



# Geo Clusteringのクイックスタート

---

## Geo Clustering for SUSE Linux Enterprise High Availability Extension 12 SP2

Geo Clustering for SUSE Linux Enterprise High Availability Extensionという別個の拡張を使用することで、無制限の距離にまたがってHigh Availabilityクラスタを使用できるようになります。Geoクラスタリングを使用すると、それぞれ1つのローカルクラスタを備えた地理的に分散された複数のサイトを運用できます。これらのクラスタ間のフェールオーバーは、より高いレベルのエンティティであるブースデーモン(`boothd`)によって管理されます。このドキュメントでは、2つのサイトからなる単一Geoクラスタ、1つのアービトラータ、およびDRBDを通じたデータレプリケーションで構成されるサンプルシナリオをセットアップする方法を説明します。

発行日: 19/10/2016

## 目次

- 1 拡張としてのインストール 3
- 2 Geoクラスタの課題 5
- 3 概念の概要 5
- 4 要件 8
- 5 サンプルシナリオと基本ステップ – 概要 9
- 6 ブースサービスの設定 10
- 7 DRBDの設定 22
- 8 すべてのサイトとアービトラータ間の設定ファイルの同期 28
- 9 クラスタリソースと制約の設定 34
- 10 DNSアップデートを通じたIPアドレス再割り当ての設定 43

- 11 Geoクラスタの管理 45
- 12 トラブルシューティング 52
- 13 最新の製品バージョンへのアップグレード 53
- A GNU Licenses 57

# 1 拡張としてのインストール


High Availability ExtensionおよびGeo Clustering for SUSE Linux Enterprise High Availability Extensionを使用するには、次のインストールパターンに含まれているパッケージが必要です。

- 高可用性
- Geo Clustering for High Availability



## 注記: アービトレータのパッケージ要件

Geoクラスタのセットアップに1つ以上のアービトレータが含まれている場合は(アービトレータを参照)、それらのアービトレータには Geo Clustering for High Availability パターンのみが必要です。このパターンのインストール方法については、1.2項「アービトレータへのパッケージのインストール」を参照してください。

両方のパターンを使用できるための条件は、お使いのシステムをSUSEカスタマセンター(またはローカル登録サーバ)に登録済みであること、かつそれぞれの製品チャネルまたはインストールメディアを拡張として追加済みであることです。拡張のインストール方法については、<http://www.suse.com/documentation/>  で入手できる『SUSE Linux Enterprise 12 SP2導入ガイド』を参照してください。「モジュール、拡張機能、サードパーティ製アドオン製品のインストール」の章を参照してください。

## 1.1 クラスタノードへのパッケージのインストール

High Availability ExtensionとGeo Clustering for SUSE Linux Enterprise High Availability Extensionの両方が拡張として追加されているが、パッケージがまだインストールされていない場合は、以下の手順を実行します。

1. コマンドラインを通じて両方のパターンからパッケージをインストールするには、Zypperを使用します。

```
sudo zypper in -t pattern ha_sles ha_geo
```

2. 代替の方法として、グラフィカルインタフェースを通じてインストールするには、YaSTを使用します。
  - a. YaSTを root ユーザとして開始し、[ソフトウェア] > [ソフトウェア管理] の順に選択します。

b. [View (表示)] > [Patterns (パターン)]の順にクリックして、以下のパターンをアクティブ化します。

- 高可用性
- Geo Clustering for High Availability

c. [同意する]をクリックして、パッケージのインストールを開始します。

## ! 重要: すべてのパーティへのソフトウェアパッケージのインストール

High AvailabilityクラスタおよびGeoクラスタに必要なソフトウェアパッケージは、クラスタノードに自動的にコピー「されません」。

- SUSE Linux Enterprise Server 12 SP2と、High AvailabilityパターンおよびGeo Clustering for High Availabilityパターンを、Geoクラスタの一部となる「すべての」マシンにインストールします。
- クラスタの一部となるすべてのノードにパッケージを手動でインストールすることを希望しない場合は、AutoYaSTを使用して既存ノードのクローンを作成します。詳細については、SUSE Linux Enterprise High Availability Extension 12 SP2の『Administration Guide』を参照してください(<http://www.suse.com/documentation/> から入手できます)。「Installation and Basic Setup」という章の「Mass Deployment with AutoYaST」という節を参照してください。Geoクラスタリング拡張が必要なすべてのマシンについて、現時点ではGeoクラスタ用のパッケージを手動でインストールする必要があります。AutoYaSTでは、Geo Clustering for SUSE Linux Enterprise High Availability Extensionはまだサポートされていません。

## 1.2 アービトラータへのパッケージのインストール

1. アービトラータとして機能するマシンにGeo Clustering for SUSE Linux Enterprise High Availability Extensionが追加されていることを確認します。
2. 各アービトラータにログインして、次のコマンドを使用してパッケージをインストールします。

```
sudo zypper in -t pattern ha_geo
```

または、YaSTを使用して Geo Clustering for High Availability パターンをインストールします。

## 2 Geoクラスタの課題

一般に、Geo環境は互いに遠く離れているため、サイト間の同期通信をサポートできません。これが次のような課題につながります。

- クラスタサイトが始動しており、実行していることを確認する方法は？
- リソースが1度だけ始動されるようにする方法は？
- クォーラムを異なるサイト間で共有可能にして、スプリットブレインシナリオを回避可能にする方法は？
- すべてのノードとサイトでCIBを最新状態に保つ方法は？
- サイト間のフェールオーバーを管理する方法は？
- 停止が必要なリソースがある場合、大幅な遅延時間に対処する方法は？

以降のセクションでは、SUSE Linux Enterprise High Availability Extensionを使用してこれらの課題に対応する方法を説明します。

## 3 概念の概要

SUSE Linux Enterprise High Availability Extensionを基にしたGeoクラスタは、各クラスタサイトが従来のクラスタ内の1つのクラスタノードに対応する「オーバーレイ」クラスタであると考えられます。オーバーレイクラスタはブース構造によって管理されます。この方式では、複数のクラスタサイトに対するクラスタリソースの高可用性が保証されます。これは、サイトをダウン状態にする必要がある場合、クラスタサイト間のフェールオーバードメインとして扱われるチケットと呼ばれるクラスタオブジェクトを使用することで可能になります。ブースは、すべてのチケットがどの時点においても1つのサイトのみによって所有されることを保証します。

以下では、Geoクラスタ向けに導入された個別のコンポーネントとメカニズムについて詳しく説明します。

## チケット

チケットは指定のクラスタサイトの特定のリソースを実行する権利を付与します。チケットは1度に1つのサイトだけが所有できます。最初はどのサイトもチケットを持ちません。クラスタ管理者が各チケットを1度付与する必要があります。その後、チケットはブースによってリソースの自動フェールオーバー用に管理されます。ただし、管理者は手動でチケットに介入して、付与または取り消しを行えます。

チケットが管理目的のために取り消された場合、そのチケットはブースによって管理されなくなります。チケットがブースによって再び管理されるようにするには、そのチケットをいずれかのサイトに再び付与する必要があります。

リソースを特定のチケットに依存させることができます。定義されたチケットがサイトで使用できる場合のみそれぞれのリソースが始動します。その逆に、チケットが削除されると、そのチケットに依存するリソースが自動的に停止します。

サイトにおけるチケットの有無はクラスタステータスとしてCIBに保存されます。特定のチケットについては、サイトに対して `true` (サイトにチケットがある) または `false` (サイトにチケットがない) の2つの状態のみとなります。特定のチケットがない状況(Geoクラスタの初期状態時)は、チケットが取り消された後の状況と同じように扱われます。どちらの状況も `false` という値で表されます。

オーバレイクラスタ内のチケットは従来のクラスタ内のリソースに似ています。ただし、従来のクラスタとは違って、チケットはオーバレイクラスタ内の唯一のリソースタイプです。これらは、設定やクローン作成の必要がないプリミティブリソースです。

## ブースクラスタチケットマネージャ

ブースは、チケットの配布を管理するインスタンスであるため、Geoクラスタのサイト間のフェールオーバープロセスとなります。それぞれの参加クラスタおよびアービトラータが、サービスである `boothd` を実行します。これは、別のサイトで実行しているブースデーモンに接続し、接続性の詳細を交換します。チケットがサイトに付与された後は、ブースメカニズムによってそのチケットを自動的に管理できます。そのチケットを保有しているサイトが稼働していない場合、ブースデーモンは、他のサイトのうち、どのサイトがそのチケットを取得するのかを投票します。短い接続の不具合を防ぐため、(明示的または、投票本体から接続が外れることで暗示的に)投票を失ったサイトはタイムアウト後にチケットを放棄する必要があります。このため、チケットは前のサイトによって放棄された後にのみ再配布されるようになっています。[Dead Man Dependency \(`loss-policy=fence`\)](#) も参照してください。

## アービトラータ

各サイトが、他のサイトとの通信に責任を持つブースインスタンスを1つ実行します。偶数のサイトを設定している場合は、サイト全体のリソースのフェールオーバーなど、決定について合意状態に達するための追加インスタンスが必要です。この場合、追加サイトで実行するアービトレー

タを1つ以上追加します。アービトレータは専用モードでブースインスタンスを実行する単一のマシンです。すべてのブースインスタンスは互いに通信するため、アービトレータによりチケットの付与または取り消しについてより信頼性の高い決定が可能となります。アービトレータはどのチケットも保有できません。

アービトレータは、2つのサイトのシナリオで特に重要です。たとえば、サイト A がサイト B と通信できなくなった場合、考えられる原因は2つあります。

- A と B の間にネットワーク障害が起こっている。
- サイト B がダウンしている。

ただし、サイト C (アービトレータ) はサイト B と通信できているため、サイト B は起動したまま、実行する必要があります。

#### チケットフェールオーバー

チケットが失われた場合、すなわち他のブースインスタンスがそのチケットの所有者と一定時間にわたって通信できなかった場合は、残りのサイトのいずれかがそのチケットを取得します。これはチケットフェールオーバーと呼ばれます。残りのメンバーが多数派を構成できない場合は、そのチケットはフェールオーバーできません。

#### Dead Man Dependency (`loss-policy="fence"`)

チケットが取り消された後、特にリソースがカスケード接続されている場合、そのチケットに依存するすべてのリソースが停止するまで長い時間がかかります。プロセスを短縮するため、クラスタ管理者は、サイトからチケットが取り消された場合に備えて、`loss-policy` を(チケット依存性と一緒)に設定できます。`loss-policy` が `fence` に設定される場合、依存するリソースをホストしているノードはフェンスされます。



#### 警告: データ喪失の可能性

一方では、`loss-policy="fence"` という設定によって、クラスタのリカバリプロセスが大幅に高速化されて、リソースをより迅速に移行することが可能になります。

その一方で、この設定が原因で、たとえば次のような未書き込みデータがすべて失われる可能性があります。

- 共有ストレージ(DRBDなど)に配置されたデータ
- 低速なネットワークリンクが原因で他方のサイトにまだ達していないレプリケーションデータベース(MariaDBやPostgreSQLなど)内のデータ

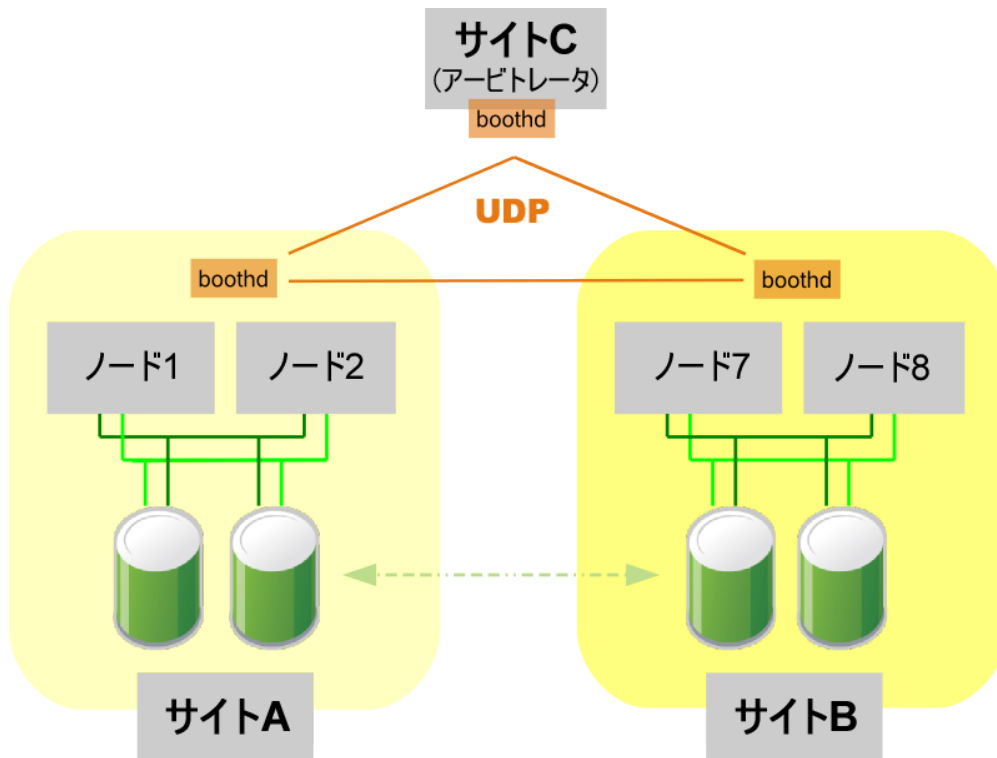


図 1 2サイトクラスタ(4ノード + アービトレータ)

おそらく最も一般的なシナリオは、2つのサイトと、3つ目のサイト上の単一のアービトレータで構成されるGeoクラスタです。これには3つのブースインスタンスが必要です。[図1「2サイトクラスタ\(4ノード + アービトレータ\)」](#)を参照してください。上限は(現時点では)16個のブースインスタンスです。

通常どおり、CIBは各クラスタ内で同期されますが、Geoクラスタのサイト間で自動的に同期されることはありません。ただし、SUSE Linux Enterprise High Availability Extension 12以降では、他のクラスタサイトへのリソース設定の転送は以前より簡単です。詳細については、[9.3項「他のクラスタサイトへのリソース設定の転送」](#)を参照してください。

## 4 要件

### ソフトウェアの必要条件

- Geoクラスタの一部となるすべてのクラスタは、SUSE Linux Enterprise High Availability Extension 12 SP2に基づいている必要があります。



- SUSE® Linux Enterprise Server 12 SP2を、すべてのアービトレータにインストールする必要があります。
- Geo Clustering for SUSE Linux Enterprise High Availability Extensionは、すべてのクラスタノードと、Geoクラスタの一部となるすべてのアービトレータにインストールする必要があります。

#### ネットワーク要件

- ブースインスタンスあたり1つのUDPポートと1つのTCPポートを通じて、サイトにアクセスする必要があります。すなわち、間に配置されているすべてのファイアウォールとIPsecトンネルをこの要件に合わせて設定する必要があります。
- 他のセットアップ判断によって、さらに多くのポートを開く必要が生じることがあります(たとえばDRDBやデータベースレプリケーション用など)。

