グループおよびロケーション依存関係を構成する (アプリケーション・ソリューションで必要とされる場合)
- PowerHA SystemMirror ファイル・コレクション機能による同期化が必要なファイルを指定する
- あらかじめ構成されているアプリケーションを必要に応じて変更する
- 構成を検証する
- アプリケーションのクラスター構成をテストする

サポートされるアプリケーション

PowerHA SystemMirror は、次のアプリケーションおよび構成モデルに対して Smart Assist を提供します。

- DB2
- DB2 - ホット・スタンバイ
- DB2 - 相互テークオーバー
- WebSphere 6.0
- WebSphere Application Server 6.0
- WebSphere Cluster Transaction Log recovery
- Deployment Manager
- Tivoli® Directory Server
- IBM HTTP Server
- Oracle 10G

一般アプリケーション用の Smart Assist

一般アプリケーション用の Smart Assist は、専用の Smart Assist のないインストール済みアプリケーションの構成を支援します。

ユーザーは、次のようないくつかの基本情報を提供します。

- 1 次ノード - デフォルトでは、ローカル・ノード
- テークオーバー・ノード - デフォルトでは、ローカル・ノード以外のすべての構成済みノード
- アプリケーション名
- アプリケーション開始スクリプト
- アプリケーション停止スクリプト
- サービス IP ラベル

一般アプリケーション用の Smart Assist は、2 ノード・クラスター構成アシスタントとほぼ同じ方法でクラスター構成を行います (ただし、構成には 3 つ以上のノードを設定できます)。ユーザーは、一般アプリケーション用の Smart Assist を使用する場合に、アプリケーションを変更、テスト、または除去することができます。

Smart Assist API

PowerHA SystemMirror には、OEM が Smart Assist を開発して独自のアプリケーションを PowerHA SystemMirror に統合できるように、「*Smart Assist developers guide*」が同梱されています。

関連情報:

PowerHA SystemMirror Smart Assist アプリケーション開発ガイド

クラスター・サービスの開始、停止、および再始動

PowerHA SystemMirror をインストールし、クラスターを構成すると、クラスター・サービスを開始できます。PowerHA SystemMirror では、クラスター・サービスの開始、停止、および再始動のためのオプションが簡素化および改善されています。PowerHA SystemMirror は、アプリケーションを中断せずに、クラスター・サービスの開始と停止の要求を処理するため、ユーザーはアプリケーションを完全に制御できません。

PowerHA SystemMirror では、次のことが可能です。

- クラスター・サービスを開始および再始動します。ユーザーがクラスター・サービスを開始したり、シャットダウン後に再始動したりするときに、PowerHA SystemMirror はデフォルトで、ユーザーが定義した方法に従ってリソースを自動的にアクティブにします。その際、アプリケーション依存関係、アプリケーションの開始スクリプトと停止スクリプト、動的属性やその他のさまざまなパラメーターが考慮されます。すなわち、PowerHA SystemMirror は、リソース・グループとそれに含まれるアプリケーションを自動的に管理 (および必要に応じて、アクティブ化) します。

さらに、PowerHA SystemMirror クラスター・サービスを開始して、自動的にはいずれのリソース・グループ (およびアプリケーション) も開始しないことを PowerHA SystemMirror に指示することもできます。アプリケーションが既に実行されている場合、クラスター・サービスを開始する前にそのアプリケーションを停止する必要がなくなりました。

注: PowerHA SystemMirror はアプリケーション・モニターとアプリケーション始動スクリプトを使用して、アプリケーションを始動する必要があるかどうかの検証、あるいはアプリケーションが既に実行されているかどうかの検証 (PowerHA SystemMirror はアプリケーションの 2 番目のインスタンスの開始を試行しません) を行います。PowerHA SystemMirror は、構成済みのアプリケーション・モニターを使用して、アプリケーションの障害を検出します。アプリケーションの重複インスタンスを開始することがないように、PowerHA SystemMirror で開始時に実行中のクラスターを検出するようにアプリケーション・モニターを構成する必要があります。もう 1 つの方法は、アプリケーション・コントローラーの重複インスタンスが開始されないようにするためのスクリプトを実行することです。

- クラスター・サービスをシャットダウンします。PowerHA SystemMirror のシャットダウン時に、リソース・グループに対して次の 3 つのアクションのいずれかを選択できます。
 - オフラインにする
 - 別のノードに移動する
 - リソース・グループを UNMANAGED 状態にする

クラスター・マネージャーはすべてのノードの状態を記憶しており、ユーザーがそれらのノードを再始動しようとしたときに適切に対応します。

関連情報:

Administering PowerHA SystemMirror

SMIT インターフェース

PowerHA SystemMirror ソフトウェアに付属の SMIT パネルを使用して実行できるタスクがいくつかあります。

そのタスクを以下に示します。

- クラスター、ノード、ネットワーク、リソース、およびイベントを構成する。
- クラスター構成のスナップショットを取り込んで復元する。
- ログ・ファイルを読み取る。
- クラスター問題を診断する。
- C-SPOC ユーティリティーを使用してクラスターを管理する。
- リソース・グループ管理タスクを実行する。
- 自動エラー通知を構成する。
- 動的アダプター・スワップを実行する。
- クラスター・パフォーマンス調整を構成する。
- カスタム・ディスク方式を構成する。

PowerHA SystemMirror システム管理 (C-SPOC)

クラスターの管理を容易にするため、PowerHA SystemMirror では、あるノードからコマンドを実行してから、変更内容を検証して他のすべてのノードと同期化する方法を提供しています。PowerHA SystemMirror システム管理ツール、Cluster Single Point of Control (C-SPOC) を使用すれば、主幹業務のジョブを停止することなく、ユーザー、ファイル、およびハードウェアを自動的に追加することができます。

C-SPOC を使用すれば、以下のタスクを実行できます。

- PowerHA SystemMirror サービスの開始または停止
- PowerHA SystemMirror 通信インターフェース管理
- PowerHA SystemMirror リソース・グループおよびアプリケーション管理
- PowerHA SystemMirror ファイル・コレクション管理
- PowerHA SystemMirror ログ表示と管理
- PowerHA SystemMirror セキュリティーとユーザー管理
- PowerHA SystemMirror 論理ボリューム管理
- PowerHA SystemMirror コンカレント論理ボリューム管理
- PowerHA SystemMirror 物理ボリューム管理
- ノード上で SMIT セッションをオープン

