



Red Hat Virtualization 4.1

テクニカルリファレンス

Red Hat Virtualization 環境の技術アーキテクチャー

Red Hat Virtualization 4.1 テクニカルリファレンス

Red Hat Virtualization 環境の技術アーキテクチャー

Red Hat Virtualization Documentation Team
Red Hat Customer Content Services
rhev-docs@redhat.com

法律上の通知

Copyright © 2016 Red Hat.

This document is licensed by Red Hat under the [Creative Commons Attribution-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/). If you distribute this document, or a modified version of it, you must provide attribution to Red Hat, Inc. and provide a link to the original. If the document is modified, all Red Hat trademarks must be removed.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux ® is the registered trademark of Linus Torvalds in the United States and other countries.

Java ® is a registered trademark of Oracle and/or its affiliates.

XFS ® is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL ® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js ® is an official trademark of Joyent. Red Hat Software Collections is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack ® Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

概要

本リファレンスガイドでは、Red Hat Virtualization 環境に採用されている概念、構成要素、技術に関して解説しています。

目次

第1章 はじめに	4
1.1. RED HAT VIRTUALIZATION MANAGER	4
1.2. RED HAT VIRTUALIZATION HOST	4
1.3. MANAGER にアクセスするためのインターフェース	6
1.4. MANAGER をサポートするコンポーネント	7
1.5. ストレージ	8
1.6. ネットワーク	9
1.7. データセンター	12
第2章 ストレージ	13
2.1. ストレージドメインの概要	13
2.2. ストレージドメインをバックアップするストレージのタイプ	13
2.3. ストレージドメインタイプ	14
2.4. 仮想ディスクイメージのストレージ形式	14
2.5. 仮想ディスクイメージ用ストレージの割り当てポリシー	15
2.6. RED HAT VIRTUALIZATION におけるストレージメタデータバージョン	15
2.7. RED HAT VIRTUALIZATION におけるストレージドメインの自動リカバリー	16
2.8. STORAGE POOL MANAGER	16
2.9. STORAGE POOL MANAGER の選択プロセス	18
2.10. RED HAT VIRTUALIZATION の排他的なリソースおよび SANLOCK	19
2.11. シンプロビジョニングとストレージのオーバーコミット	20
2.12. 論理ボリュームの拡張	20
第3章 ネットワーク	22
3.1. ネットワークアーキテクチャー	22
3.2. ネットワークの基礎用語	22
3.3. ネットワークインターフェースコントローラー	22
3.4. ブリッジ	22
3.5. ボンディング	23
3.6. ボンディング用のスイッチ設定	24
3.7. 仮想ネットワークインターフェースカード	24
3.8. 仮想 LAN (VLAN)	26
3.9. ネットワークラベル	26
3.10. クラスタネットワーク	27
3.11. 論理ネットワーク	28
3.12. 必須ネットワーク、任意ネットワーク、仮想マシンネットワーク	30
3.13. 仮想マシンの接続性	31
3.14. ポートミラーリング	31
3.15. ホストのネットワーク構成	32
3.16. ブリッジの構成	32
3.17. VLAN の構成	32
3.18. ブリッジとボンディングの構成	33
3.19. 複数のブリッジ、複数の VLAN、NIC の構成	34
3.20. 複数のブリッジ、複数の VLAN、ボンディングの構成	34
第4章 電源管理	36
4.1. 電源管理とフェンシングについて	36
4.2. RED HAT VIRTUALIZATION におけるプロキシーを使用した電源管理	36
4.3. 電源管理	36
4.4. フェンシング	37
4.5. ホストのソフトフェンシング	38
4.6. 複数の電源管理フェンスエージェントの使用	39

第5章 負荷分散、スケジューリング、移行	40
5.1. 負荷分散、スケジューリング、移行	40
5.2. 負荷分散ポリシー	40
5.3. 負荷分散ポリシー: VM_EVENLY_DISTRIBUTED	40
5.4. 負荷分散ポリシー: EVENLY_DISTRIBUTED	41
5.5. 負荷分散ポリシー: POWER_SAVING	41
5.6. 負荷分散ポリシー: NONE	41
5.7. 高可用性仮想マシンの確保	41
5.8. スケジューリング	41
5.9. 移行	42
第6章 ディレクトリーサービス	43
6.1. ディレクトリーサービス	43
6.2. ローカル認証: 内部ドメイン	43
6.3. GSSAPI を使用したリモート認証	43
第7章 テンプレートとプール	45
7.1. テンプレートとプール	45
7.2. テンプレート	45
7.3. プール	46
第8章 仮想マシンのスナップショット	47
8.1. スナップショット	47
8.2. RED HAT VIRTUALIZATION でのライブスナップショット	47
8.3. スナップショットの作成	48
8.4. スナップショットのプレビュー	49
8.5. スナップショットの削除	50
第9章 ハードウェアのドライバーとデバイス	52
9.1. 仮想ハードウェア	52
9.2. RED HAT VIRTUALIZATION における不変のデバイスアドレス	52
9.3. CPU (CENTRAL PROCESSING UNIT)	52
9.4. システムデバイス	53
9.5. ネットワークデバイス	53
9.6. グラフィックデバイス	53
9.7. ストレージデバイス	54
9.8. サウンドデバイス	54
9.9. シリアルドライバー	54
9.10. バルーンドライバー	54
第10章 最小要件および技術的な制限事項	55
10.1. 最小要件およびサポート制限事項	55
10.2. データセンターの制限事項	55
10.3. クラスターの制限事項	55
10.4. ストレージドメインの制限事項	55
10.5. RED HAT VIRTUALIZATION MANAGER の制限事項	56
10.6. ハイパーバイザーホストの要件	57
10.7. ゲストの要件とサポート制限事項	60
10.8. SPICE の制限事項	60
10.9. その他の参考資料	61