#### その他の要件と推奨事項

- すべてのサイト上のすべてのクラスタノードはクラスタ外のNTPサーバと同期する必要があります。詳細については、『SUSE Linux Enterprise Server 12 SP2 管理ガイド』(<http://www.suse.com/documentation/> から入手可能)を参照してください。特に、「Time Synchronization with NTP」の章を参照してください。  
ノードが同期されていない場合、ログファイルまたはクラスタレポートは分析が難しくなります。

## 5 サンプルシナリオと基本ステップ – 概要

以降の節では、下記のようなサンプルシナリオを採用します。

例 1 シナリオ: 2つのサイトからなるクラスタ、1つのアービトレータ、およびDRBDを介したデータレプリケーション

- 2つのサイト(amsterdamとberlin)および1つのアービトレータからなるGeoクラスタ。
- 各サイトには、もう一方のサイトにルーティングされるプライベートネットワークがあります。
  - amsterdam: 192.168.201.x
  - berlin: 192.168.202.x
- 各サイトでは、2つのノードからなるクラスタが運用されています。
  - クラスタ amsterdam を構成するノードは alice と bob
  - クラスタ berlin を構成するノードは charly と doro

- 障害復旧用の非同期モードのDRBDを使用して、データがサイト間で複製されます。
- ブース設定と他の重要な設定ファイルは、Csync2を使用してクラスタサイトとアービトラータの間で同期されます。

このシナリオをセットアップするには、以下の基本ステップを実行する必要があります。

### ブースサービスの設定

1. デフォルトのブースセットアップまたは複数のテナント用のブースセットアップのどちらを使用するかを選択。
2. すべてのサイトとアービトラータに対するブース設定の同期。
3. ブース用のクラスタリソースの設定(チケットの依存関係、ブースの制約とリソースを参照)。
4. 他のクラスタサイトへのリソース設定の転送。
5. ブースサービスの有効化と開始。

### DRBDの設定

1. DRBDの設定(DRBDの設定を参照)。
2. DRBD用のクラスタリソースの設定(DRBDのリソースと制約を参照)。
3. 他のクラスタサイトへのリソース設定の転送。
4. DRBD設定ファイルの同期(Csync2を使用した変更内容の同期を参照)。

### すべてのサイトとアービトラータ間の設定ファイルの同期

1. Csync2のセットアップ(Geoクラスタ用のCsync2設定を参照)。
2. サイトおよびアービトラータ間でのすべての関連設定ファイルの初期同期(Csync2を使用した変更内容の同期を参照)。

## 6 ブースサービスの設定

デフォルトのブース設定は `/etc/booth/booth.conf` です。このファイルは、Geoクラスタのすべてのサイト(アービトラータを含む)上で同じである必要があります。すべてのサイトとアービトラータの間でブース設定を常に同期させるには、Csync2を使用します(6.3項「すべてのサイトとアービトラータに対するブース設定の同期」を参照)。



## 注記: /etc/boothおよびファイルの所有権

/etc/booth ディレクトリとこのディレクトリ内のすべてのファイルは、ユーザ hacluster とグループ haclient に属している必要があります。このディレクトリから新しいファイルをコピーする際は必ず、cp コマンドで -p オプションを使用して所有権を維持してください。または、新しいファイルを作成する際は、その後に chown hacluster:haclient FILE を使用してユーザとグループを設定します。

複数のGeoクラスタを含むセットアップの場合は、同じアービトラータを「共有」できます(SUSE Linux Enterprise High Availability Extension 12以降)。複数のブース設定ファイルを提供することで、複数のブースインスタンスを同じアービトラータ上で開始することができ、それぞれのブースインスタンスを異なるポート上で動作させることができます。このようにすることで、「1台の」マシンを「複数の」Geoクラスタ用のアービトラータとして機能させることができます。複数のGeoクラスタ向けにブースを設定する方法の詳細については、[6.2項「複数のテナント用のブースセットアップ」](#)を参照してください。

悪質なユーザによってブースサービスが妨害されることを防止するために、ブースと通信する際に共有鍵に基づいた認証を要求するように設定できます。詳細については、[⑤](#) 内の[例2「ブース設定ファイル」](#)を参照してください。各種のブースサーバと通信するすべてのホストはこの鍵を必要とします。したがって、Csync2設定に鍵ファイルを含めるか、鍵ファイルをすべての関係先の間で手動で同期させる必要があります。

## 6.1 デフォルトのブースセットアップ

ブースに必要なすべてのパラメータを設定するには、ブース設定ファイルを手動で編集するか、YaSTの[Geo Cluster (Geoクラスタ)] モジュールを使用して編集します。このYaSTモジュールにアクセスするには、コマンドラインから yast2 geo-cluster コマンドを使用してこのモジュールを起動します(またはYaSTを起動して、[High Availability (高可用性)] > [Geo Cluster (Geoクラスタ)] の順に選択します)。

例 2 ブース設定ファイル

```
transport = UDP ①
port = 9929 ②
arbitrator = 147.2.207.14 ③
site = 192.168.201.151 ④
site = 192.168.202.151 ④
authfile = /etc/booth/authkey ⑤
ticket = "ticket-nfs" ⑥
    expire = 600 ⑦
    timeout = 10 ⑧
    retries = 5 ⑨
```

```

renewal-freq = 30 ⑩
before-acquire-handler ⑪ = /etc/booth/ticket-nfs ⑫ ms_drbd_nfs ⑬
acquire-after = 60 ⑭
ticket = "ticketA" ⑥
  expire = 600 ⑦
  timeout = 10 ⑧
  retries = 5 ⑨
  renewal-freq = 30 ⑩
  before-acquire-handler ⑪ = /etc/booth/ticket-A ⑫ db-1 ⑬
  acquire-after = 60 ⑭
ticket = "ticketB" ⑥
  expire = 600 ⑦
  timeout = 10 ⑧
  retries = 5 ⑨
  renewal-freq = 30 ⑩
  before-acquire-handler ⑪ = /etc/booth/ticket-B ⑫ db-8 ⑬
  acquire-after = 60 ⑭

```

- ① サイト間の通信に使用するトランスポートプロトコル。UDPのみがサポートされていますが、将来的には他のトランスポート層も追従する予定です。したがって、現時点ではこのパラメータは省略可能です。
- ② 各サイトのブースインスタンス間の通信用に使用されるポート。デフォルトポート(9929)を使用しない場合は、異なるサービス用にまだ使用されていないポートを選択してください。ノードおよびアービトラータのファイアウォールで必ずポートを解放します。ブースクライアントはTCPを使用して `boothd` と通信します。ブースは常にUDPポートとTCPポートの両方をバインドおよびリッスンします。
- ③ アービトラータとして使用するマシンのIPアドレス。Geoクラスタのセットアップで使用するアービトラータごとに、エントリを追加します。
- ④ サイト上の `boothd` 用に使用されるIPアドレス。Geoクラスタのセットアップで使用するサイトごとに、エントリを追加します。各サイトの正しい仮想IPアドレス( `IPaddr2` )を挿入してください。そうしない場合、ブースメカニズムが正常に機能しません。ブース用の仮想IPアドレスを設定する方法の詳細については、[手順7「boothdのリソースグループの設定」](#)を参照してください。ブースは、IPv4アドレスとIPv6アドレスの両方に対応しています。
- ⑤ このパラメータはオプションです。共有鍵に基づいてクライアントとサーバのブース認証を可能にします。このパラメータは、鍵ファイルのパスを指定します。

#### 鍵の要件

- 鍵はバイナリでもテキストでもかまいません。  
鍵がテキストの場合は、先頭と末尾の空白、および改行文字は無視されます。
- 鍵の長さは8文字から64文字の間である必要があります。

- 鍵はユーザ hacluster とグループ haclient に属している必要があります。
  - 鍵は、その鍵ファイルの所有者のみが読み取り可能である必要があります。
- ⑥ ブースによって管理されるチケット。各チケットで、ticket エントリを追加します。この例では、チケット ticket-nfs は、後でNFSとDRBDのフェールオーバー用に使用されます。詳細については、7項「DRBDの設定」を参照してください。
  - ⑦ このパラメータはオプションです。チケットの有効期限を秒単位で定義します。チケットを付与されたサイトは定期的にチケットを更新します。ブースが、指定された有効期限内にチケット更新に関する情報を受け取らない場合、チケットは取り消され、他のサイトに付与されます。有効期限が指定されていない場合、デフォルトでチケットは 600 秒後に期限切れになります。このパラメータを120秒未満の値に設定しないでください。
  - ⑧ このパラメータはオプションです。タイムアウト期間を秒単位で指定します。この期間内に応答を受信できなかった場合、ブースはパケットを再送信します。他のブースメンバー(すべてのアービトラータとサイト)にパケットが届くのに十分な長さのタイムアウト期間を指定する必要があります。
  - ⑨ このパラメータはオプションです。ブースがパケットの送信を再試行する回数を指定します。この回数に達すると、ブースは他のサイトによる確認を待つことを放棄します。2 以下の値は無効であり、その場合はブースは起動されなくなります。
  - ⑩ このパラメータはオプションです。チケットの更新頻度期間を設定します。チケットの更新は、デフォルトでは有効期限の半分の時間ごとに行われます。ネットワークの信頼性が長期間の間に低下することが多い場合は、更新頻度を高くすることをお勧めします。毎回の更新前に、before-acquire-handler が実行されます。
  - ⑪ このパラメータはオプションです。このパラメータは1つ以上のスクリプトをサポートしています。複数のスクリプトを使用するためには、各スクリプトは異なるチェックを担当できます(たとえば、クラスタ状態、データセンターの接続性、環境正常性センサーなど)。すべてのスクリプトを /etc/booth.d/TICKET\_NAME ディレクトリに保管して、これらのスクリプトに正しい所有権(ユーザ hacluster およびグループ haclient)が設定されていることを確認します。このディレクトリ名をパラメータ before-acquire-handler の値として割り当てます。  
このディレクトリ内のスクリプトはアルファベット順に実行されます。すべてのスクリプトは、boothd によってチケットの取得または更新が試行される前に呼び出されます。そのチケットが付与または更新されるためには、「すべての」スクリプトが正常に実行される必要があります。セマンティクスは単一スクリプトの場合と同じです。すなわち、0 以外の終了コードでは、boothd はそのチケットを放棄します。
  - ⑫ /usr/share/booth/service-runnable スクリプトは、サンプルとして製品に含まれています。このスクリプトを使用するには、このスクリプトをそれぞれの「チケット」ディレクトリにリンクします。

```
root # ln -s /usr/share/booth/service-runnable /etc/booth.d/TICKET_NAME
```

`/etc/booth.d/TICKET_NAME` ディレクトリに `service-runnable` スクリプトが含まれているとします。このシンプルなスクリプトは、`crm_simulate` をベースにしています。このスクリプトを使用して、特定のクラスタリソースを現在のクラスタサイト上で実行「できる」かどうかを判定できます。すなわちこのスクリプトは、クラスタがそのリソースを実行できるだけの十分に正常な状態であるのかどうかを確認します(すべてのリソース依存関係が満たされているのか、クラスタパーティションにクォラムが設定されているのか、ダーティノードがないのかなど)。たとえば、依存性チェーン内のサービスがすべての使用可能なノード上で `INFINITY` という failcount 値を保有している場合、そのサービスをそのサイトで実行することはできません。その場合は、そのチケットを要求することは無駄です。

- 13 `before-acquire-handler` (ここでは `service-runnable` スクリプト)によって判定されるリソース。それぞれのチケットによって保護されるリソースを参照する必要があります。この例では、リソース `db-1` は `ticketA` によって保護され、`db-8` は `ticketB` によって保護されます。DRBD(`ms_drbd_nfs`)のリソースはチケット `ticket-nfs` によって保護されます。
- 14 このパラメータはオプションです。チケットが失われた後、ブースはここで指定された時間だけさらに待ってから、そのチケットを取得します。この目的は、そのチケットを失ったサイトが、リソースを停止するかノードをフェンスすることで、リソースを放棄可能にすることです。一般的な遅延時間としては `60` 秒が考えられますが、最終的には保護されるリソースとフェンシング設定に依存します。デフォルトの設定は `0` です。  
リソースの停止や降格またはノードのフェンシングにかかる時間が不明な場合は(`loss-policy` に依存)、このパラメータを使用して、リソースが2つのサイトで同時に実行されることを防止してください。

#### 手順 1 ブース設定ファイルの手動編集

1. `root` または同等の権限でクラスタノードにログインします。
2. サンプルのブース設定ファイル `/etc/booth/booth.conf.example` を `/etc/booth/booth.conf` にコピーします。

```
root # cp -p /etc/booth/booth.conf.example /etc/booth/booth.conf
```
3. 例2「ブース設定ファイル」に従って `/etc/booth/booth.conf` を編集します。
4. 変更を検証して、ファイルを保存します。
5. すべてのクラスタノードとアービトラータ上で、ブース用に設定したファイアウォール内のポートを開きます。例2「ブース設定ファイル」の ② の箇所を参照してください。

#### 手順 2 YASTを使用したブースのセットアップ



1. root または同等の権限でクラスターノードにログインします。
2. YaSTの [Geo Cluster (Geoクラスター)] モジュールを開始します。
3. [編集] をクリックして既存のブース設定ファイルを編集するか、[追加] をクリックして新しいブース設定を作成します。
  - a. 表示される画面で、以下のパラメータを設定します。
    - 設定ファイル: ブース設定ファイルの名前。YaSTによって提示されるデフォルトの名前は booth です。この結果として、ブース設定が /etc/booth/booth.conf に書き込まれます。6.2項「複数のテナント用のブースセットアップ」の説明に従って別々のGeoクラスター用に複数のブースインスタンスをセットアップする必要がある場合にのみ、この値を変更してください。
    - トランスポート: サイト間の通信に使用するトランスポートプロトコル。UDPのみがサポートされていますが、将来的には他のトランスポート層も追従する予定です。例2「ブース設定ファイル」の ❶ の箇所も参照してください。
    - ポート: 各サイトのブースインスタンス間の通信用に使用されるポート。例2「ブース設定ファイル」の ❷ の箇所も参照してください。
    - アービトラータ: アービトラータとして使用するマシンのIPアドレス。例2「ブース設定ファイル」の ❸ の箇所も参照してください。  
[アービトラータ] を指定するには、[追加] をクリックします。開かれるダイアログで、アービトラータのIPアドレスを入力して、[OK] をクリックします。
    - サイト: サイト上の boothd 用に使用されるIPアドレス。例2「ブース設定ファイル」の ❹ の箇所も参照してください。  
お使いのGeoクラスターの [サイト] を指定するには、[追加] をクリックします。開かれるダイアログで、1つのサイトのIPアドレスを入力して、[OK] をクリックします。
    - チケット: ブースによって管理されるチケット。例2「ブース設定ファイル」の ❺ の箇所も参照してください。

[チケット]を指定するには、[追加]をクリックします。開かれるダイアログで、固有の[チケット]名を入力します。同じパラメータと値を使用した複数のチケットを設定する必要がある場合は、すべてのチケット用のデフォルトのパラメータと値を指定した「チケットテンプレート」を作成して、設定内容を保存します。そのためには、\_\_default\_\_を[チケット]名として使用します。

- 認証: ブースの認証を有効にするには、[認証]をクリックして、開かれるダイアログで、[Enable Security Auth (セキュリティ認証を有効にする)]をアクティブにします。既存の鍵がすでにある場合は、その鍵のパスとファイル名を[Authentication file (認証鍵)]に指定します。新しいGeoクラスタ用の鍵ファイルを生成するには、[Generate Authentication Key File (認証鍵ファイルの生成)]をクリックします。鍵が作成されて、[Authentication file (認証ファイル)]で指定された場所に書き込まれます。  
さらに、チケットに対してオプションのパラメータを指定できます。概要については、[例 2「ブース設定ファイル」](#)の 7 から 14 の箇所を参照してください。  
[OK]をクリックして、変更を確認します。]

The screenshot shows the 'Geo Clustering Settings' window. It has a sidebar on the left with 'Geo Clustering Settings' and 'Firewall Settings'. The main area is titled 'Geo Clustering Settings'. It contains fields for 'Environment Setting File (G)' (booth), 'Transfer' (UDP), and 'Port (P)' (9929). Below these are two lists: 'Arbitrator (A)' with IP '147.2.207.14' and 'Site (S)' with IP '192.168.201.151'. There are buttons for 'Add (A)', 'Edit (E)', and 'Delete (D)' for both lists. Below these is a table of tickets:

チケット	タイムアウト	再試行	重み	有効期限	acquire-after	before-acquire-handler
ticketA	10	5		600	60	/usr/share/booth/service-runnable
ticketB	10	5		600		/usr/share/booth/service-runnable

Buttons for 'Add (A)', 'Edit (E)', and 'Delete (D)' are below the table. At the bottom right are buttons for '認証 (U)', 'キャンセル (N)', 'OK (O)', 'キャンセル (Q)', and '完了 (F)'.