C-SPOC ユーティリティーは、最大 16 ノードまでのクラスター内の共有 LVM コンポーネントの保守を単純化します。C-SPOC コマンドは、単一ノードに作用する標準 AIX コマンドと同等の機能を、クラスター環境に提供します。C-SPOC は、繰り返しタスクを自動化することにより、エラーの原因となりうる要素を排除し、処理速度を速めます。

C-SPOC 機能がないと、システム管理者は、各クラスター・ノード上で管理用タスクを個別に実行する必要があります。例えば、ユーザーを追加するには、通常、各クラスター・ノード上でこのタスクを実行する必要があります。C-SPOC ユーティリティーを使用すると、1 つのノード上で実行されたコマンドがほかのクラスター・ノード上でも実行されます。そのため、C-SPOC は、管理上のオーバーヘッドを最小化

し、ノード状態不一致の可能性を少なくします。C-SPOC を使用すると、単独ノード上で C-SPOC コマンドを発行するだけで、そのユーザーはすべての指定クラスター・ノードに追加されます。

C-SPOC はまた、論理ボリューム・コンポーネントの管理およびクラスター・サービスの制御の効率を向上させます。C-SPOC ユーティリティを使用すると、単独ノードから各ノード上のクラスター・サービスを始動または停止させることができます。

次の図は、2 ノード構成と、1 つのクラスター・ノードからクラスター・サービスが開始されるときに行われるコマンド、スクリプト、およびノード間の対話を示したものです。C-SPOC コマンドはすべて、プレフィックス `cl_` で始まります。

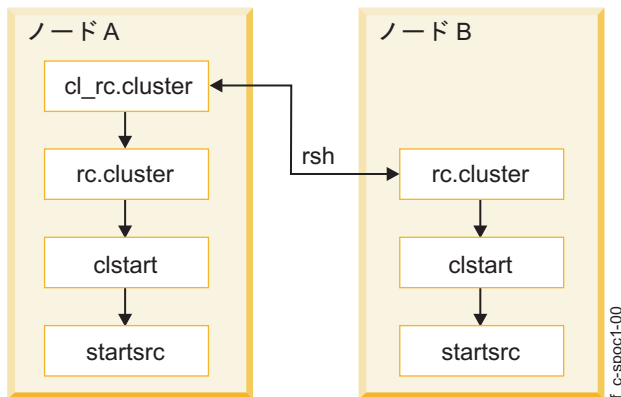


図 32. C-SPOC ユーティリティによってクラスター始動時に使用されるコマンドのフロー

C-SPOC は、SMIT のメニューおよびパネルからアクセス可能な、C-SPOC 独自のクラスター管理コマンド・セットによりこの機能を提供します。C-SPOC を使用するには、「PowerHA SystemMirror SMIT」メニューから「Cluster System Management (クラスター・システム管理)」オプションを選択します。

クラスター・スナップショット・ユーティリティ

クラスター・スナップショット・ユーティリティを使用すれば、後で復元したいクラスター構成を保管しておくことができます。

また、システムやクラスター構成問題を診断する際に役立つ追加のシステム情報やクラスター情報を保存しておくこともできます。クラスターに関する追加情報を格納する独自のカスタム・スナップショット・メソッドを作成することができます。

クラスター・スナップショットを使用すれば、スナップショットにログ・ファイルを保存する処理をスキップできます。クラスター・スナップショットは、クラスター構成情報を記録するために使用できますが、クラスターはクラスターの運用を記録するだけで、構成情報は記録しません。デフォルトで、PowerHA SystemMirror は、クラスター・スナップショットを作成するときにクラスター・ログ・ファイルを収集しませんが、SMIT でログを収集するように指定することはできます。ログ収集をスキップすると、スナップショット・ユーティリティの実行時間が高速化され、スナップショットのサイズが減少します。

カスタマイズされたイベント処理

個々のサイトのニーズに合わせてイベント処理を調整するために複数のイベント前処理およびイベント後処理を定義することができます。

関連情報:

Administering PowerHA SystemMirror

リソース・グループ管理ユーティリティ

リソース・グループ管理ユーティリティ、clRGmove により、クラスター内のリソース・グループを管理し、PowerHA SystemMirror の障害リカバリー機能を強化することができます。

このユーティリティを使用すると、クラスター・サービスを停止することなく、あらゆるタイプのリソース・グループを (そのリソース (IP アドレス、アプリケーション、およびディスクなど) と一緒に) オンラインまたはオフラインにしたり、別のノードに移動したりすることができます。リソース・グループ管理により、クラスターの管理作業が効率化し、クラスター・ハードウェア・リソースの利用率が向上します。

また、リソース・グループ管理を使用すれば、クラスターをリポートしたり稼働中のノードを停止したりせずに、選択的な保守を行うことができます。例えば、このユーティリティを利用して任意のリソース・グループのノードを解放し、特定のクラスター・ノード上のシステム保守を実行することができます。

リソース・グループ管理ユーティリティは、現在ノードが所有している他のリソース・グループに影響を与えません。現在のノードがリソース・グループを解放すると、ノードのフォールオーバーのときと同じように、宛先ノードがそのグループを取得します (リソース・グループ間にロケーション依存関係が構成されている場合、PowerHA SystemMirror はそれが指定どおりであることを確認します)。

リソース・グループ管理を使用して、次のタスクを実行します。

- 作業用クラスター内で、非コンカレント・リソース・グループをあるノードから別のノード (およびあるサイトから別のサイト) に一時的に移動する。
- クラスター内の 1 つまたはすべてのノードでリソース・グループをオンラインまたはオフラインにする。

グループを移動すると、再度それを移動するまで、そのグループは移動先のノードにとどまります。

「Fallback to Highest Priority Node (最高優先順位ノードへのフォールバック)」フォールバック・ポリシーを持つグループを移動すると、そのグループは、優先順位が最も高い「新規」の一時ノードにフォールバックまたは戻されます (PowerHA SystemMirror が、後続のクラスター・イベント時に他のノード上でそのグループを回復する必要がある場合)。

そのグループを再度移動したい場合、PowerHA SystemMirror は、グループをホストできる、さらに優先順位の高いノードの存在を検出したかどうかをインテリジェントな方法 (宛先ノードのピック・リスト) でユーザーに通知します。ユーザーはグループをそのノードに移動することをいつでも選択できます