図 2 サンプルのチケットの依存関係

- [OK]をクリックして現在のブース設定画面を閉じます。YaSTでは、作成されたブース設定ファイルの名前が表示されます。
- YaSTモジュールを閉じる前に、[Firewall Configuration (ファイアウォール設定)]カテゴリに切り替えます。
  - ブース用に設定したポートを開くには、[Open Port in Firewall (ファイアウォール内のポートを開く)]を有効にします。



## ❗ 重要: ローカルマシン専用のファイアウォール設定

ファイアウォール設定は現在のマシンのみに適用されます。ファイアウォール設定によって、/etc/booth/booth.conf または他のいずれかのブース設定ファイルで指定されたすべてのポートについてUDP/TCPポートが開かれます(6.2項「複数のテナント用のブースセットアップ」を参照してください)。

Geoクラスタセットアップの他のすべてのクラスタノード上およびアービトラータ上でそれぞれのポートを必ず開いてください。この操作は、手動で行うか、Csync2を使用して以下のファイルを同期させることで行います。

- /etc/sysconfig/SuSEfirewall2
- /etc/sysconfig/SuSEfirewall2.d/services/booth

6. [Finish (完了)]をクリックして、すべての設定を確定してYaSTモジュールを閉じます。ステップ 3.aで指定した[Configuration File (設定ファイル)]の NAME に応じて、設定は /etc/booth/NAME.conf に書き込まれます。

## 6.2 複数のテナント用のブースセットアップ

複数のGeoクラスタを含むセットアップの場合は、同じアービトラータを「共有」できます(SUSE Linux Enterprise High Availability Extension 12以降)。複数のブース設定ファイルを提供することで、複数のブースインスタンスを同じアービトラータ上で開始することができ、それぞれのブースインスタンスを異なるポート上で動作させることができます。このようにすることで、「1台の」マシンを「複数の」Geoクラスタ用のアービトラータとして機能させることができます。

2つのGeoクラスタがあるとします。1つはEMEA(ヨーロッパ/中東/アフリカ)にあり、もう1つはアジア太平洋地域(APAC)にあります。

両方のGeoクラスタ用に同じアービトラータを使用するには、/etc/booth ディレクトリ内に2つの設定ファイル /etc/booth/emea.conf および /etc/booth/apac.conf を作成します。これらの設定ファイル間の違いは、以下のパラメータのみにする必要があります。

- ブースインスタンスの通信用に使用されるポート。
- このアービトラータが使用される対象となる個々のGeoクラスタに属しているサイト。

例 3 /etc/booth/apac.conf

```
transport = UDP ❶
```

```

port = 9133 ②
arbitrator = 147.2.207.14 ③
site = 192.168.2.254 ④
site = 192.168.1.112 ④
authfile = /etc/booth/authkey-apac ⑤
ticket = "tkt-db-apac-intern" ⑥
    timeout = 10
    retries = 5
    renewal-freq = 60
    before-acquire-handler ⑪ = /usr/share/booth/service-runnable ⑫ db-apac-intern ⑬
ticket = "tkt-db-apac-cust" ⑥
    timeout = 10
    retries = 5
    renewal-freq = 60
    before-acquire-handler = /usr/share/booth/service-runnable db-apac-cust

```

例 4 /etc/booth/emea.conf

```

transport = UDP ①
port = 9150 ②
arbitrator = 147.2.207.14 ③
site = 192.168.201.151 ④
site = 192.168.202.151 ④
authfile = /etc/booth/authkey-emea ⑤
ticket = "tkt-sap-crm" ⑥
    expire = 900
    renewal-freq = 60
    before-acquire-handler ⑪ = /usr/share/booth/service-runnable ⑫ sap-crm ⑬
ticket = "tkt-sap-prod" ⑥
    expire = 600
    renewal-freq = 60
    before-acquire-handler = /usr/share/booth/service-runnable sap-prod

```

- ① サイト間の通信に使用するトランスポートプロトコル。UDPのみがサポートされていますが、将来的には他のトランスポート層も追従する予定です。したがって、現時点ではこのパラメータは省略可能です。
- ② 各サイトのブースインスタンス間の通信用に使用されるポート。これらの設定ファイルでは、同じアービトレータ上の複数のブースインスタンスを開始可能にするために異なるポートが使用されます。
- ③ アービトレータとして使用するマシンのIPアドレス。上記の例では、異なるGeoクラスタ用に同じアービトレータを使用しています。
- ④ サイト上の boothd 用に使用されるIPアドレス。両方のブース設定ファイルで指定されているサイトは異なります。これらのサイトは2つの異なるGeoクラスタに属しているからです。

- ⑤ このパラメータはオプションです。共有鍵に基づいてクライアントとサーバのブース認証を可能にします。このパラメータは、鍵ファイルのパスを指定します。テナントごとに異なる鍵ファイルを使用します。

#### 鍵の要件

- 鍵はバイナリでもテキストでもかまいません。  
鍵がテキストの場合は、先頭と末尾の空白、および改行文字は無視されます。
  - 鍵の長さは8文字から64文字の間である必要があります。
  - 鍵はユーザ `hacluster` とグループ `haclient` に属している必要があります。
  - 鍵は、その鍵ファイルの所有者のみが読み取り可能である必要があります。
- ⑥ ブースによって管理されるチケット。理論的には、異なるブース設定ファイルで同じチケット名を指定できます。これらのチケットは、異なるブースインスタンスによって管理されている異なるGeoクラスタの一部であるため、干渉することはありません。ただし、上記の例で示すように、Geoクラスタごとに異なるチケット名を使用することをお勧めします(より適切な概要のために)。
- ⑪ このパラメータはオプションです。このパラメータが設定されている場合は、指定されたコマンドが、`boothd` によってチケットの取得または更新が試行される前に呼び出されます。`0` 以外の終了コードでは、`boothd` はそのチケットを放棄します。
- ⑫ ここで指定されている `service-runnable` スクリプトは、サンプルとして製品に含まれています。これは `crm_simulate` をベースにしたシンプルなスクリプトです。このスクリプトを使用して、特定のクラスタリソースを現在のクラスタサイト上で実行「できる」かどうかを判定できます。すなわちこのスクリプトは、クラスタがそのリソースを実行できるだけの十分に正常な状態であるのかどうかを確認します(すべてのリソース依存関係が満たされているのか、クラスタパーティションにクォーラムが設定されているのか、ダーティノードがないのかなど)。たとえば、依存性チェーン内のサービスがすべての使用可能なノード上で `INFINITY` というfailcount値を保有している場合、そのサービスをそのサイトで実行することはできません。その場合は、そのチケットを要求することは無駄です。
- ⑬ `before-acquire-handler` (ここでは `service-runnable` スクリプト)によって判定されるリソース。それぞれのチケットによって保護されるリソースを参照する必要があります。

#### 手順 3 異なるGEOクラスタに対する同じアービトラータの使用

1. `/etc/booth` 内に複数の異なるブース設定ファイルを作成します(例3「`/etc/booth/apac.conf`」と例4「`/etc/booth/emea.conf`」を参照)。これらのファイルは、手動で作成するか、YaSTを使用して作成します(手順2「YaSTを使用したブースのセットアップ」を参照)。
2. アービトラータ上で、`/etc/booth` 内のいずれかのブース設定ファイルで指定されたポートを開きます。

3. アービトレータが使用される対象である個別のGeoクラスタに属しているノード上で、それぞれのブースインスタンス用に使用されるポートを開きます。
4. 同じブース設定を使用するすべてのクラスタノードとアービトレータの間で、それぞれのブース設定ファイルを同期させます。詳細については、6.3項「すべてのサイトとアービトレータに対するブース設定の同期」を参照してください。
5. アービトレータ上で、マルチテナンシーセットアップ向けに個別のブースインスタンスを開始します(アービトレータ上のブースサービスの開始を参照)。
6. 個別のGeoクラスタ上で、ブースサービスを開始します(クラスタサイト上のブースサービスの開始を参照)。

## 6.3 すべてのサイトとアービトレータに対するブース設定の同期



### 注記: すべてのサイトとアービトレータ上で同じブース設定の使用

ブースを正常に機能させるためには、同一Geoクラスタ内のすべてのクラスタノードとアービトレータが同じブース設定を使用する必要があります。

Csync2を使用して、ブース設定を同期させることができます。詳細については、8.1項「Geoクラスタ用のCsync2設定」および8.2項「Csync2を使用した変更内容の同期」を参照してください。

何らかのブース設定が変更された場合は、その変更内容に応じてすべての関連先で設定ファイルを更新して、ブースサービスを再起動してください(6.5項「実行中のブースの再設定」を参照)。

## 6.4 ブースサービスの有効化と開始

### クラスタサイト上のブースサービスの開始

各クラスタサイト用のブースサービスは、手順7「[boothdのリソースグループの設定](#)」で設定されたブースリソースグループによって管理されます。サイトあたり1つのブースサービスインスタンスを開始するには、各クラスタサイト上でそれぞれのブースリソースグループを開始します。

### アービトレータ上のブースサービスの開始

SUSE Linux Enterprise 12以降では、ブースアービトレータはsystemdによって管理されます。ユニットファイルの名前は `booth@.service` です。`@` は、パラメータを指定してこのサービスを実行できることを示しています。この場合は、パラメータは設定ファイル名です。アービトレータ上のブースサービスを「有効」にするには、次のコマンドを使用します。

```
root # systemctl enable booth@booth
```

コマンドラインからサービスを有効にした後に、YaST Services Managerを使用してそのサービスを管理できます(そのサービスが無効化されていない場合)。サービスを無効化した場合は、次回にsystemdを再起動したときにYaSTのサービスリストからそのサービスが削除されます。ただし、ブースサービスを「開始」するためのコマンドは、ブースセットアップに応じて異なります。

- デフォルトのセットアップ(6.1項「デフォルトのブースセットアップ」を参照)を使用している場合は、/etc/booth/booth.conf のみが設定されます。その場合は、各アービトラータにログインして、次のコマンドを使用してください。

```
root # systemctl start booth@booth
```

- ブースをマルチテナンシーモードで実行している場合は(6.2項「複数のテナント用のブースセットアップ」を参照)、複数のブース設定ファイルを /etc/booth 内で設定済みです。個別のブースインスタンス向けのサービスを開始するには systemctl start booth@ NAME を使用します。ここで、NAME はそれぞれの設定ファイル /etc/booth/NAME.conf の名前です。  
たとえば、/etc/booth/emea.conf および /etc/booth/apac.conf というブース設定ファイルがある場合は、アービトラータにログインして、次のコマンドを実行します。

```
root # systemctl start booth@emea  
root # systemctl start booth@apac
```

これにより、ブースサービスがアービトラータモードで起動します。その他すべてのブースデーモンと通信できますが、クラスタサイトで実行するブースデーモンとは異なり、チケットは付与されません。ブースアービトラータは選択のみに参加します。それ以外については、ブースアービトラータは休止状態になります。

## 6.5 実行中のブースの再設定

ブースサービスがすでに実行されている最中にブース設定を変更する必要がある場合は、以下の手順を実行します。

1. 希望に応じてブース設定ファイルを編集します。
2. Geoクラスタの一部であるすべてのクラスタノードとアービトラータに対して、更新後のブース設定ファイルを同期させます。詳細については、8項「すべてのサイトとアービトラータ間の設定ファイルの同期」を参照してください。

3. アービトレータ上とクラスタサイト上のブースサービスを再起動します(6.4項「ブースサービスの有効化と開始」を参照)。この操作によって、サイトにすでに付与されているチケットが影響を受けることはありません。

## 7 DRBDの設定

全体的なシナリオの詳細については、5項「サンプルシナリオと基本ステップ – 概要」を参照してください。ルーティングされたIPv4接続またはIPv6接続で接続された2つのクラスタサイトがあるとしします。この接続の伝送速度は、数Mbit/秒から最大10Gbit/秒の範囲です。この場合は、待ち時間が長い場合、これらのサイトにまがたってクラスタファイルシステムを使用することはできません。しかし、一方のサイトがダウンした場合に素早いフェールオーバーを可能にするために、DRBDを使用してデータを複製できます(アクティブ/パッシブセットアップ)。DRBDは、別々のサイトに配置されたホスト間でブロックデバイス(ハードディスク、パーティション、論理ボリュームなど)のコンテンツをミラーリングすることで、ストレージデータを複製するためのソフトウェアです。フェールオーバーはブースサービスを介して管理されます(ブースクラスタチケットマネージャを参照)。

### 7.1 DRBDのシナリオと基本ステップ

図3「DRBDのセットアップとリソースのスタッキング」では、以下で設定するセットアップとリソースをグラフィカルに表示しています。

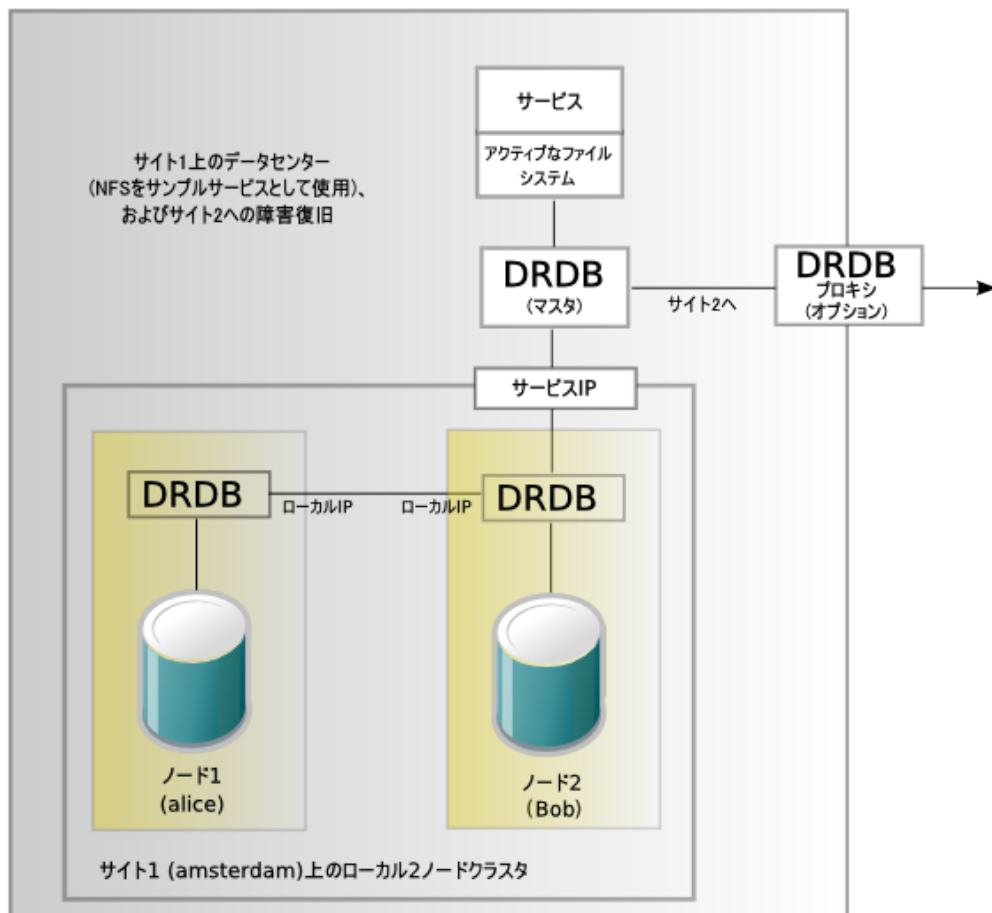


図 3 DRBDのセットアップとリソースのスタッキング

#### シナリオ - 詳細

- ファイルシステムはNFSを介してGeoクラスタ全体にわたって提供されます。
- DRBDの下位のストレージ層としてLVMが使用されます。
- 「スタックされた」DRBDリソース:
  - 上位層のDRBDは、サイトあたり1つのノード上で実行され、Geoクラスタのもう一方のサイトにデータを複製します。
  - 下位層のDRBDは、データのローカルレプリケーションを実行します(1つのクラスタサイトのノード間で)。サイトあたり1つのノード上でいずれかの下位DRBDデバイスをアクティブ化した後に、サービスのIPアドレス(クラスタリソースとして設定されるもの)が開始されます。



- サイトamsterdam上で、DRBDは、同期レプリケーションプロトコルであるプロトコル C で実行されています。DRBDではLAN内のローカルIPアドレスが使用されます。
- サービスのIPアドレスは、サービス自体用だけに使用されるのではなく、レプリケーションのために上位DRBDデバイス(2次状態で実行されているもの)がアクセスできる固定ポイントとしても使用されます。
- ファイルシステムサービスを実行する必要があるサイト上で、上位層DRBDは `ocf:linbit:drbd` リソースエージェントによって `primary` モードに設定されます。この結果として、各種のアプリケーションはこの場所のファイルシステムをマウントして使用できます。
- 希望に応じて、Geoクラスタのもう一方のサイトへのDRBD接続では、その間でDRBDプロキシを使用できます。

このセットアップシナリオの場合は、以下の基本ステップを実行する必要があります。

1. DRBD設定ファイルを編集して、各Geoクラスタサイトの設定と、サイト間のDRBD接続の設定を追加します。詳細については、7.2項「DRBDの設定」にある例を参照してください。
2. クラスタリソースを設定します(9.1項「DRBDのリソースと制約」を参照)。
3. ブースを設定します(6項「ブースサービスの設定」を参照)。
4. 各ローカルクラスタ内とGeoクラスタサイト間のDRBDとブース設定ファイルの同期を設定します。詳細については、8項「すべてのサイトとアービトラータ間の設定ファイルの同期」を参照してください。

## 7.2 DRBDの設定

DRBD 8.3以降では、DRBD設定ファイルは複数の個別ファイルに分割されています。これらのファイルは `/etc/drbd.d/` ディレクトリに配置する必要があります。以下のいくつかのDRBD設定断片では、シナリオ ー 詳細で説明したシナリオに対応する基本的なDRBD設定を示しています。すべての断片は、単一のDRBDリソース設定ファイルに追加できます( `/etc/drbd.d/nfs.res` など)。その後、このファイルをCsync2を使用して同期できます(8.1項「Geoクラスタ用のCsync2設定」を参照)。以下のDRBD設定の断片は必要最小限のものです。すなわち、これらの設定断片にはパフォーマンスチューニングオプションなどは一切含まれていません。DRBDのチューニング方法の詳細については、SUSE Linux Enterprise High Availability Extensionの『Administration Guide』(<http://www.suse.com/documentation/> から入手可能)で「DRBD」の章を参照してください。

例 5 サイト1(AMSTERDAM)用のDRBD設定断片

```
resource nfs-lower-amsterdam ① {
    disk          /dev/volgroup/lv-nfs; ②
```



```

meta-disk    internal; ③
device       /dev/drbd0; ④
protocol     C; ⑤
net {
    shared-secret "2a9702a6-8747-11e3-9ebb-782bcbd0c11c"; ⑥
}

on alice { ⑦
    address      192.168.201.111:7900; ⑧
    node-id 0; ⑨
}
on bob { ⑦
    address      192.168.201.112:7900; ⑧
    node-id 1; ⑨
}
connection-mesh { ⑩
    hosts alice bob;
}
}

```

- ① それぞれのサービス(ここではNFS)への関連付けを可能にするリソース名。サイト名も含めることにより、名前の競合を生じさせることなく、全体的なDRBD設定をサイト間で同期させることができます。
- ② ノード間で複製されるデバイス。この例では、LVMはDRBDの下のストレージ層として使用されており、ボリュームグループ名は volgroup です。
- ③ meta-diskパラメータには、通常、値 internal が含まれますが、メタデータを保持する明示的なデバイスを指定することもできます。詳細については、<http://www.drbd.org/users-guide-emb/ch-internals.html#s-metadata> を参照してください。
- ④ DRBD用デバイス名とそのマイナー番号。ローカルレプリケーション用の下位層DRBDと、Geoクラスタサイト間のレプリケーション用の上位層DRBDを区別するために、デバイスマイナー番号として 0 と 10 が使用されています。
- ⑤ DRBDは、同期レプリケーションプロトコルであるプロトコル C で実行されています。プライマリノード上のローカル書き込み操作が完了したと見なされるタイミングは、ローカルとリモートの両方のディスク書き込みが確認された後です。このため、単一ノードが失われた場合でもデータは一切失われないことが保証されます。両方のノード(または両ノードのストレージサブシステム)が同時に破損して復旧不能になった場合は、当然ながら、このレプリケーションプロトコルを採用している場合でも、データの喪失を避けることはできません。
- ⑥ 共有秘密を使用して接続ペアが検証されます。接続ペアごとに異なる共有秘密が必要です。uuidgen プログラムを使用して、他と重複しない共有秘密値を得ることができます。
- ⑦ on セクションでは、この設定文が適用されるホストを記述します。

- ⑧ それぞれのノードのローカルIPアドレスとポート番号。各DRBDリソースは個別のポートを必要とします。
- ⑨ 複数のノードを設定する際は、ノードIDが必要です。ノードIDは、別々のノードを区別するための固有の負でない整数です。
- ⑩ 同一メッシュのすべてのノードを設定します。hosts パラメータには、同じDRBDセットアップを共有するすべてのホスト名が含まれます。

#### 例 6 サイト2(BERLIN)用のDRBD設定断片

サイト2(berlin)の設定は、サイト1の設定とほとんど同じです。したがって、ほとんどのパラメータの値をそのまま使用できます(ボリュームグループや論理ボリュームの名前を含む)。ただし、以下のパラメータの値は変更する必要があります。

- DRBDリソースの名前(①)
- ノードの名前とローカルIPアドレス(⑦と⑧)
- 共有秘密(⑥)

```
resource nfs-lower-berlin ① {  
    disk          /dev/volgroup/lv-nfs; ②  
    meta-disk     internal; ③  
    device        /dev/drbd0; ④  
    protocol      C; ⑤  
    net {  
        shared-secret "2e9290a0-8747-11e3-a28c-782bcbd0c11c"; ⑥  
    }  
  
    on charly { ⑦  
        address          192.168.202.111:7900; ⑧  
        node-id 0;  
    }  
    on doro { ⑦  
        address          192.168.202.112:7900; ⑧  
        node-id 1;  
    }  
    connection-mesh {  
        hosts charly doro;  
    }  
}
```

- ① それぞれのサービス(ここではNFS)への関連付けを可能にするリソース名。サイト名も含めることにより、名前の競合を生じさせることなく、全体的なDRBD設定をサイト間で同期させることができます。

- ② ノード間で複製されるデバイス。この例では、LVMはDRBDの下ストレージ層として使用されており、ボリュームグループ名は `volgroup` です。
- ③ meta-diskパラメータには、通常、値 `internal` が含まれますが、メタデータを保持する明示的なデバイスを指定することもできます。詳細については、<http://www.drbd.org/users-guide-emb/ch-internals.html#s-metadata> を参照してください。
- ④ DRBD用デバイス名とそのマイナー番号。ローカルレプリケーション用の下位層DRBDと、Geoクラスタサイト間のレプリケーション用の上位層DRBDを区別するために、デバイスマイナー番号として `0` と `10` が使用されています。
- ⑤ DRBDは、同期レプリケーションプロトコルであるプロトコル `C` で実行されています。プライマリノード上のローカル書き込み操作が完了したと見なされるタイミングは、ローカルとリモートの両方のディスク書き込みが確認された後です。このため、単一ノードが失われた場合でもデータは一切失われないことが保証されます。両方のノード(または両ノードのストレージサブシステム)が同時に破損して復旧不能になった場合は、当然ながら、このレプリケーションプロトコルを採用している場合でも、データの喪失を避けることはできません。
- ⑥ 共有秘密を使用して接続ペアが検証されます。接続ペアごとに異なる共有秘密が必要です。`uuidgen` プログラムを使用して、他と重複しない共有秘密値を得ることができます。
- ⑦ `on` セクションでは、この設定文が適用されるホストを記述します。
- ⑧ それぞれのノードのローカルIPアドレスとポート番号。各DRBDリソースは個別のポートを必要とします。

例 7 サイト間接続用のDRBD設定断片

```
resource nfs-upper ① {
    disk            /dev/drbd0; ②
    meta-disk       internal;
    device          /dev/drbd10; ③
    protocol        A; ④
    net {
        shared-secret "3105dd88-8747-11e3-a7fd-782bcbd0c11c"; ⑤
        ping-timeout 20; ⑥
    }


    stacked-on-top-of      nfs-lower-amsterdam { ⑦
        address            192.168.201.151:7910; ⑧
    }
    stacked-on-top-of      nfs-lower-berlin { ⑦
        address            192.168.202.151:7910; ⑧
    }
}
```

- ① それぞれのサービス(ここではNFS)への関連付けを可能にするリソース名。これは上位層DRBDの設定です。上位層DRBDは、Geoクラスタのもう一方のサイトにデータを複製します。
- ② 複製するストレージディスクは、DRBDデバイス `/dev/drbd0` です。`/dev/drbd/by-res/nfs-lower-site-N/0` を代わりに使用できますが、これはサイト固有となるため、サイトごとの設定内に移動する必要があります。すなわち、それぞれの `stacked-on-top-of` キーワードの下に移動する必要があります。
- ③ DRBD用デバイス名とそのマイナー番号。ローカルレプリケーション用の下位層DRBDと、Geoクラスタサイト間のレプリケーション用の上位層DRBDを区別するために、デバイスマイナー番号として `0` と `10` が使用されています。
- ④ DRBDは、長距離レプリケーション用に使用される非同期レプリケーションプロトコルであるプロトコル `A` で実行されています。プライマリノード上のローカル書き込み操作が完了したと見なされるタイミングは、ローカルディスク書き込みが完了して、レプリケーションパケットがローカルTCP送信バッファに配置されたときです。強制フェールオーバーが実行された場合は、データが失われる可能性があります。スタンバイノード上のデータはフェールオーバー後も同じままです。ただし、クラッシュ前に実行された最新の更新の内容は失われる可能性があります。
- ⑥ 待ち時間が長くなるため、pingタイムアウトを `20` に設定してください。
- ⑥ 共有秘密を使用して接続ペアが検証されます。接続ペアごとに異なる共有秘密が必要で、`uuidgen` プログラムを使用して、他と重複しない共有秘密値を得ることができます。
- ⑦ ホスト名を渡す代わりに、ここでは下位デバイスの上にスタックするようにDRBDに指示しています。これは、下位デバイスが `Primary` である必要があることを示しています。
- ⑧ どのクラスタノードに下位DRBDデバイス `Primary` があるのか不明な状態で、Geoクラスタのもう一方のサイトへのTCP/IP接続を可能にするために、NFS向けに設定されたサービスIPアドレスを使用します。サービスIPアドレスの設定方法については、[手順5「DRBDセットアップ向けのリソースの設定」](#)を参照してください。


## 8 すべてのサイトとアービトラータ間の設定ファイルの同期

クラスタ内のすべてのノードとGeoクラスタ全体にわたって重要な設定ファイルを複製するには、`Csync2`を使用します。`Csync2`は、同期グループ別にソートされた任意の数のホストを操作できます。各同期グループは、メンバーホストの独自のリストとその包含/除外パターン(同期グループ内でどのファイルを同期するか定義するパターン)を持っています。グループ、各グループに属するホスト名、および各グループの包含/除外ルールは、`Csync2`設定ファイル `/etc/csync2/csync2.cfg` で指定されます。

Csync2は、認証には、同期グループ内でIPアドレスと事前共有キーを使用します。管理者は、同期グループごとに1つのキーファイルを生成し、そのファイルをすべてのグループメンバにコピーする必要があります。


Csync2は、TCPポート(デフォルトでは 6556)を介して他のサーバにアクセスして、リモートCsync2インスタンスを開始します。Csync2の詳細については、<http://oss.linbit.com/csync2/paper.pdf>  を参照してください。

## 8.1 Geoクラスタ用のCsync2設定

YaSTを使用して個別のクラスタ向けにCsync2を設定する方法については、SUSE Linux Enterprise High Availability Extensionの『Administration Guide』で、「Installation and Basic Setup」の章の「Transferring the Configuration to All Nodes」という節を参照してください。ただしYaSTは、Geoクラスタに必要な設定のような複雑なCsync2設定を処理することはできません。 4「Geoクラスタ用のCsync2設定の例」に示すような以下の設定については、設定ファイルを編集することでCsync2を手動で設定してください。

ローカルクラスタ内だけでなく、さまざまな地域に分散されたサイト間でもファイルを同期させるようにCsync2を調整するには、Csync2の設定で以下の2つの同期グループを定義する必要があります。

- グローバルグループ ha\_global (Geoクラスタに属しているすべてのサイトとアービトレータの間で、グローバルに同期される必要のあるファイル用)
- ローカルクラスタサイトのグループ ha\_local (ローカルクラスタ内で同期される必要のあるファイル用)