PowerHA SystemMirror ファイル・コレクション管理

ボリューム・グループと同じように、各クラスター・ノードに配置された特定のファイルは、PowerHA SystemMirror (および他のアプリケーション) が正しく動作するように同期に保つ必要があります。このようなファイルには、イベント・スクリプト、アプリケーション・スクリプト、いくつかの AIX および PowerHA SystemMirror 構成ファイルが含まれます。PowerHA SystemMirror ファイル・コレクション管理は、ファイルのリストがクラスター越しに同期に保たれることを要求する簡単な方法を提供します。

PowerHA SystemMirror ファイル・コレクションを使用すると、更新されたファイルを手動で各クラスター・ノードにコピーし、ファイルが正しくコピーされたことを確認し、各ノードが同じバージョンを持っていることを確認する必要がありません。

また、これらのファイルの 1 つ以上が、1 つ以上のクラスター・ノードで不注意に削除または破損された場合、問題を判別するには時間と労力が必要です。PowerHA SystemMirror ファイル・コレクションを使用すると、このシナリオは緩和されます。PowerHA SystemMirror は、ファイル・コレクション内のフ

ファイルがいつ削除されたか、またはファイル・サイズがゼロに変更されたかを検出し、管理者に通知するためにメッセージをログに記録します。2つの定義済みの PowerHA SystemMirror ファイル・コレクションがデフォルトでインストールされます。

- **Configuration_Files。** `/etc/hosts` や `/etc/services` などの必須システム・ファイルのコンテナ。
- **PowerHA SystemMirror_Files。** PowerHA SystemMirror 構成内のすべてのユーザー構成可能ファイルのコンテナ。これは特別なファイル・コレクションであり、低レベル層となっているファイル・コレクション伝搬ユーティリティーが、PowerHA SystemMirror 構成データベース (ODM) クラス内のすべてのユーザー構成可能ファイルを参照するのに使用します。

関連情報:

Administering PowerHA SystemMirror

PowerHA SystemMirror インストール・ガイド

モニター・ツール

PowerHA SystemMirror には、さまざまなツールがモニター用に用意されています。

ここで説明するユーティリティーの多くは、`clhosts` ファイルを使用して、PowerHA SystemMirror クラスタ・ノード間の通信を可能にしています。

関連概念:

118 ページの『`clhosts` ファイルの概要』

`Clinfo`、`HAView`、`clstat` ユティリティーなど、このトピックで説明する多くのモニターが `clhosts` ファイルを使用します。`clhosts` ファイルには、PowerHA SystemMirror クラスタ・ノード間の通信を可能にする IP アドレスの情報が含まれています。`clhosts` ファイルは、PowerHA SystemMirror クラスタのすべてのサーバーおよびクライアント上にあります。

関連情報:

Administering PowerHA SystemMirror

クラスタ・マネージャ

クラスタ・マネージャは SNMP 情報を用意して SNMP クライアントに通知します。

このデーモンは、ノードやインターフェースのクラスタ状態の変化に関するクラスタ情報を収集するものです。クラスタ情報は、SNMP コマンドを用いて、または SNMP ベースのクライアント・プログラムによって、検索することができます。

関連概念:

60 ページの『クラスタ・マネージャおよび SNMP モニター・プログラム』

PowerHA SystemMirror クラスタは動的であり、時間の経過と共に各種の推移が可能です。

クラスタ情報プログラム

クラスタ情報プログラム (`Clinfo`) は、SNMP からクラスタ情報を収集し、このプログラムと通信するクライアントにクラスタの状態の変化を通知します。

関連概念:

61 ページの『クラスタ情報プログラム』

クラスタ情報プログラム (`Clinfo`) の `clinfo` デーモンは、SNMP ベースのモニターです。

アプリケーション・モニター

アプリケーション・モニターによって、特定のアプリケーションとプロセスをモニターするアプリケーション・コントローラー用の複数のモニターを構成できます。そして、プロセスまたは他のアプリケーション障害の予想しない終了の検出について取る処置を定義することができます。

関連概念:

71 ページの『単一障害点としてのアプリケーションの除去』

PowerHA SystemMirror クラスタを作成する主要な理由は、主幹業務アプリケーション用に可用性の高い環境を提供することです。

clam_nfsv4 アプリケーション・モニター

ネットワーク・ファイル・システム・バージョン 4 (NFSV4) のリソースがリソース・グループ内に指定されると、clam_nfsv4 アプリケーション・モニターは PowerHA SystemMirror バージョン 7.1 によって自動的に追加されます。

clam_nfsv4 アプリケーション・モニター・スクリプトは、デフォルトでは 1 分間に 1 回ずつ実行されます。モニターの再始動カウントはゼロです。clam_nfsv4 アプリケーション・モニターには、フォールオーバーの障害アクションがあります。フォールオーバーが発生すると、モニターによってログに記録されたエラーによって NFSV4 アプリケーション・サーバーがオフラインの状態になり、アプリケーション・サーバー上のリソース・グループは次の最高優先順位のノードに移動します。

clam_nfsv4 アプリケーション・モニターを除去することはできませんが、障害アクションを通知に変更することによって使用不可にすることができます。

clam_nfsv4 アプリケーション・モニターは、クロスマウント構成のノードが正常にマウントされているかを検査します。また、モニターは NFS サーバーがクロスマウント構成を通して出されたクライアント要求に応答できるかも検査します。

clam_nfsv4 アプリケーション・モニターは、以下のサブシステムが実行中か検査します。

- nfsrgyd
- nfsd
- rpc.mountd
- rpc.statd
- rpc.lockd

サブシステム機能に実行中でないものがあると、clam_nfsv4 アプリケーション・モニターは NFSV4 アプリケーション・サーバーがオフラインであると識別します。

リソース・グループを含む NFS ノードを登録する必要があります。また、リソース・グループを正常にエクスポートするよう NFS を構成する必要もあります。

関連情報:

PowerHA SystemMirror での NFS の使用

PowerHA SystemMirror での NFS クロスマウント

clam_nfsv4 アプリケーション・モニターのトラブルシューティング

「Show cluster applications (クラスター・アプリケーションの表示)」 SMIT オプション

「Show Cluster Applications (クラスター・アプリケーション表示)」 SMIT オプションは、クラスター構成のアプリケーション中心のビューを提供します。

このユーティリティーは、「アプリケーション・ダウン」タイプのビューに既存のインターフェースと情報を表示します。

クラスター・ステータス・ユーティリティー (clstat)

クラスター・ステータス・ユーティリティー `/usr/es/sbin/cluster/clstat` は、クラスター状況をモニターします。このユーティリティーは、主要クラスター・コンポーネントの状況 (クラスター自体、クラスター内のノード、ノードに接続されているネットワーク・インターフェース、および各ノード上のリソース・グループ) を報告します。

クラスターが動作中、停止中、または不安定な状態のいずれであるかを報告します。また、ノードが動作中、停止中、結合中、結合解除中、および再構成中のいずれであるかについても報告し、クラスター内のノードの数も報告します。`clstat` ユーティリティーには、ASCII、Motif、X Window System、および HTML インターフェースが用意されています。`clstat` は ASCII SMIT から実行できます。

クラスター全体としては、`clstat` は、クラスター状態とクラスター・ノードの数を示します。各ノードについて、`clstat` は、ノードに付加された各サービス・ネットワーク・インターフェースの IP ラベルとアドレス、およびそのインターフェースがアップかダウンかを表示します。`clstat` はリソース・グループ状態も表示します。

クラスター状況情報を ASCII または X Window System 表示モード、あるいは Web ブラウザーで表示できます。

IP バージョン 6 ループバック・アドレスが `/usr/es/sbin/cluster/etc/clhosts` ファイルで指定されていて、IP バージョン 6 アドレスがクラスターで使用されていない場合、`clstat` で示されるクラスター状況情報は絶えず変化し続ける可能性があります。`clstat` の状態が UP から DOWN へ、DOWN から UP へと繰り返し変化する場合は、`/usr/es/sbin/cluster/etc/clhosts` ファイルから `:::1` エントリーを除去してください。

`/usr/es/sbin/cluster/etc/clhosts` ファイルで `:::1` エントリーを保持する場合は、以下の手順を実行してください。

1. `:::1` エントリーが `/usr/es/sbin/cluster/etc/clhosts` ファイルにあることを確認します。
2. 次の行を `/etc/snmpdv3.conf` ファイルに追加して、IP バージョン 6 アドレスをサポートするように Simple Network Management Protocol (SNMP) バージョン 3 を構成します。
`COMMUNITY public public noAuthNoPriv :::0 -`
3. `snmpd` デーモン、`clinfoES` デーモン、およびすべての従属デーモンを再始動します。

注: `clstat` ユーティリティーは、クラスターについての情報を検索するのに `Clinfo` API を使用します。したがって、`clstat` の出力を表示するには、クライアント・システム上で `Clinfo` が稼働していることを確認してください。

アプリケーション可用性分析ツール

アプリケーション可用性分析ツールは、PowerHA SystemMirror に定義されたアプリケーション・コントローラーでアプリケーションについてのアップタイム統計を測定します。

PowerHA SystemMirror ソフトウェアは、このツールでモニターすることを選択したアプリケーションについて包括的な情報を収集し、タイム・スタンプを取り、ログに記録します。SMIT で時間枠を選択すると、その時間枠における特定のアプリケーションのアップ時間とダウン時間の統計をツールで表示できます。

永続ノード IP ラベル

永続的ノード IP ラベル は、PowerHA SystemMirror クラスター・サービスがノード上で停止している場合でも、そのノードに接続できるようにする有用な管理ツールです。

永続ノード IP ラベルを定義した場合、PowerHA SystemMirror はそのノードに IP アドレスを割り当てようとします。永続ノード IP ラベルをノード上のネットワークに割り当てると、クラスター・ネットワーク上のノード・バインド IP アドレスを持つことができます。ノード・バインド IP アドレスを使用すると、管理目的でクラスター内の特定のノードにアクセスできます。永続ノード IP ラベルは、クラスター・ネットワーク上の特定のノードに割り当てることができる IP エイリアスです。

- 必ず同じノードにとどまります (ノード結合 です)。
- すでにサービス IP ラベルが定義されたネットワーク・インターフェース・カード上で共存します。
- そのノード上に追加の物理ネットワーク・インターフェース・カードをインストールする必要はありません。
- いずれのリソース・グループ にも属しません。

各ノードの各ネットワークにつき 1 つの永続ノード IP ラベルを持つことができます。

PowerHA SystemMirror 検証および同期化

PowerHA SystemMirror の検証および同期化のプロセスで検証される内容は、AIX システム・ファイルに対する特定の PowerHA SystemMirror の変更が正しいこと、クラスターとそのリソースが正しく構成されていること、セキュリティー (セットアップされている場合) が正しく構成されていること、クラスター・トポロジー、ネットワーク構成、および PowerHA SystemMirror リソースの所有権とテークオーバーについてすべてのノードが一致していることです。

検証では、カスタム・クラスター・スナップショット・メソッドが存在するかどうか、各クラスター・ノード上で実行できるかどうかをもた知らせます。

クラスターを構成、再構成、または更新した後で、クラスター検証手順を実行する必要があります。検証に成功すれば、構成は自動的に同期化されます。同期化した結果は、アクティブなクラスターで即時に有効となります。

検証ユーティリティーは実行されると、各ノード上の PowerHA SystemMirror 構成データベース内に詳細情報レコードを保持します。各ノードのサブディレクトリーには、最後に検証に成功したとき (pass)、最後よりも 1 つ前に検証に成功したとき (pass.prev)、および最後に検証に失敗したとき (fail) の情報が格納されています。

ユーティリティーのメッセージ出力では、エラーが発生した場所 (例えば、ノード、デバイス、コマンドなど) を示します。

自動クラスター構成モニターによる検証:

PowerHA SystemMirror は、自動的なクラスター構成モニターを提供します。デフォルトで PowerHA SystemMirror は、真夜中に 24 時間ごとに 1 回、アルファベット順の先頭にくるノード上で自動的に検証 ユーティリティーを実行します。

クラスター構成が無効になった場合、クラスター管理者に通知されます。クラスター確認が選択したクラスター・ノードで完了したとき、このノードは他のクラスター・ノードに通知します。各ノードは、日付、時刻、どのノードが検証を実行したか、および `/var/hacmp/log/clutils.log` ファイル内の確認の結果に関する情報を格納します。選択されたノードが利用不能になった場合、またはクラスター検証を完了できない場合、`var/hacmp/log/clutils.log` ファイル内のレポートの欠如でこれを検出できます。クラスター確認が完了し、構成エラーを検出した場合は、可能性のある問題についてユーザーに通知がなされます。

- 検証の終了状況は、クラスター検証プロセス完了についての情報と共にクラスター内で公開されます。
- ブロードキャスト・メッセージが、クラスター越しに送信され、`stdout` に表示されます。これらのメッセージは、ユーザーに、検出された構成エラーについて通知します。