これら2つの同期グループ用の複数のCsync2設定ファイルの概要については、 4「Geoクラスタ用のCsync2設定の例」を参照してください。

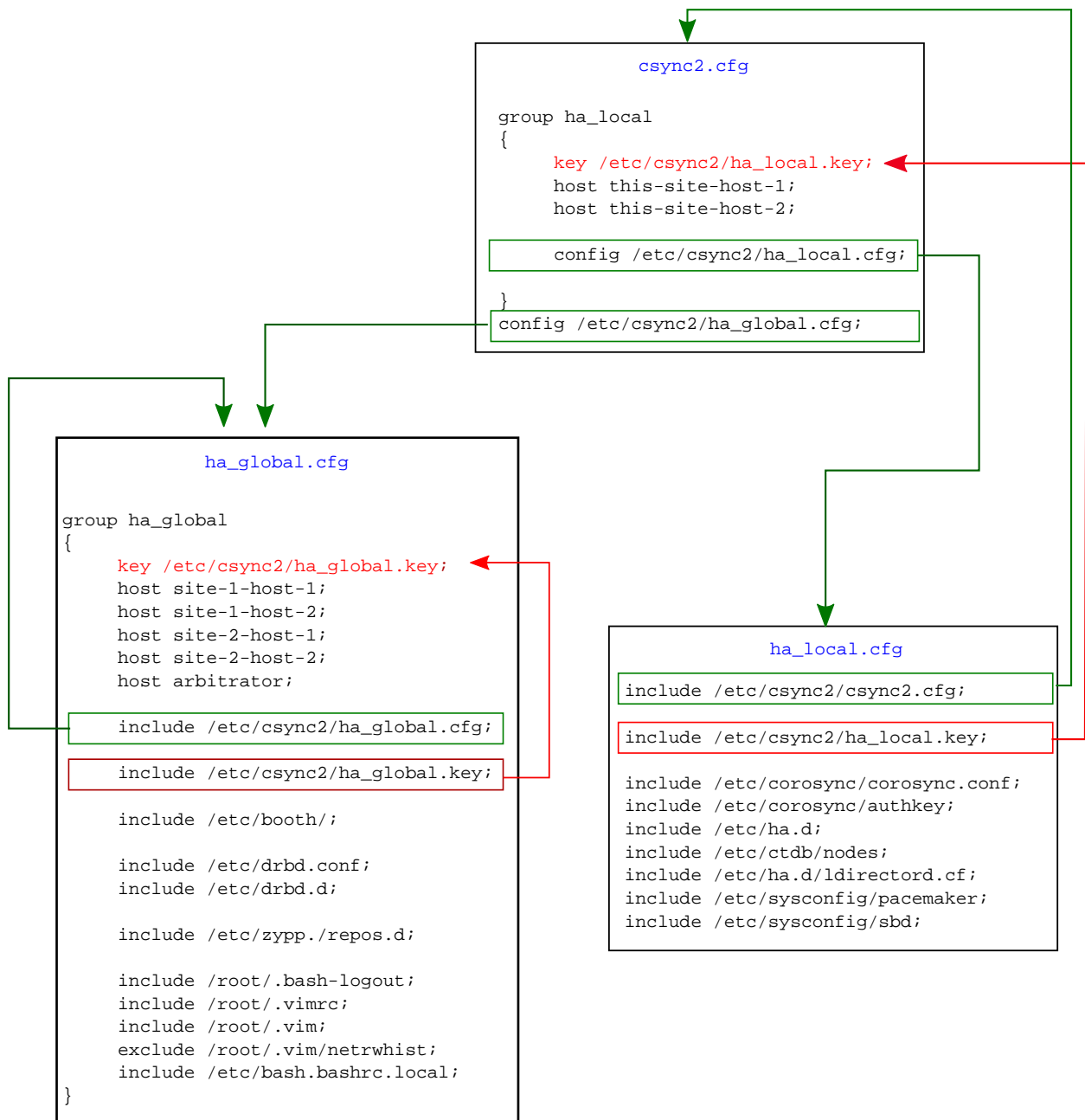


図 4 GEOクラスタ用のCSYNC2設定の例

認証鍵ファイルとそれらの参照は赤色で表示されています。Csync2設定ファイルの名前は青色で表示されており、それらの参照は緑色で表示されています。詳細については、[Csync2の設定例: 設定ファイル](#)を参照してください。

#### CSYNC2の設定例: 設定ファイル

##### /etc/csync2/csync2.cfg

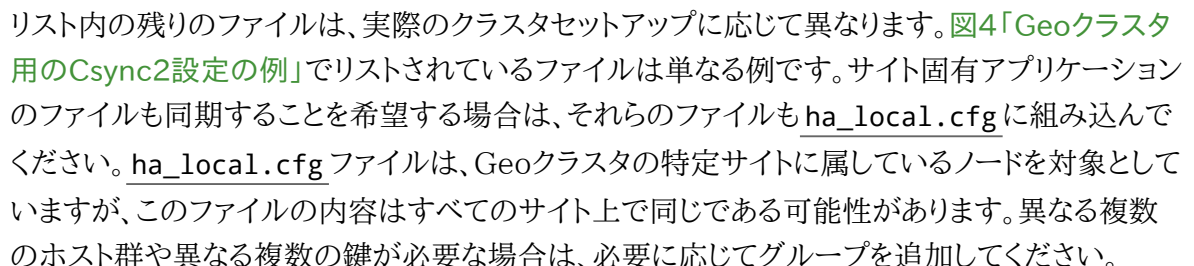
メインのCsync2設定ファイル。このファイルは意図的に簡潔になっており、以下のみで構成されています。

- 同期グループ ha\_local の定義。このグループは2つのノード( this-site-host-1 と this-site-host-2 )で構成されており、/etc/csync2/ha\_local.key を使用して認証を行います。このグループのみについて同期されるファイルのリストは、別の Csync2 設定ファイルである /etc/csync2/ha\_local.cfg に記載されています。このファイルは、config ステートメントを使用して組み込まれています。
- 別の Csync2 設定ファイルである /etc/csync2.cfg/ha\_global.cfg の参照。このファイルも config ステートメントを使用して組み込まれています。

#### /etc/csync2/ha\_local.cfg

このファイルはローカルクラスタのみに関するものです。このファイルでは、ha\_local 同期グループ内でのみ同期されるファイルのリストが指定されます。これらのファイルはクラスタごとに固有であるからです。特に重要なものは以下のとおりです。

- /etc/csync2/csync2.cfg: このファイルには、ローカルクラスタノードのリストが含まれています。
- /etc/csync2/ha\_local.key: ローカルクラスタ内の Csync2 同期用に使用される認証鍵です。
- /etc/corosync/corosync.conf: このファイルでは、ローカルクラスタノード間の通信チャンネルが定義されます。
- /etc/corosync/authkey: Corosync 認証鍵です。

リスト内の残りのファイルは、実際のクラスタセットアップに応じて異なります。 図4「Geoクラスタ用のCsync2設定の例」でリストされているファイルは単なる例です。サイト固有アプリケーションのファイルも同期することを希望する場合は、それらのファイルも ha\_local.cfg に組み込んでください。ha\_local.cfg ファイルは、Geoクラスタの特定サイトに属しているノードを対象としていますが、このファイルの内容はすべてのサイト上で同じである可能性があります。異なる複数のホスト群や異なる複数の鍵が必要な場合は、必要に応じてグループを追加してください。

#### /etc/csync2.cfg/ha\_global.cfg

このファイルでは、Csync2 同期グループ ha\_global が定義されています。このグループは、複数のサイトにわたる「すべての」クラスタノード(アービトラータを含む)にまたがっています。Csync2 同期グループごとに異なる鍵を使用することが推奨されているため、このグループは /etc/csync2/ha\_global.key を使用して認証を行います。include ステートメントでは、ha\_global 同期グループ内で同期されるファイルのリストを指定します。特に重要なものは以下のとおりです。



- /etc/csync2/ha\_global.cfgと/etc/csync2/ha\_global.key: ha\_global 同期グループ用の設定ファイルと、このグループ内の同期用の認証鍵です。
- /etc/booth/: ブース設定が格納されているデフォルトディレクトリです。1つのブースセットアップを複数のテナント用に使用している場合は、このディレクトリには複数のブース設定ファイルが格納されています。ブース用に認証を使用している場合は、鍵ファイルもこのディレクトリに配置すると便利です。
- /etc/drbd.confと/etc/drbd.d (クラスタセットアップ内でDRBDを使用している場合): DRBDの設定は、リソース設定ファイルに含まれているホスト名から得られるため、グローバルに同期させることができます。
- /etc/zypp/repos.de: パッケージリポジトリは、すべてのクラスタノード上で同じである可能性が高いです。

リスト内の他のファイル(/etc/root/\*)は、利便性のため(クラスタ管理を簡易化するため)に組み込むことができるファイルの例です。



## 注記

csync2.cfg ファイルと ha\_local.key ファイルはサイト固有であるため、クラスタサイトごとに異なる内容でこれらのファイルを作成する必要があります。これらのファイルの内容は、同じクラスタに属しているノード上では同じですが、別のクラスタ上では異なります。各 csync2.cfg ファイルには、対象となるサイトに属しているホスト(クラスタノード)のリストに加えて、サイト固有の認証鍵が含まれている必要があります。

アービトラータは csync2.cfg ファイルも必要とします。ただし、アービトラータは ha\_global.cfg を参照するだけでかまいません。

## 8.2 Csync2を使用した変更内容の同期

Csync2を使用してファイルを正常に同期するには、以下の前提条件が満たされている必要があります。

- 同じ同期グループに属しているすべてのマシンで同じCsync2設定を使用できる。
- 各同期グループのCsync2認証鍵をそのグループのすべてのメンバーで使用できる。
- 「すべての」ノードとアービトラータ上でCsync2が実行されている。

したがって、Csync2を初めて実行する前に、以下の準備を行う必要があります。



1. 同期グループごとに1台のマシンにログインして、各グループ用の認証鍵を生成します。

```
root # csync2 -k NAME_OF_KEYFILE
```

ただし、同じグループの他のメンバー上で鍵ファイルを再生成「しない」でください。

図4「Geoクラスタ用のCsync2設定の例」を基準にした場合、これにより生成される鍵ファイルは、/etc/csync2/ha\_global.keyと、サイトあたり1つのローカル鍵(/etc/csync2/ha\_local.key)です。

2. 各鍵ファイルを各同期グループの「すべての」メンバーにコピーします。図4「Geoクラスタ用のCsync2設定の例」を基準にして、以下の手順を実行します。
  - a. /etc/csync2/ha\_global.keyを「すべての」関連先(Geoクラスタのすべてのサイト上のすべてのクラスタノードとアービトラータ)にコピーします。この鍵ファイルは、ha\_global.cfgで指定されているha\_globalグループに含まれるすべてのホスト上で使用できる必要があります。
  - b. 各サイトのローカル鍵ファイル(/etc/csync2/ha\_local.key)を、Geoクラスタの各サイトに属しているすべてのクラスタノードにコピーします。
3. サイト固有の /etc/csync2/csync2.cfg 設定ファイルを、Geoクラスタの各サイトに属しているすべてのクラスタノードとアービトラータにコピーします。
4. すべてのノードとアービトラータ上で次のコマンドを実行して、Csync2サービスがブート時に自動的に開始されるようにします。

```
root # systemctl enable csync2.socket
```

5. すべてのノードとアービトラータ上で次のコマンドを実行して、Csync2サービスを今すぐ開始します。

```
root # systemctl start csync2.socket
```

#### 手順 4 CSYNC2を使用したファイルの同期

1. 最初にすべてのファイルを一度同期させるには、設定の「コピー元」であるマシン上で次のコマンドを実行します。

```
root # csync2 -xv
```

これにより、すべてのファイルが同期グループの他のメンバーにプッシュされることで、すべてのファイルが一度同期されます。すべてのファイルが正常に同期されると、Csync2がエラーなしで終了します。

同期対象の1つ以上のファイルが(現在のマシン上だけでなく)他のマシン上で変更されている場合は、Csync2から衝突が報告されます。次の出力とよく似た出力が表示されます。

```
While syncing file /etc/corosync/corosync.conf:  
ERROR from peer site-2-host-1: File is also marked dirty here!  
Finished with 1 errors.
```

2. 現在のマシン上のファイルバージョンが「最適な」バージョンであることが分かっている場合は、そのファイルを強制的に使用して再同期を行うことで、衝突を解決できます。

```
root # csync2 -f /etc/corosync/corosync.conf  
root # csync2 -x
```

Csync2オプションの詳細については、`csync2 -help`を実行してください。



### 注記: 変更後の同期のプッシュ

Csync2は変更のみをプッシュします。Csync2はマシン間でファイルを絶えず同期しているわけではありません。

同期が必要なファイルを更新するたびに、変更を加えたマシン上で `csync2 -xv` を実行することで、変更を同じ同期グループの他のマシンにプッシュする必要があります。変更されていないファイルが配置された他のマシン上でこのコマンドを実行しても、何も起こりません。

## 9 クラスタリソースと制約の設定

特定のクラスタセットアップのために定義する必要があるリソースや制約の他に、Geoクラスタでは、次に説明するように、追加のリソースと制約が必要になります。以下の例で示しているcrmシェル (crmsh) またはHA Web Konsole (Hawk2) を使用して、これらのリソースや制約を設定できます。

この節では、Geoクラスタに固有のタスクを主に扱います。推奨されるクラスタ管理ツールの概要と、このツールを使用してリソースと制約を設定する方法に関する一般的な指示については、SUSE Linux Enterprise High Availability Extensionの『Administration Guide』(<http://www.suse.com/documentation/> から入手可能)で、以下のいずれかの章を参照してください。

- Hawk2: 「クラスタリソースの設定と管理(Webインタフェース)」の章
- crmsh: 「クラスタリソースの設定と管理(コマンドライン)」の章

## ！ 重要: サイト間でCIBが同期されない

CIBは、Geoクラスタのクラスタサイト間で自動的に同期「されません」。したがって、サイトごとにGeoクラスタ全体にわたって高可用性が求められるすべてのリソースを適宜設定する必要があります。

他のクラスタサイトへの設定の転送を簡易化するために、サイト固有のパラメータを持つリソースを、そのリソースが実行されているクラスタサイトの名前に応じてそれらのパラメータの値が決定されるような形で設定できます。

このことを実現するために、各サイトのクラスタ名をそれぞれの `/etc/corosync/corosync.conf` ファイルで指定する必要があります。たとえば、サイト1(`amsterdam`)の `/etc/corosync/corosync.conf` には、次のエントリが含まれている必要があります。

```
totem {
  [...]
  cluster_name: amsterdam
}
```

1つのサイト上でリソースを設定した後に、すべてのクラスタサイト上で必要なリソースにタグ付けて、それらのリソースを現在のCIBからエクスポートして、別のクラスタサイトのCIBにインポートできます。詳細については、9.3項「他のクラスタサイトへのリソース設定の転送」を参照してください。

## 9.1 DRBDのリソースと制約

DRBDのセットアップを完了するには、いくつかのリソースと制約を設定して(手順 5を参照)、それらのリソースと制約を他のクラスタサイトに転送する必要があります(9.3項「他のクラスタサイトへのリソース設定の転送」を参照)。