修正アクションの確認:

クラスターの検証では、さまざまなユーザー構成 PowerHA SystemMirror サーバー・コンポーネントに対して一連の検査が実行されます。それぞれの検査で、クラスターの整合性の問題または構成エラーの検出を試みます。

PowerHA SystemMirror の運用には重要であるものの、PowerHA SystemMirror ソフトウェア自体の一部ではない情報がすべてのクラスター・ノードに適切に伝搬されない場合、エラー状態が発生します。

デフォルトで、検証は、標準構成と拡張構成で共に有効な自動修正アクション・モードで実行されます。これは、検証を実行する際に推奨されるモードです。必要に応じて、拡張構成に対して自動修正アクション・モードを無効にすることができます。ただし、自動修正アクション・モードで検証を行えば、クライアント・ベースの `clhosts` ファイル (このファイルは多くのモニターで使用されます) を作成するタスクなど、多くの構成タスクを自動化できます。

検証で以下のいずれかの条件が検出されると、エラー・チェックを継続する前に修正アクションに許可を与えることができます。

- PowerHA SystemMirror 共用ボリューム・グループ・タイム・スタンプはすべてのノードでは一致しません。
- ノード上の `/etc/hosts` ファイルに PowerHA SystemMirror 管理の IP ラベルと IP アドレスの一部が含まれていません。
- ファイルシステムは、リソース・グループの一部であるノード上には作成されませんが、ディスクは利用できます。
- ディスクは利用できますが、ボリューム・グループはノードにインポートされていません。
- 必要な `/etc/services` エントリーがノードにありません。
- 必要な PowerHA SystemMirror `snmpd` エントリーがノードにありません。

検証中にエラーが検出されて修正アクションが起動された場合、ユーティリティーは最初のパスを終了した後で、すべてのチェックを再び実行します。同じチェックが再び不合格で、オリジナルの問題がエラーの場合、エラーはログ記入され、検証は失敗します。オリジナルの条件が警告の場合は、検証は成功します。

関連概念:

118 ページの『`clhosts` ファイルの概要』

Clinfo、HAView、clstat ユーティリティーなど、このトピックで説明する多くのモニターが `clhosts` ファイルを使用します。`clhosts` ファイルには、PowerHA SystemMirror クラスター・ノード間の通信を可能にする IP アドレスの情報が含まれています。`clhosts` ファイルは、PowerHA SystemMirror クラスターのすべてのサーバーおよびクライアント上にあります。

カスタム検証メソッド:

SMIT からは、クラスター構成についての個々のチェックを実行するカスタム定義検証メソッドの追加、変更、および削除も行えます。コマンド行または SMIT インターフェースから検証を実行し、クラスター・イベントに応じてカスタマイズしたりリモート通知メソッドを発行することができます。

clhosts ファイルの概要:

Clinfo、HAView、clstat ユーティリティーなど、このトピックで説明する多くのモニターが **clhosts** ファイルを使用します。**clhosts** ファイルには、PowerHA SystemMirror クラスター・ノード間の通信を可能にする IP アドレスの情報が含まれています。**clhosts** ファイルは、PowerHA SystemMirror クラスターのすべてのサーバーおよびクライアント上にあります。

ファイルがある場所による **clhosts** ファイルの違いを、以下の表にまとめます。

表 5. **clhosts** ファイル

clhosts ファイル	説明
サーバー・ベースのファイル	このファイルは、すべての PowerHA SystemMirror サーバー・ノード上の <code>/usr/es/sbin/cluster/etc/</code> ディレクトリー内にあります。PowerHA SystemMirror のインストール中に、デフォルトの IP バージョン 4 (IPv4) ループバック・アドレス 127.0.0.1 が自動的にファイルに追加されます。インストール時に IP バージョン 6 (IPv6) アドレスが構成されると、デフォルトの IPv6 ループバック・アドレス <code>::1</code> も追加されます。この IP アドレスで通常定義される名前ループバックおよびエイリアス・ローカル・ホストは必要ありません。
クライアント・ベースのファイル	このファイルは、すべての PowerHA SystemMirror クライアント・ノード上の <code>/usr/es/sbin/cluster/etc/</code> ディレクトリー内にあります。自動訂正機能で検証を実行すると、このファイルはサーバー上に clhosts.client というファイル名で自動的に生成されます。このファイルをクライアント・ノードに手動でコピーし、 clhosts という名前に変更する必要があります。このファイルには既知のすべての IP アドレスを含み、127.0.0.1、 <code>::1</code> 、ループバック、またはローカル・ホスト・アドレスを含んでいてはなりません。

モニター・デーモンが起動されると、ローカルの `/usr/es/sbin/cluster/etc/clhosts` ファイルを読み込み、以下のようにして通信に使用できるノードを調べます。

- PowerHA SystemMirror サーバー・ノード上で稼働するデーモンの場合は、ローカルのサーバー・ベースの **clhosts** ファイルはループバック・アドレス (127.0.0.1 and `::1`) のみを必要とします。これは、PowerHA SystemMirror のサーバー部分がインストールされる際に、サーバー・ベースの **clhosts** ファイルに自動的に追加されます。
- PowerHA SystemMirror クライアント・ノード上で稼働するデーモンの場合は、ローカルのクライアント・ベースの **clhosts** ファイルに、PowerHA SystemMirror サーバー・ノード用の IP アドレスのリストを含む必要があります。このようにして、特定の PowerHA SystemMirror サーバー・ノードが使用できない場合 (電源オフなど) でも、クライアント・ノード上のデーモンが他の PowerHA SystemMirror サーバー・ノードと通信できます。

PowerHA SystemMirror 検証ユーティリティーは、クライアント・ベースの **clhosts** ファイルに IP アドレスを取り込む場合に役立ちます。IP アドレスを取り込むには、使用可能なすべての PowerHA SystemMirror サーバー・ノードを検出し、サーバー・ノード上に `/usr/es/sbin/cluster/etc/clhosts.client` ファイルを作成し、そのファイルに PowerHA SystemMirror サーバー・ノードの IP アドレスを取り込みます。

クラスター上の PowerHA SystemMirror の検証と同期化を終了した後で、この **clhosts.client** ファイルを `/usr/es/sbin/cluster/etc/clhosts` として各クライアント・ノードに手作業でコピーする必要があります (名前を変更して、**.client** という拡張子を除去します)。

関連概念:

116 ページの『PowerHA SystemMirror 検証および同期化』

PowerHA SystemMirror の検証および同期化のプロセスで検証される内容は、AIX システム・ファイルに対する特定の PowerHA SystemMirror の変更が正しいこと、クラスターとそのリソースが正しく構成されていること、セキュリティ (セットアップされている場合) が正しく構成されていること、クラスター・トポロジー、ネットワーク構成、および PowerHA SystemMirror リソースの所有権とテークオーバーについてすべてのノードが一致していることです。

トラブルシューティング・ツール

通常、機能している PowerHA SystemMirror クラスターに介入する必要はほとんどありません。ただし問題が発生した場合は、診断および回復のスキルが必須です。そのため、トラブルシューティングを行うときは、問題を素早く特定し、PowerHA SystemMirror ソフトウェアの知識を応用して、クラスターを完全な稼働状態まで復元する必要があります。

関連概念:

121 ページの『クラスター・テスト・ツール』

クラスター・テスト・ツールは、PowerHA SystemMirror クラスター構成をテストし、ノードがアクセス不可能になった場合、ネットワークがアクセス不可能になった場合、リソース・グループがあるノードから別のグループに移動した場合などのような、一連の指定の事情の下でクラスターがどのように動作するかについて評価できるユーティリティです。

ログ・ファイル

PowerHA SystemMirror ソフトウェアは、それが生成するメッセージをシステム・コンソールといくつかのログ・ファイルに書き出します。各ログ・ファイルの詳細さのレベルはそれぞれ異なりますから、システム管理者は各種のログ・ファイルを見ることによって、PowerHA SystemMirror 処理の種々の局面を検討することができます。

主なログ・ファイルには以下のものがあります。

- **/var/hacmp/adm/cluster.log** ファイルでは、クラスター・イベントが追跡されます。
- **/var/hacmp/log/hacmp.out** ファイルでは、構成スクリプトの実行時に生成される出力を記録します。イベント・サマリーは、クラスター・マネージャーが開始したイベントの詳細出力のあとに出力され、**hacmp.out** ファイルで重要な情報を調べる助けになります。さらに、イベント・サマリーは、**hacmp.out** ファイル内の対応するイベントに HTML リンクを提供します。
- **/var/hacmp/adm/history/cluster.mmddyyyy** ログ・ファイルには、毎日のクラスター・ヒストリーが記録されます。
- **/var/hacmp/clverify/clverify.log** ファイルには、検証時の詳細メッセージ出力が記録されます。クラスターの検証では、さまざまな PowerHA SystemMirror 構成に対して一連の検査が実行されます。それぞれの検査で、クラスターの整合性の問題またはエラーの検出を試みます。検証ユーティリティのメッセージ出力では、エラーが発生した場所 (例えば、ノード、デバイス、コマンドなど) を示します。

PowerHA SystemMirror では、ログ・ファイルのパラメーターを表示、リダイレクト、保存、および変更できるため、そのパラメーターを調整して自分のニーズに合わせるすることができます。

関連情報:

Administering PowerHA SystemMirror

PowerHA SystemMirror チューナブル値のリセット

クラスターを構成およびテストする間、クラスター・パフォーマンスに影響を及ぼす PowerHA SystemMirror チューナブル値の 1 つについて値を変更することができます。または、構成の他の面を変更せずに、これらのチューナブル値をデフォルト設定にリセットすることもできます。

サード・パーティーのクラスター管理者またはコンサルタントは、自分が構成したのではないクラスターの管理を引き継ぐよう依頼され、チューナブル値をデフォルトにリセットする必要があることがあります。SMIT インターフェースを使用して、クラスター・チューナブル値をリセットできます。PowerHA SystemMirror は、リセットの前に、クラスター・スナップショットを取ります。値がデフォルトにリセットされた後で、カスタマイズしたクラスター設定に戻る場合は、クラスター・スナップショットを適用できます。クラスター・チューナブル値をリセットすると、クラスター構成データベース内の情報がリセットされます。リセットまたは除去される情報は以下のカテゴリーで構成されます。

- ユーザーが提供する情報 (例えば、イベント前処理およびイベント後処理スクリプトと、ネットマスクのようなネットワーク・パラメーター)。クラスター・チューナブル値のリセットは、既に構成してあるイベント前処理およびイベント後処理スクリプトを除去しないことに注意してください。しかし、チューナブル値をリセットすると、PowerHA SystemMirror のイベント前処理とイベント後処理スクリプトの知識は構成から除去され、これらのスクリプトはクラスター内でリソースを管理する PowerHA SystemMirror によってもはや使用されません。必要であれば、これらのスクリプトを再使用するように PowerHA SystemMirror を再構成することができます。
- 構成と同期化の間に PowerHA SystemMirror によって自動的に生成された情報。これにはノードとネットワーク ID、ネットマスクのようなオペレーティング・システムから発見された情報が含まれます。通常、ユーザーは生成された情報を見ることができません。

関連情報:

Administering PowerHA SystemMirror

PowerHA SystemMirror インストール・ガイド

クラスター・ステータス情報ファイル

(各クラスター・ノードから見た) クラスター構成のレコードを保管するために PowerHA SystemMirror クラスター・スナップショット・ユーティリティを使用すると、オプションでこのユーティリティに数多くの標準 AIX コマンドおよび PowerHA SystemMirror コマンドを実行させ、そのクラスターのステータス情報を取得できます。この情報は、スナップショット・ディレクトリー内の **.info** 拡張子で識別されるファイルに格納されます。

この snapshots ディレクトリーは、SNAPSHOTPATH 環境変数の値により定義されます。デフォルトで、クラスター・スナップショット・ユーティリティには、**cllsif**、**cllsnw**、**df**、**ls**、および **netstat** のようなコマンドからの出力が含まれます。**.info** ファイルに格納する追加情報を指定するカスタム・スナップショット方式を作成できます。

クラスター・スナップショットを使用すれば、スナップショットにログ・ファイルを保存する処理をスキップできます。クラスター・スナップショットは、クラスター構成情報を記録するために使用できますが、クラスターはクラスターの運用を記録するだけで、構成情報は記録しません。デフォルトで、PowerHA SystemMirror は、クラスター・スナップショットを作成するときにクラスター・ログ・ファイルを収集しませんが、SMIT でログを収集するように指定することはできます。ログ・コレクションをスキップすると、スナップショットのサイズが減少し、スナップショット・ユーティリティの実行が高速化します。クラスター・スナップショットのサイズは構成に左右されます。例えば、基本的な 2 ノード構成にはおおよそ 40KB が必要です。