### 手順 5 DRBDセットアップ向けのリソースの設定

1. クラスタ `amsterdam` のいずれかのノード上で、シェルを起動して、`root` または同等の権限でログインします。
2. 「`crm configure`」と入力して、対話式 `crm` シェルに切り替えます。
3. NFS向けの(サイト固有の)サービスIPアドレスを基本プリミティブとして設定します。

```
crm(live)configure# primitive ip_nfs ocf:heartbeat:IPaddr2 \
  params iflabel="nfs" nic="eth1" cidr_netmask="24"
  params rule #cluster-name eq amsterdam ip="192.168.201.151" \
```

```
params rule #cluster-name eq berlin ip="192.168.202.151" \  
op monitor interval=10
```

4. ファイルシステムリソースとNFSサーバ用リソースを設定します。

```
crm(live)configure# primitive nfs_fs ocf:heartbeat:Filesystem \  
params device="/dev/drbd/by-res/nfs/0" directory="/mnt/nfs" \  
fstype="ext4" \  
crm(live)configure# primitive nfs_service systemd:nfs-server
```

5. DRBD用の以下のプリミティブとマルチステートリソースを設定します。

```
crm(live)configure# primitive drbd_nfs ocf:linbit:drbd \  
params drbd_resource="nfs-upper" \  
op monitor interval="31" role="Slave" \  
op monitor interval="30" role="Master" \  
crm(live)configure# primitive drbd_nfs_lower ocf:linbit:drbd \  
params rule #cluster-name eq amsterdam \  
drbd_resource="nfs-lower-amsterdam" \  
params rule #cluster-name eq berlin \  
drbd_resource="nfs-lower-berlin" \  
op monitor interval="31" role="Slave" \  
op monitor interval="30" role="Master" \  
crm(live)configure# ms ms_drbd_nfs drbd_nfs \  
meta master-max="1" master-node-max="1" \  
clone-max="1" clone-node-max="1" notify="true" \  
crm(live)configure# ms ms_drbd_nfs_lower drbd_nfs_lower \  
meta master-max="1" master-node-max="1" \  
clone-max="2" clone-node-max="1" notify="true"
```

6. 以下のコロケーション制約と順序付け制約が適用されるグループを追加します。

```
crm(live)configure# group g_nfs nfs_fs nfs_service \  
crm(live)configure# colocation col_nfs_ip_with_lower \  
inf: ip_nfs:Started ms_drbd_nfs_lower:Master \  
crm(live)configure# colocation col_nfs_g_with_upper \  
inf: g_nfs:Started ms_drbd_nfs:Master \  
crm(live)configure# colocation col_nfs_upper_with_ip \  
inf: ms_drbd_nfs:Master ip_nfs:Started \  
crm(live)configure# order o_lower_drbd_before_ip_nfs \  
inf: ms_drbd_nfs_lower:promote ip_nfs:start \  
crm(live)configure# order o_ip_nfs_before_drbd \  
inf: ip_nfs:start ms_drbd_nfs:promote \  
crm(live)configure# order o_drbd_nfs_before_svc \  
inf: ms_drbd_nfs:promote g_nfs:start
```

7. **show** で変更内容をレビューします。

8. すべて正しければ、`commit` で変更を送信し、`exit` でcrmライブ設定を終了します。  
設定がCIBに保存されます。

## 9.2 チケットの依存関係、ブースの制約とリソース

ブースのセットアップを完了するには、以下の手順を実行して、リソースのフェールオーバーとブースに必要なリソースと制約を設定する必要があります。

- リソースのチケット依存関係の設定
- boothdのリソースグループの設定
- 順序の制約の追加

リソースの設定は、それぞれのクラスタサイトで使用できる必要があります。9.3項「他のクラスタサイトへのリソース設定の転送」の説明に従って、リソースの設定を他のサイトに転送します。

### 手順 6 リソースのチケット依存関係の設定

Geoクラスタでは、特定のチケットに依存するリソースを指定できます。この特殊なタイプの制約と合わせて、チケットが取り消された場合のそれぞれのリソースの動作を定義する `loss-policy` を設定できます。属性 `loss-policy` は次の値を持つことができます。

- `fence`: 該当するリソースを実行するノードをフェンスします。
  - `stop`: 該当するリソースを停止します。
  - `freeze`: 該当するリソースに何も実行しません。
  - `demote`: `master` モードで実行する該当のリソースを `slave` モードに降格します。
1. クラスタamsterdamのいずれかのノード上で、シェルを起動して、`root` または同等の権限でログインします。
  2. 「`crm configure`」と入力して、対話式crmシェルに切り替えます。
  3. 特定のチケットにどのリソースが依存するかを指定する制約を設定します。たとえば、7.1項「DRBDのシナリオと基本ステップ」で概示したDRBDのシナリオ向けに以下の制約が必要です。

```
crm(live)configure# rsc_ticket nfs-req-ticket-nfs ticket-nfs: \
ms_drbd_nfs:Master loss-policy=demote
```

このコマンドは、nfs-req-ticket-nfs というIDを持つ制約を作成します。この制約により、マルチステートリソースの ms\_drbd\_nfs が ticket-nfs に依存することが指定されます。ただし、このリソースのマスタモードのみがこのチケットに依存します。ticket-nfs が取り消された場合は、ms\_drbd\_nfs は自動的に slave モードに降格されて、その結果としてDRBDが Secondary モードになります。このようにして、サイトがこのチケットを持たない場合でも、DRBDのレプリケーションが引き続き実行されることが保証されます。

4. さらに多くのチケットに依存するその他のリソースが必要な場合は、rsc\_ticket で必要なだけ制約を作成します。
5. show で変更内容をレビューします。
6. すべて正しければ、commit で変更を送信し、exit でcrmライブ設定を終了します。設定がCIBに保存されます。

#### 例 8 プリミティブのチケット依存関係

以下ではもう1つの例として、プリミティブリソース rsc1 を ticketA に依存させる制約を示します。

```
crm(live)configure# rsc_ticket rsc1-req-ticketA ticketA: \  
rsc1 loss-policy="fence"
```

ticketA が取り消された場合は、このリソースを実行しているノードはフェンスされます。

#### 手順 7 boothdのリソースグループの設定

各サイトは、その他のブースデーモンと通信する boothd のインスタンスを1つ実行する必要があります。デーモンはあらゆるノードで起動できるため、プリミティブリソースとして設定する必要があります。(可能な場合) boothd リソースが同じノード上にとどまるようにするには、リソースの固着性を設定に追加します。各デーモンは持続的なIPアドレスを必要とするため、他のプリミティブは仮想IPアドレスで設定します。両方のプリミティブをグループ化します。

1. クラスタ amsterdam のいずれかのノード上で、シェルを起動して、root または同等の権限でログインします。
2. 「crm configure」と入力して、対話式crmシェルに切り替えます。
3. 以下のように入力して、両方のプリミティブリソースを作成して、これらのリソースを g-booth という1つのグループに追加します。

```
crm(live)configure# primitive ip-booth ocf:heartbeat:IPaddr2 \  
params iflabel="ha" nic="eth1" cidr_netmask="24" \  
params rule #cluster-name eq amsterdam ip="192.168.201.151" \  

```

```
params rule #cluster-name eq berlin ip="192.168.202.151"
crm(live)configure# primitive booth ocf:pacemaker:booth-site \
meta resource-stickiness="INFINITY" \
params config="nfs" op monitor interval="10s"
crm(live)configure# group g-booth ip-booth booth
```

この設定では、各ブースデーモンは、デーモンが実行しているノードとは関係なく、個々のIPアドレスで使用できます。

4. `show` で変更内容をレビューします。
5. すべて正しければ、`commit` で変更を送信し、`exit` でcrmライブ設定を終了します。設定がCIBに保存されます。

#### 手順 8 順序の制約の追加

チケットがサイトに付与されにもかかわらず、そのサイトのすべてのノードが `boothd` リソースグループを何らかの理由でホストできない場合は、地理的に分散したサイト間で「スプリットブレイン」が発生する可能性があります。その場合は、そのチケットの別のサイトへのフェールオーバーを安全に管理するための `boothd` インスタンスが使用できなくなっていると考えられます。チケットの同時実行違反(チケットは複数のサイトに同時に付与されている)の可能性を回避するため、順序の制約を追加します。

1. クラスタ `amsterdam` のいずれかのノード上で、シェルを起動して、`root` または同等の権限でログインします。
2. 「`crm configure`」と入力して、対話式 `crm` シェルに切り替えます。
3. 順序の制約の作成

```
crm(live)configure# order o-booth-before-nfs inf: g-booth ms_drbd_nfs:promote
```

順序付け制約の `o-booth-before-nfs` は、`g-booth` リソースグループが開始された後にのみ、リソース `ms_drbd_nfs` をマスタモードに昇格できることを指定します。

4. 特定のチケットに依存するその他のリソースについて、さらに順序の制約を定義します。
5. `show` で変更内容をレビューします。
6. すべて正しければ、`commit` で変更を送信し、`exit` でcrmライブ設定を終了します。設定がCIBに保存されます。

#### 例 9 プリミティブの順序付け制約

特定のチケットに依存するリソースがマルチステートリソースではないが、プリミティブである場合は、順序付け制約は次のようになります。



```
crm(live)configure# order o-booth-before-rsc1 inf: g-booth rsc1
```

この制約では、g-booth リソースグループの後にのみ rsc1 (ticketA に依存)を開始できることが指定されます。

## 9.3 他のクラスタサイトへのリソース設定の転送

9.1項と9.2項の説明に従って1つのクラスタサイトのリソースを設定した場合、まだ作業は完了していません。そのリソース設定をGeoクラスタの他のサイトに転送する必要があります。

この転送を簡易化するために、「すべての」クラスタサイト上で必要なリソースにタグ付けして、それらのリソースを現在のCIBからエクスポートして、別のクラスタサイトのCIBにインポートできます。**手順 9「他のクラスタサイトへのリソース設定の転送」**では、この操作の実行方法を例示しています。これは以下の前提条件に基づいています。

### 前提条件

- クラスタ amsterdam とクラスタ berlin という2つのサイトからなるGeoクラスタを使用している。
- 各サイトのクラスタ名はそれぞれの /etc/corosync/corosync.conf ファイルで指定されている。

```
totem {  
    [...]  
    cluster_name: amsterdam  
}
```

この操作は手動で実行することも(/etc/corosync/corosync.confを編集)、YaSTクラスタモジュールを使用して実行することもできます。後者の方法については、SUSE Linux Enterprise High Availability Extension 12 SP2の『Administration Guide』(<http://www.suse.com/documentation/> から入手可能)を参照してください。「Installation and Basic Setup」という章の「Defining the First Communication Channel」の手順を参照してください。

- 9.1項「DRBDのリソースと制約」と9.2項「チケットの依存関係、ブースの制約とリソース」の説明に従って、DRBDとブースに必要なリソースを設定済みである。

### 手順 9 他のクラスタサイトへのリソース設定の転送

1. クラスタ amsterdam のいずれかのノードにログインします。
2. 次のコマンドを実行してこのクラスタを開始します。



```
root # systemctl start pacemaker
```

3. 「`crm configure`」と入力して、対話式crmシェルに切り替えます。
4. Geoクラスタ全体にわたって必要なリソースと制約にタグ付けします。
  - a. 現在のCIB設定を確認します。

```
crm(live)configure# show
```

- b. 次のコマンドを入力して、`geo_resources` というタグが付けられたGeoクラスタ関連リソースをグループ化します。

```
crm(live)configure# tag geo_resources: \  
ip_nfs nfs_fs nfs_service drbd_nfs drbd_nfs_lower ms_drbd_nfs \  
ms_drbd_nfs_lower g_nfs ① \  
col_nfs_ip_with_lower col_nfs_g_with_upper col_nfs_upper_with_ip ① \  
o_lower_drbd_before_ip_nfs o_ip_nfs_before_drbd \  
o_drbd_nfs_before_svc ① \  
nfs-req-ticket-nfs ip-booth booth g-booth o-booth-before-nfs ② \  
[...] ③
```

タグ付けによって、それらのリソース間にコロケーションや順序付けの関係が生じるわけではありません。

- ① DRBDのリソースと制約。9.1項「DRBDのリソースと制約」を参照してください。
    - ② boothdのリソースと制約。9.2項「チケットの依存関係、ブースの制約とリソース」を参照してください。
    - ③ Geoクラスタのすべてのサイト上で必要な、現在のセットアップの他のリソース。
  - c. `show` で変更内容をレビューします。
  - d. 設定が希望どおりの場合は、`submit` を使用して変更内容を送信して、`exit` を実行してcrmライブシェルを終了します。
5. タグ付けされたリソースと制約を `exported.cib` という名前のファイルにエクスポートします。

```
root # crm configure show tag:geo_resources geo_resources > exported.cib
```

`crm configure show tag: TAGNAME` コマンドを実行すると、`TAGNAME` で指定したタグに属しているすべてのリソースが表示されます。

6. クラスタ `berlin` のいずれかのノードにログインして、以下の手順を実行します。

- a. 次のコマンドを実行してこのクラスタを開始します。

```
root # systemctl start pacemaker
```

- b. exported.cib ファイルをクラスタ amsterdam からこのノードにコピーします。
- c. タグ付けされたリソースと制約を exported.cib ファイルからクラスタ berlin のCIBにインポートします。

```
root # crm configure load update PATH_TO_FILE/exported.cib
```

crm configure load コマンドの update パラメータを使用している場合は、crmshはこのファイルの内容を現在のCIB設定内に統合しようとします(現在のCIBをこのファイルの内容に置き換えるのではなく)。

- d. 次のコマンドを使用して、更新されたCIB設定を表示します。

```
root # crm configure show
```

インポートされたリソースと制約がCIBに表示されます。

この設定によって得られる結果は次のとおりです。

- ticket-nfs をクラスタ amsterdam に付与する場合、リソース ip\_nfs をホストしているノードには 192.168.201.151 というIPアドレスが割り当てられます。
- ticket-nfs をクラスタ berlin に付与する場合、リソース ip\_nfs をホストしているノードには 192.168.202.151 というIPアドレスが割り当てられます。

#### 例 10 リソース内のサイト依存パラメータの参照

手順 5の例に基づいて、別のリソースのサイト固有パラメータを参照するリソースを作成することもできます(ip\_nfs のIPパラメータなど)。次の手順に従います。

1. クラスタ amsterdam 上で、ip\_nfs のIPパラメータを参照するダミーリソースを作成します。このリソースは、これらのパラメータをこのリソースの state パラメータの値として使用します。

```
crm(live)configure# primitive dummy1 ocf:pacemaker:Dummy \  
  params rule #cluster-name eq amsterdam \  
  @ip_nfs-instance_attributes-0-ip:state \  
  params rule #cluster-name eq berlin \  
  @ip_nfs-instance_attributes-1-ip:state \  
;
```

```
op monitor interval=10
```

2. dummy1 リソースが ticket-nfs に依存することを指定する制約も追加します。

```
crm(live)configure# rsc_ticket dummy1-dep-ticket-nfs \  
ticket-nfs: dummy1 loss-policy=stop
```

3. これらのリソースと制約にタグ付けします。

```
crm(live)configure# tag geo_resources_2: dummy1 \  
dummy1-dep-ticket-nfs
```

4. show コマンドを使用して変更内容を確認して、submit コマンドを使用して変更内容を送信して、exit コマンドを使用してcrmライブシェルを終了します。
5. geo\_resources\_2 というタグが付けられたリソースをクラスタ amsterdam からエクスポートして、クラスタ berlin のCIBにインポートします(手順 9のステップ 5からステップ 6.dを参考にしてください)。

この設定によって得られる結果は次のとおりです。

- ticket-nfs をクラスタ amsterdam に付与する場合は、dummy リソースをホストしているノード上に、/var/lib/heartbeat/cores/192.168.201.151 というファイルが作成されます。
- ticket-nfs をクラスタ berlin に付与する場合は、dummy リソースをホストしているノード上に、/var/lib/heartbeat/cores/192.168.202.151 というファイルが作成されます。

## 10 DNSアップデートを通じたIPアドレス再割り当ての設定

Geoクラスタの1つのサイトがダウンして、チケットフェールオーバーが発生した場合は、通常は状況に応じてネットワークルーティングを調整する必要があります(または各チケットについてネットワークフェールオーバーを事前に設定しておく必要があります)。チケットにバインドされているサービスの種類に応じて、ルーティングを再設定するための代替手段があります。ダイナミックDNSアップデートを使用して代わりにサービスのIPアドレスを変更できます。

このシナリオの場合は、以下の前提条件が満たされている必要があります。

- フェールオーバーする必要のあるサービスがホスト名にバインドされている必要があります。
- DNSサーバがダイナミックDNSアップデートをサポートするように設定されている必要があります。BIND/namedを使用してこのように設定する方法については、[named](#) のドキュメントまたは <http://www.semicomplete.com/articles/dynamic-dns-with-dhcp/> を参照してください。DNSのセットアップ方法の詳細については(ゾーンデータの動的更新を含む)、『SUSE Linux Enterprise Administration Guide』の「The Domain Name System」の章を参照してください。<http://www.suse.com/documentation/> から入手できます。
- 次の例では、更新されるゾーンの共有鍵(TSIG鍵)によってDNSアップデートが保護されていると想定しています。この鍵は、[dnssec-keygen](#) を使用して作成できます。

```
root # dnssec-keygen -a hmac-md5 -b 128 -n USER geo-update
```

詳細については、[dnssec-keygen](#) のマニュアルページまたは『SUSE Linux Enterprise Administration Guide』(<http://www.suse.com/documentation/> から入手可能)を参照してください。「The Domain Name System」という章の「Secure Transactions」という節を参照してください。

#### 例11「ダイナミックDNSアップデートに対応するためのリソース設定」で

は、[ocf:heartbeat:dnsupdate](#) リソースエージェントを使用して [nsupdate](#) コマンドを管理する方法を例示しています。このリソースエージェントはIPv4とIPv6の両方をサポートしています。

#### 例 11 ダイナミックDNSアップデートに対応するためのリソース設定

```
crm(live)configure# primitive dns-update-ip ocf:heartbeat:dnsupdate params \
  hostname="www.domain.com" ① ip="192.168.3.4" ② \
  keyfile="/etc/whereever/Kgeo-update*.key" ③ \
  server="192.168.1.1" ④ serverport="53" ⑤
```

- ① チケットとともにフェールオーバーする必要のあるサービスにバインドされたホスト名。このホスト名のIPアドレスはダイナミックDNSを通じて更新される必要があります。
- ② 移行されるサービスをホストしているサーバのIPアドレス。ここで指定されるIPアドレスはクラスタ管理下であってもかまいません。これによりローカルフェールオーバーは処理されませんが、チケットのフェールオーバー後に外部の第三者が適切なサイトに誘導されるようになります。
- ③ [dnssec-keygen](#) を使用して生成された公開鍵ファイルのパス。
- ④ アップデートの送信先となるDNSサーバのIPアドレス。サーバが指定されていない場合は、デフォルトでは正しいゾーンのマスタサーバが使用されます。
- ⑤ DNSサーバとの通信に使用されるポート。このオプションが有効となるのは、DNSサーバが指定されている場合のみです。

上記のリソース設定を使用した場合は、リソースエージェントは障害が発生したGeoクラスタサイトをDNSレコードから削除して、ダイナミックDNSアップデートを通じてサービスのIPアドレスを変更します。

## 11 Geoクラスタの管理

ブースがGeoクラスタ内の特定のチケットを管理するためには、まずチケットを手動でサイトに付与する必要があります。

### 11.1 コマンドラインから

`booth client`コマンドの概要で説明するように、`booth client`コマンドラインツールを使用して、チケットの付与、一覧表示、または取り消しを行います。`booth client`コマンドは、`boothd`が実行されているマシン上だけでなく、クラスタ内の任意のマシン上で実行できます。`booth client`コマンドは、ブース設定ファイルとローカルに指定されたIPアドレスを参照することで、「ローカル」クラスタを見つけようとしています。ブースクライアントが接続する必要のあるサイトを指定していない場合は(`-s` オプションを使用)、ブースクライアントは常にローカルサイトに接続します。



#### 注記: 構文の変更

ブースクライアントコマンドの構文は、SUSE Linux Enterprise High Availability Extension 11から簡易化されました。たとえば、`client` キーワードを `list`、`grant`、または `revoke` 操作について省略できます(`booth list`)。また、チケットを指定する際に `-t` オプションを省略できます。

以前の構文も引き続き使用できます。詳細については、ブースのマニュアルページの [Synopsis](#) のセクションを参照してください。ただし、このマニュアルに記載されている例では簡易化された構文を使用しています。

#### `booth client`コマンドの概要

##### すべてのチケットのリスト表示

```
root # booth list
ticket: ticketA, leader: none
ticket: ticketB, leader: 10.2.12.101, expires: 2014-08-13 10:28:57
```

`-s` オプションを使用して特定のサイトを指定していない場合は、チケットに関する情報がローカルブースインスタンスに対して要求されます。

## チケットをサイトに付与する

```
root # booth grant -s 192.168.201.151 ticketA
booth[27891]: 2014/08/13_10:21:23 info: grant request sent, waiting for the
result ...
booth[27891]: 2014/08/13_10:21:23 info: grant succeeded!
```

この場合、ticketAがサイト 192.168.201.151 に付与されます。-s オプションを省略した場合は、ブースは自動的に現在のサイト(ブースクライアントを実行しているサイト)に接続して、grant 操作を要求します。

チケットを付与する前に、このコマンドではサニティチェックが実行されます。同じチケットが別のサイトにすでに付与されている場合、そのことに関する警告が表示され、まず現在のサイトからチケットを取り消すように促されます。

## サイトからチケットを取り消す

```
root # booth revoke ticketA
booth[27900]: 2014/08/13_10:21:23 info: revoke succeeded!
```

ブースは当該チケットが現在付与されているサイトを確認して、ticketA に対する revoke 操作を要求します。取り消し操作は即座に実行されます。

grant 操作および revoke 操作(特定の状況下の場合)は、明確な操作の結果を返すまでにしばらく時間がかかることがあります。クライアントは、通常はチケットの timeout 値に達するまで結果を待ちますが、-w オプションが指定されている場合は、クライアントは無期限に結果を待ち続けます。正確なステータスを確認するには、ログファイルを参照するか、crm\_ticket -L コマンドを使用します。



### 警告: crm\_ticket および crm site ticket

ブースサービスが何らかの理由で実行されていない場合、crm\_ticket または crm site ticket でチケットを手動で管理することもできます。両コマンドはクラスターノードでのみ使用できます。これらのコマンドは、同じチケットが他のサイトにすでに付与されているかどうかを確認「できない」ため、手動で管理する場合は、十分に注意してこれらのコマンドを使用してください。詳細については、マニュアルページを参照してください。

ブースが起動し、実行されている限り、手動の管理には booth client のみを使用します。

最初にサイトにチケットを付与した後、ブースメカニズムがチケットを自動的に引き継いで管理します。チケットを持つサイトがサービス外である場合、チケットは有効期限後に自動的に取り消され、他のサイトに付与されます。そのチケットに依存するリソースは、チケットを持つ新しいサイトにフェールオーバーします。前にリソースを実行していたノードは、制約で設定した loss-policy によって処理されます。

#### 手順 10 チケットの手動管理

ticket-nfs をサイト amsterdam (仮想IPアドレス 192.168.201.151) からサイト berlin (仮想IPアドレス 192.168.202.151) に手動で移動する場合は、以下の手順を実行します。

1. 次のコマンドを使用して、ticket-nfs をスタンバイに設定します。

```
root # crm_ticket -t ticket-nfs -s
```

2. ticket-nfs に依存するリソースがすべて停止するか、完全に降格するまで待機します。
3. 次のコマンドを使用して、ticket-nfs をサイト amsterdam から取り消します。

```
root # booth revoke -s 192.168.201.151 ticket-nfs
```

4. チケットが元のサイトから取り消されたら、次のコマンドを使用して、このチケットをサイト berlin に付与します。

```
root # booth grant -s 192.168.202.151 ticket-nfs
```

## 11.2 HA Web Konsole (Hawk2)の使用

Hawk2は、クラスタとGeoクラスタを管理するためのWebベースのユーザインタフェースです。

### 11.2.1 Hawk2ダッシュボードを使用した複数クラスタの監視

[Dashboard (ダッシュボード)]に表示されるクラスタ情報はサーバ側に保存されており、クラスタノード間で同期されます(クラスタノード間でのパスワード不要のSSHアクセスが設定されている場合)。詳細については、SUSE Linux Enterprise High Availability Extensionの『Administration Guide』(<http://www.suse.com/documentation/> から入手可能)を参照してください。「Running Cluster Reports Without root Access」という章の「Configuring a Passwordless SSH Account」という節を参照してください。ただし、Hawk2を実行するマシンは、その目的のためにクラスタの一部である必要はなく、別個の無関係のシステムで構いません。



前提条件

- Hawk2の[Dashboard (ダッシュボード)]で監視するすべてのクラスタでは、SUSE Linux Enterprise High Availability Extension 12 SP2を実行している必要があります。それより前のバージョンのSUSE Linux Enterprise High Availability Extensionを実行しているクラスタは監視できません。
- すべてのクラスタノードにあるHawk2の自己署名証明書を独自の証明書(または公式認証局によって署名された証明書)で置き換えていない場合は、「すべての」クラスタの「すべての」ノードで、少なくとも1回はHawkにログインします。証明書を検証します(または、ブラウザで例外を追加して警告をスキップします)。

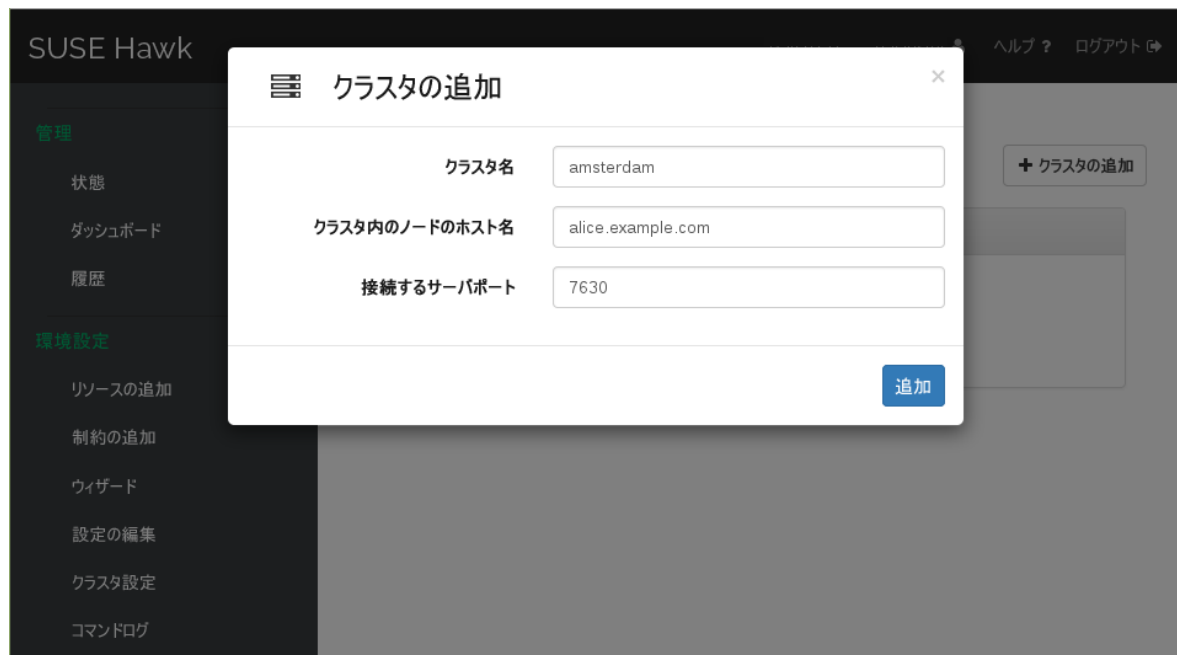
1. 複数のクラスタを監視するマシン上で、Hawk2 Webサービスを開始します。

```
root # systemctl start hawk
```

2. Webブラウザを起動して、Hawk2を実行するマシンのIPアドレスまたはホスト名をURLとして入力します。

```
https://HAWKSERVER:7630/
```

3. Hawk2 Webインタフェースにログインします。
4. 左のナビゲーションバーから、[Dashboard (ダッシュボード)]を選択します。
5. 複数のクラスタのダッシュボードを追加するため、[Add Cluster (クラスタの追加)]をクリックします。



- a. [Dashboard (ダッシュボード)]でクラスタを識別するためのカスタムクラスタ名を[クラスタ名]に入力します。
- b. クラスタ内のノードのホスト名を[Hostname of a node in the cluster (クラスタ内のノードのホスト名)]に入力します。
- c. [Add (追加)]をクリックして、変更を確認します。
- d. ダッシュボードにクラスタをさらに追加するには、もう一度[Add Cluster (クラスタの追加)]をクリックして、次のクラスタの詳細を入力します。

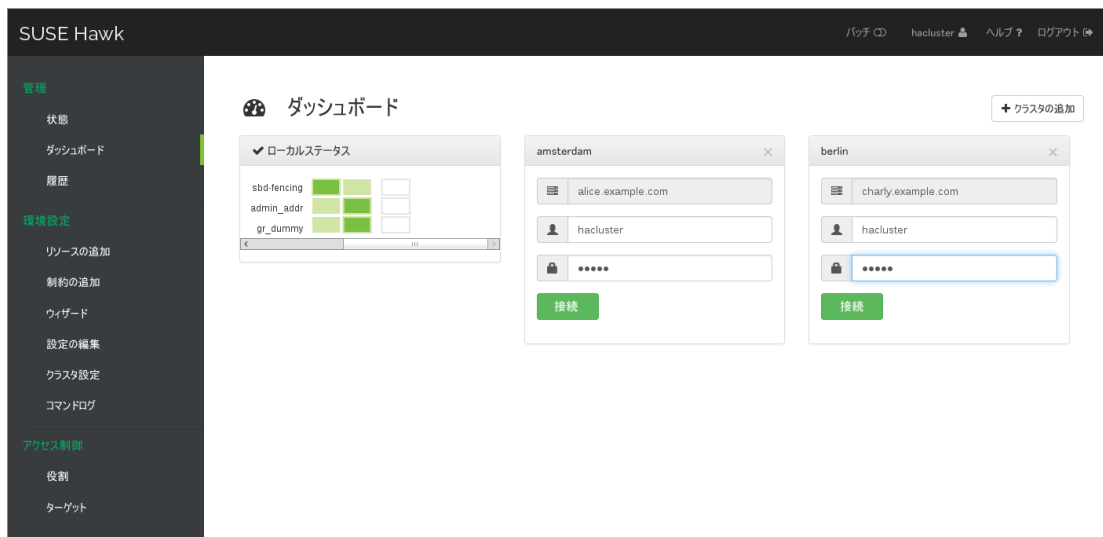


図 5 HAWK2 - ダッシュボード

6. ダッシュボードからクラスタを削除するには、クラスタの名前の横にある x アイコンをクリックします。
7. クラスタの詳細を表示するには、パスワードを入力して、[Connect (接続)] をクリックします。それぞれのノード、リソース、およびチケットの状態と共にサマリが表示されます。クラスタ名をクリックすると、クラスタの [Status (状態)] 画面が表示されます。ここから、通常の方法でクラスタを管理できます。