自動エラー通知

AIX エラー通知機能を使用すると、PowerHA SystemMirror ソフトウェアでは特にモニターされないイベントを検出することができます。例えば、ディスク・アダプターの障害などを検出ことができ、イベント発生時に行われる処理を指定することができます。

通常、エラー通知メソッドは、1 つずつ手動で定義します。PowerHA SystemMirror は、重大なエラーに関して SMIT インターフェースからワンステップで自動的に「オン」に設定できる事前指定通知メソッドのセットを提供します。これにより、それぞれの通知メソッドを手動で定義する必要がなくなり、時間と労力が大幅に節約されます。

カスタム・リモート通知

クラスター・イベントに応じてカスタマイズした通知メソッドを発行するために、SMIT インターフェースによって `notification` メソッドを定義することができます。テキスト・メッセージ通知を、携帯電話を含む任意のアドレス宛てに送信したり、E メール・アドレス宛てにメールで送信したりできます。

リモート通知メソッドの構成後に、テスト・メッセージを送信して構成が正しく行われていることを確認できます。

各種イベントのために、さまざまなテキスト・メッセージとダイヤル先電話番号を伴った通知メソッドを任意の数だけ構成できます。通知をトリガーする可能性のあるイベントすべてに応えるのに十分な情報を、関連テキスト・メッセージが伝える限り、同じ通知メソッドを、複数の異なるイベントのために使用できます。

イベント・プリアンブルおよび要約

クラスター・イベントについて詳しくは、`hacmp.out` ファイルに記録されます。このファイルの冗長出力には何行ものイベント情報が含まれ、各イベントの詳細の最後には簡潔な要約が記されています。クラスターに最近何が発生したかを素早く効率的にチェックするには、SMIT の「**Display Event Summaries** (イベント要約の表示)」パネルを使用して、現在と以前の `hacmp.out` ファイルのイベント要約部分の編集を見ることができます。

また、編集したイベント要約を指定したファイルに保管することもできます。オプションで、イベント・サマリーは、`hacmp.out` ファイル内の対応するイベントに HTML リンクを提供します。クラスター・マネージャーは、各イベントの処理にどのリソース・グループがエンキューされたかを示すプリアンブルも出力します。これによって、後続の処理の順序を知ることができます。

関連情報:

PowerHA SystemMirror のトラブルシューティング

クラスター・テスト・ツール

クラスター・テスト・ツールは、PowerHA SystemMirror クラスター構成をテストし、ノードがアクセス不可能になった場合、ネットワークがアクセス不可能になった場合、リソース・グループがあるノードから別のグループに移動した場合などのような、一連の指定の事情の下でクラスターがどのように動作するかについて評価できるユーティリティです。

テストを開始し、席をはずしている間にそれを実行し、あとで戻ってテストの結果を評価することができます。

トポロジーとリソース・グループ管理用に基本クラスター・テストの自動化スイートを実行する場合、SMIT から自動化されたテスト・スイートを実行できます。PowerHA SystemMirror 管理者の経験があり、クラスター・テストを環境に調整する必要がある場合は、SMIT から実行できるカスタム・テストを作成することもできます。

随時ツールを実行することを推奨します。例えば、最初に PowerHA SystemMirror を構成してからクラスターを実稼働環境に移行する前、クラスターのサービス停止中にクラスター構成変更を行ったあと、またはクラスターが正常に機能していると思われる場合にも定期的な間隔で、ツールを実行します。

特記事項

本書は米国が提供する製品およびサービスについて作成したものです。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119

Armonk, NY 10504-1785

US

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

記載されている性能データとお客様事例は、例として示す目的でのみ提供されています。実際の結果は特定の構成や稼働条件によって異なります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者にお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

表示されている IBM の価格は IBM が小売り価格として提示しているもので、現行価格であり、通知なしに変更されるものです。卸価格は、異なる場合があります。

本書はプランニング目的としてのみ記述されています。記述内容は製品が使用可能になる前に変更になる場合があります。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名前はすべて架空のものであり、名前や住所が類似する個人や企業が実在しているとしても、それは偶然にすぎません。

著作権使用許諾:

本書には、様々なオペレーティング・プラットフォームでのプログラミング手法を例示するサンプル・アプリケーション・プログラムがソース言語で掲載されています。お客様は、サンプル・プログラムが書かれているオペレーティング・プラットフォームのアプリケーション・プログラミング・インターフェースに準拠したアプリケーション・プログラムの開発、使用、販売、配布を目的として、いかなる形式においても、IBM に対価を支払うことなくこれを複製し、改変し、配布することができます。このサンプル・プログラムは、あらゆる条件下における完全なテストを経ていません。従って IBM は、これらのサンプル・プログラムについて信頼性、利便性もしくは機能性があることをほめかしたり、保証することはできません。これらのサンプル・プログラムは特定物として現存するままの状態を提供されるものであり、いかなる保証も提供されません。IBM は、お客様の当該サンプル・プログラムの使用から生ずるいかなる損害に対しても一切の責任を負いません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物には、次のように、著作権表示を入れていただく必要があります。

© (お客様の会社名) (西暦年).

このコードの一部は、IBM Corp. のサンプル・プログラムから取られています。

© Copyright IBM Corp. _年を入れる_.

プライバシー・ポリシーに関する考慮事項

サービス・ソリューションとしてのソフトウェアも含めた IBM ソフトウェア製品（「ソフトウェア・オフアリング」）では、製品の使用に関する情報の収集、エンド・ユーザーの使用感の向上、エンド・ユーザーとの対話またはその他の目的のために、Cookie はじめさまざまなテクノロジーを使用することがあります。多くの場合、ソフトウェア・オフアリングにより個人情報が収集されることはありません。IBM の「ソフトウェア・オフアリング」の一部には、個人情報を収集できる機能を持つものがあります。ご使用の「ソフトウェア・オフアリング」が、これらのCookie およびそれに類するテクノロジーを通じてお客様による個人情報の収集を可能にする場合、以下の具体的事項を確認ください。

この「ソフトウェア・オフアリング」は、Cookie もしくはその他のテクノロジーを使用して個人情報を収集することはありません。

この「ソフトウェア・オフアリング」が Cookie およびさまざまなテクノロジーを使用してエンド・ユーザーから個人を特定できる情報を収集する機能を提供する場合、お客様は、このような情報を収集するにあたって適用される法律、ガイドライン等を遵守する必要があります。これには、エンドユーザーへの通知や同意の要求も含まれますがそれらには限られません。