ノードまたはリソースに関するステータス変更は、[ Dashboard ] 内でほぼ即時に反映されます。

## 11.2.2 Hawk2を使用したチケットの管理

Hawk2でチケットを表示できるのは、それらのチケットが少なくとも1回付与または取り消された場合や、それらのチケットがチケットの依存関係で参照されている場合です。

Hawk2では、以下のチケットステータスが表示されます。

- [Granted (付与済み)]: 現在のサイトに付与されているチケット。
- [Elsewhere (他のサイト)]: 別のサイトに付与されているチケット。
- [Revoked (取り消し済み)]: 取り消されたチケット。これ以外に、Hawk2でチケットが取り消し済みとして表示されるケースは、それらのチケットがチケットの依存関係で参照されているが、どのサイトにもまだ付与されていない場合です。



## 注記: 現在のサイトへのチケットの付与とチケットの取り消し

Hawk2を使用してすべてのサイトのチケットを表示できますが、Hawk2によってトリガされた付与操作や取り消し操作は、現在のサイト(Hawk2を使用して現在接続しているサイト)のみに適用されます。Geoクラスタの別のサイトにチケットを付与するには、そのサイトに属しているいずれかのクラスタノード上でHawk2を起動します。

どのサイトにもまだ付与されていないチケットのみを付与できます。

### 手順 12 HAWK2を使用したチケットの表示、付与、および取り消し

1. Webブラウザを起動して、Hawk2にログインします。
2. 左のナビゲーションバーで、[Status (状態)]を選択します。  
クラスタノードとリソースに関する情報とともに、Hawk2には[チケット]カテゴリも表示されます。このカテゴリには、チケットのステータス、チケットを参照している制約のID、およびチケット名が表示されます。[チケット]カテゴリには、チケットを管理するための2つのカラムも含まれています。
3. 特定のチケットの詳細を表示するには、そのチケットの横にある[詳細]アイコンをクリックします。そのチケットがすでに付与されているかどうか、設定されている[Loss Policy (喪失ポリシー)] (チケットがサイトから取り消された場合の処理)、およびそのチケットに依存しているリソースがウィンドウに表示されます。

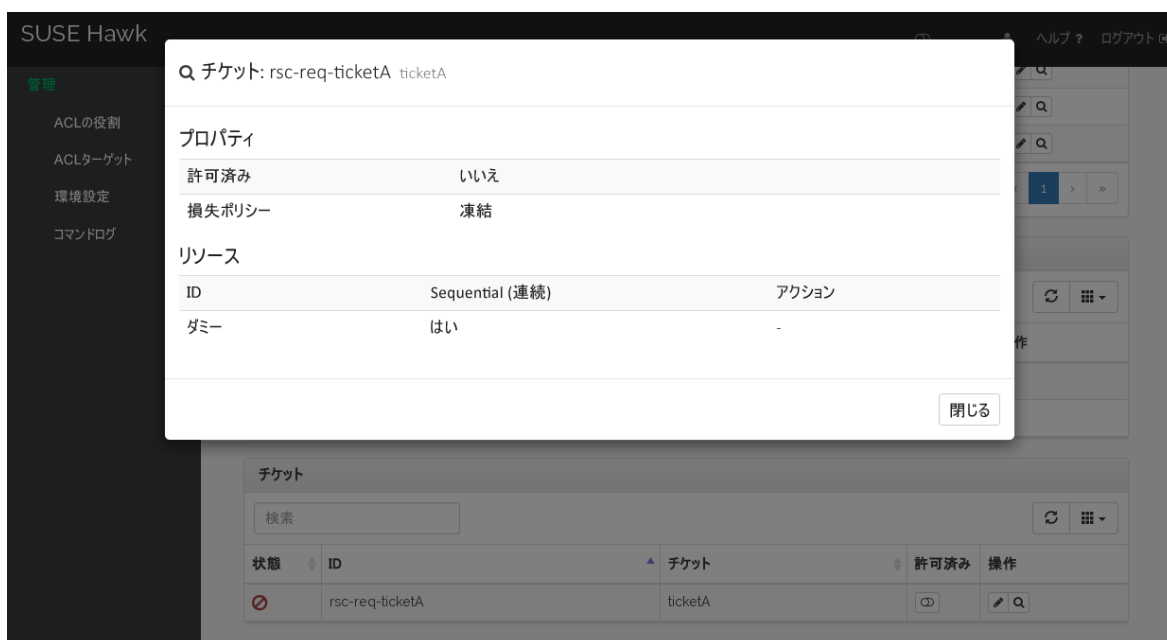


図 6 HAWK2 — チケットの詳細

4. 付与済みのチケットを現在のサイトから取り消すには、またはチケットを現在のサイトに付与するには、そのチケットの横にある[Granted (付与済み)]カラムのスイッチをクリックします。クリックすると、使用可能な操作が表示されます。Hawk2で確認プロンプトが表示されたら、選択内容を確定します。

いかなる理由でもそのチケットを付与することも取り消すこともできない場合は、Hawk2にはエラーメッセージが表示されます。チケットが正常に付与または取り消された場合は、そのチケットの[Status (ステータス)]が更新されます。

## 12 トラブルシューティング

boothではCRMと同じロギングメカニズムが使用されます。したがって、ログレベルを変更すると、boothログにも影響します。boothログメッセージにはチケットに関する情報も格納されています。

boothログメッセージとbooth設定ファイルは、どちらも `crm report` に含まれています。

boothの予期せぬ動作や何らかの問題が発生した場合は、`sudo journalctl -n`を使用してログインデータを確認するか、`crm report`を使用して詳細なクラスタレポートを作成します。

SSHを介して単一のホストからすべてのサイト(およびアービトラータ)上のクラスタノードにアクセスできる場合は、同じ `crm report` 内でこれらのクラスタノードのすべてからログファイルを収集できます。`crm report` を `-n` オプション付きで呼び出す場合は、`-n` で指定したすべてのホストからログファイルが取得されます(`-n` を指定しない場合は、このコマンドは各クラスタからノードのリストを取得しようとします)。たとえば、2つの2ノードクラスタ(`192.168.201.111|192.168.201.112` および `192.168.202.111|192.168.202.112`)とアービトラータ(`147.2.207.14`)から得られるログファイルが含まれた単一の `crm report` を作成するには、次のコマンドを使用します。

```
root # crm report -n "147.2.207.14 192.168.201.111 192.168.201.112 \
192.168.202.111 192.168.202.112" -f 10:00 -t 11:00 db-incident
```

問題がboothのみに関するものであり、boothが実行されているクラスタノード(サイト内)がわかっている場合は、それら2つのノードとアービトラータのみを指定してください。

単一のホストからすべてのサイトにアクセスできない場合は、アービトラータ上および個別サイトのクラスタノード上で、同じ期間を指定して `crm report` を個別に実行してください。アービトラータ上のログファイルを収集するには、単一ノード操作に `-S` オプションを使用する必要があります。

```
amsterdam # crm report -f 10:00 -t 11:00 db-incident-amsterdam
berlin # crm report -f 10:00 -t 11:00 db-incident-berlin
arbitrator # crm report -S -f 10:00 -t 11:00 db-incident-arbitrator
```

ただし、ログファイルが必要な対象であるすべてのマシンに関する単一の crm report を作成することが推奨されます。

## 13 最新の製品バージョンへのアップグレード

クラスタのアップグレード方法に関する一般的な指示については、SUSE Linux Enterprise High Availability Extension 12 SP2の『Administration Guide』を参照してください。<http://www.suse.com/documentation/> で入手できます。「Upgrading Your Cluster and Updating Software Packages」という章では、アップグレードプロセスを開始する前に行うべき準備についても説明しています。

表 1 SLE HAおよびSLE HA GEOでサポートされるアップグレードパス

アップグレード元とアップグレード先	アップグレードパス	詳細情報の参照先
SLE HA 11 SP3から SLE HA (Geo) 12	オフラインマイグレーション	<ul style="list-style-type: none"><li>● 基本システム: 『SUSE Linux Enterprise Server 12導入ガイド』の「Updating and Upgrading SUSE Linux Enterprise」のパート</li><li>● SLE HA: SUSE Linux Enterprise High Availability Extension 12 SP1の『Administration Guide』、「Upgrading Your Cluster and Updating Software Packages」の章の「Offline Migration」の節</li><li>● SLE HA Geo: <b>オフラインマイグレーション</b></li></ul>
SLE HA (Geo) 11 SP4からSLE HA (Geo) 12 SP1	オフラインマイグレーション	<ul style="list-style-type: none"><li>● 基本システム: 『SUSE Linux Enterprise Server 12 SP1導入ガイド』の「Updating and Upgrading SUSE Linux Enterprise」のパート</li><li>● SLE HA: SUSE Linux Enterprise High Availability Extension 12 SP1の『Administration Guide』、「Upgrading</li></ul>

アップグレード元とアップグレード先	アップグレードパス	詳細情報の参照先
		<p>Your Cluster and Updating Software Packages」の章の「Offline Migration」の節</p> <ul style="list-style-type: none"> <li>● SLE HA Geo: <a href="#">アービトレータのアップグレード</a></li> </ul>
SLE HA (Geo) 12からSLE HA (Geo) 12 SP1	ローリングアップグレード	<ul style="list-style-type: none"> <li>● 基本システム: 『SUSE Linux Enterprise Server 12 SP1導入ガイド』の「Updating and Upgrading SUSE Linux Enterprise」のパート</li> <li>● SLE HA: SUSE Linux Enterprise High Availability Extension 12 SP1の『Administration Guide』、「Upgrading Your Cluster and Updating Software Packages」の章の「Rolling Upgrade」の節</li> <li>● SLE HA Geo: <a href="#">アービトレータのアップグレード</a></li> </ul>
SLE HA (Geo) 12 SP1からSLE HA (Geo) 12 SP2	ローリングアップグレード	<ul style="list-style-type: none"> <li>● 基本システム: 『SUSE Linux Enterprise Server 12 SP2導入ガイド』の「Updating and Upgrading SUSE Linux Enterprise」のパート</li> <li>● SLE HA: 『Administration Guide for SUSE Linux Enterprise High Availability Extension 12 SP2』、「Upgrading Your Cluster and Updating Software Packages」の章の「Rolling Upgrade」の節</li> </ul>



アップグレード元とアップグレード先	アップグレードパス	詳細情報の参照先
		<ul style="list-style-type: none"> <li>• SLE HA Geo: <a href="#">アービトレータのアップグレード</a></li> <li>• DRBD 8 to DRBD 9: 『Administration Guide for SUSE Linux Enterprise High Availability Extension 12 SP2』、「DRBD」の章の「Migrating from DRBD 8 to DRBD 9」の節</li> </ul>

## 13.1 オフラインマイグレーション

この項は次のシナリオに適用されます。

- SLE HA Geo 11 SP3からSLE HA Geo 12へのアップグレード

SUSE Linux Enterprise High Availability Extension 11から11 SP3のブースバージョン(v0.1)は、Paxosアルゴリズムに基づいていました。SUSE Linux Enterprise High Availability Extension 11 SP4およびSUSE Linux Enterprise High Availability Extension 12/12 SPxのブースバージョン(v0.2)は大まかにraftに基づいているため、v0.1を実行しているものとは互換性がありません。したがって、旧ブースバージョンを実行しているシステムから新しいブースバージョンを実行しているシステムへのローリングアップグレードは不可能です。代わりに、[手順13「クラスタ全体のオフラインマイグレーション」](#)で説明されているように、すべてのクラスタノードをオフラインにしてから、クラスタ全体を移行する必要があります。

新しいマルチテナンシー機能が原因で、新しいアービトレータ初期化スクリプトはPaxos v0.1アービトレータのステータスを停止することもテストすることもできません。v0.2へのアップグレード時には、アービトレータは実行されている場合は停止されます。OCFリソースエージェント `ocf:pacemaker:booth-site` は、ブースv0.1サイトデーモンを停止および監視できます。

### 手順 13 クラスタ全体のオフラインマイグレーション

1. クラスタノードのアップグレードについては、『Administration Guide for SUSE Linux Enterprise High Availability Extension 12 SP1』で、「Upgrading Your Cluster and Updating Software Packages」の章の「Offline Migration」の節の指示に従ってください。
2. クラスタサイト外でアービトレータを使用する場合は、次のようにします。

- a. 各アービトレータを希望のSUSE Linux Enterprise Serverバージョンにアップグレードします。個々のアップグレードプロセスの詳細を確認するには、表1「SLE HAおよびSLE HA Geoでサポートされるアップグレードパス」を参照してください。
  - b. Geoクラスタリング拡張を追加して、パッケージをインストールします(1.2項「アービトレータへのパッケージのインストール」を参照)。
3. ブースの構文とコンセンサスアルゴリズムが変更されたため、最新の要件に適合するようにブース設定ファイルを更新する必要があります。以前は、セミコロン(;)を区切り文字として使用して、チケット名の後ろに有効期限と重みを付加することで、これらを希望に応じて指定可能でした。新しい構文では、すべてのチケットオプションについて別個のトークンが用意されています。詳細については、6項「ブースサービスの設定」を参照してください。デフォルトと異なる有効期限や重みを指定しなかった場合に、マルチテナンシー機能を使用しない場合は、以前の `/etc/booth/booth.conf` を引き続き使用できます。
4. すべてのクラスタサイトとアービトレータの間で更新後のブース設定ファイルを同期させます。
5. クラスタサイトとアービトレータ上でブースサービスを開始します(6.4項「ブースサービスの有効化と開始」を参照)。

## 13.2 アービトレータのアップグレード

この項は次のシナリオに適用されます。

- SLE HA Geo 11 SP4からSLE HA Geo 12 SP1へのアップグレード
- SLE HA Geo 12からSLE HA Geo 12 SP1へのアップグレード
- SLE HA Geo 12 SP1からSLE HA Geo 12 SP2へのアップグレード

### 手順 14 クラスタ外のアービトレータのアップグレード

クラスタサイト外でアービトレータを使用する場合は、各アービトレータについて以下の手順を実行します。

1. 目的のターゲットバージョンのSUSE Linux Enterprise Serverへのアップグレードを実行します。個々のアップグレードプロセスの詳細を確認するには、表1「SLE HAおよびSLE HA Geoでサポートされるアップグレードパス」を参照してください。
2. Geoクラスタリング拡張を追加して、パッケージをインストールします(1.2項「アービトレータへのパッケージのインストール」を参照)。

# A GNU Licenses

This appendix contains the GNU Free Documentation License version 1.2.

## GNU Free Documentation License

Copyright (C) 2000, 2001, 2002 Free Software Foundation, Inc. 51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA. Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

### 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

### 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

### 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## ADDENDUM: How to use this License for your documents

```
Copyright (c) YEAR YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.