このような目的での Cookie などの各種テクノロジーの使用については、『IBM オンラインでのプライバシー・ステートメントのハイライト』(<http://www.ibm.com/privacy/jp/ja/>)、『IBM オンラインでのプライバシー・ステートメント』(<http://www.ibm.com/privacy/details/jp/ja/>) の『クッキー、ウェブ・ビーコン、その他のテクノロジー』というタイトルのセクション、および『IBM Software Products and Software-as-a-Service Privacy Statement』(<http://www.ibm.com/software/info/product-privacy>) を参照してください。

商標

IBM、IBM ロゴおよび [ibm.com](http://www.ibm.com) は、世界の多くの国で登録された International Business Machines Corp. の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

UNIX は、The Open Group の米国およびその他の国における登録商標です。

索引

日本語, 数字, 英字, 特殊文字の順に配列されています。なお, 濁音と半濁音は清音と同等に扱われています。

[ア行]

アプリケーション 44
 障害として除去 71
 モニター 114
アプリケーション可用性分析ツール 71, 116
アプリケーションの可用性
 概説 66
アプリケーション・モニター 71
 clam_nfsv4 114
永続ノード IP ラベル 116

[カ行]

概説
 アプリケーションの可用性 66
カスタム・リモート通知 121
クライアント 18
クラスター
 イベント 87
 拡張 30
 クライアント 18
 サイト 19
 スナップショット 111
 セキュリティ 107
 ソフトウェア 55
 ディスク・デバイス 13
 テスト・ツール 121
 ネットワーク 13, 32
 ハードウェア 55
 物理コンポーネント 11
 マルチサイト・ソリューション 28
 リソース・グループ 45
 リンク 29
 C-SPOC 83
 IP アドレス・テークオーバー 35
 node 12, 18
クラスター情報プログラム 61
クラスター・サービス
 開始 109
 再始動 109
 停止 109
クラスター・ステータス情報ファイル 120
クラスター・マネージャー 57
 SNMP 60

検査
 カスタム・メソッド 118
 自動クラスター構成モニターの使用 117
 修正アクションの使用 117
構成
 オプション 106
 拡張構成パス 107
 サイト間 LVM ミラー 100
 スタンバイ 90
 例 90, 92
 多層アプリケーション 97
 テークオーバー 94
 相互 94
 ワンサイド 94
 2 ノード相互 96
 8 ノード相互 96
 動的 LPAR 101
 標準構成パス 107
 リソース・グループ・ロケーション依存関係 99
 Smart Assist 107
高速ディスク・テークオーバー 85
コンカレント・リソース・マネージャー 65

[サ行]

サービス IP アドレス 35, 45
サービス IP ラベル 35, 45
災害時リカバリー 86
最小限に抑える
 スケジュールされたダウン時間 77
 テークオーバー時間 85
 予定外のダウン時間 84
最大限に行う
 災害時リカバリー 86
サイト 19, 54
 ソリューション 28
サイト間 LVM ミラーリング 86
サブネット 34
自動エラー通知 121
除去
 単一障害点 66
 アプリケーション 71
 通信インターフェース 72
 ディスク 76
 ディスク・アダプター 76
 ネットワーク 74
 node 67
スナップショット 111
セキュリティ 107
ソフトウェア 55, 56
 コンポーネント 58

ソフトウェア (続き)
補足 65

[タ行]

チューナブル値
リセット 120
通信インターフェース 34
障害として除去 72
通信デバイス 34
テープ 45
ディスク
障害として除去 76
ディスク・アクセス
コンカレント 63
非コンカレント 63
ディスク・アダプター
障害として除去 76
ディスク・デバイス 13
動的自動再構成 78
トラブルシューティング
イベント要約 121
イベント・プリアンブル 121
カスタム・リモート通知 121
クラスター・ステータス情報ファイル 120
クラスター・テスト・ツール 121
自動エラー通知 121
チューナブル値のリセット 120
ログ・ファイル 119

[ナ行]

ネットワーク 13, 32
サービス IP アドレス 35
サービス IP ラベル 35
サブネット 34
障害として除去 74
通信インターフェース 34
通信デバイス 34
ハートビート 37
物理的 32
論理 32
IP エイリアス 35
ネットワーク・スイッチ
マルチキャスト 39

[ハ行]

ハードウェア 55
ハートビート 37
Point-to-Point ネットワーク 39
TCP/IP ネットワーク 38
ファイルシステム 43
ファイル・コレクション 112
物理ネットワーク 32

分割構成 20
ボリューム・グループ 43

[マ行]

マージ・ポリシー 22
マルチキャスト 39
経路指定 41
ネットワーク・スイッチ 39
パケット通信 39
IGMP (Internet Group Management Protocol) 40
モニター
アプリケーション 114
アプリケーション可用性分析ツール 116
永続ノード IP ラベル 116
クラスター情報プログラム 113
クラスター・ステータス・ユーティリティ 115
クラスター・マネージャー 113
clhosts ファイル 118

[ラ行]

リセット
チューナブル値 120
リソース
アプリケーション 44
サービス IP アドレス 45
サービス IP ラベル 45
テープ 45
ファイルシステム 43
ボリューム・グループ 43
論理ボリューム 43
リソース・グループ 45
管理 81
管理ユーティリティ 112
サイト 54
始動 47, 49
整定時間 49
動的ノード優先順位 49
フォールオーバー 47, 49
フォールバック 47, 49
フォールバック・タイマー・フォールバック 49
ポリシー 48
ロケーション依存関係 53
リポジトリ・ディスク
障害 17
マルチパス 16
例
スタンバイ構成 90, 92
ログ 119
アナライザー 55
論理ネットワーク 32
論理ボリューム 43

C

clam_nfsv4 114
clhosts 118
Clinfo 113
clinfo 61
clRGmove 81, 112
clstat 115
C-SPOC 83, 110

D

DARE
参照：動的自動再構成

H

hacmp.out 121

I

IP アドレス・テークオーバー 35
IP エイリアス 35

N

NFS 62
node 12, 18
障害として除去 67

P

PowerHA SystemMirror
マルチキャスト 39

S

Smart Assist 107
SMIT 110
SNMP 60, 113

[特殊文字]

/.rhosts 107



Printed in Japan

日本アイ・ビー・エム株式会社

〒103-8510 東京都中央区日本橋箱崎町19-21