第1章 はじめに

1.1. RED HAT VIRTUALIZATION MANAGER

Red Hat Virtualization Manager により、仮想化環境を一元管理することができます。Red Hat Virtualization Manager へのアクセスには、複数の異なるインターフェースを使用することができます。各インターフェースは、それぞれ異なる方法で仮想化環境へのアクセスを円滑化します。

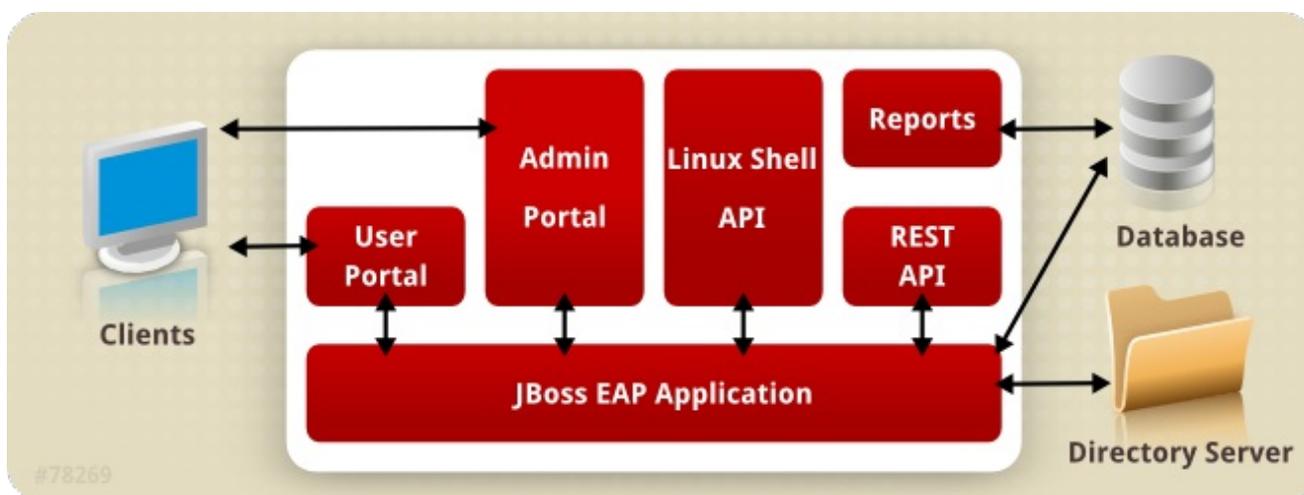


図1.1 Red Hat Virtualization Manager のアーキテクチャー

Red Hat Virtualization Manager はグラフィカルインターフェースおよび *Application Programming Interface* (API) を提供しています。各インターフェースは Manager つまり *Red Hat JBoss Enterprise Application Platform* の埋め込みインスタンスにより提供されるアプリケーションに接続します。Red Hat Virtualization Manager をサポートするコンポーネントは、Red Hat JBoss Enterprise Application Platform の他にも数多くあります。

1.2. RED HAT VIRTUALIZATION HOST

Red Hat Virtualization 環境にはホストが 1 台または複数アタッチされます。ホストとは、仮想マシンが使用する物理ハードウェアを提供するサーバーです。

Red Hat Virtualization Host (RHVH) は、仮想化ホスト作成用にカスタマイズされた特殊なインストールメディアを使用してインストールする、最適化されたオペレーティングシステムを実行します。

Red Hat Enterprise Linux ホストは、標準の Red Hat Enterprise Linux オペレーティングシステムを実行するサーバーで、インストール後に、ホストとしての使用を許可するように設定されています。

どちらのホストのインストール方法を使用しても、同じように仮想化環境内の他のリソースと対話するホストが作成されるので、いずれも **ホスト** と呼ばれます。

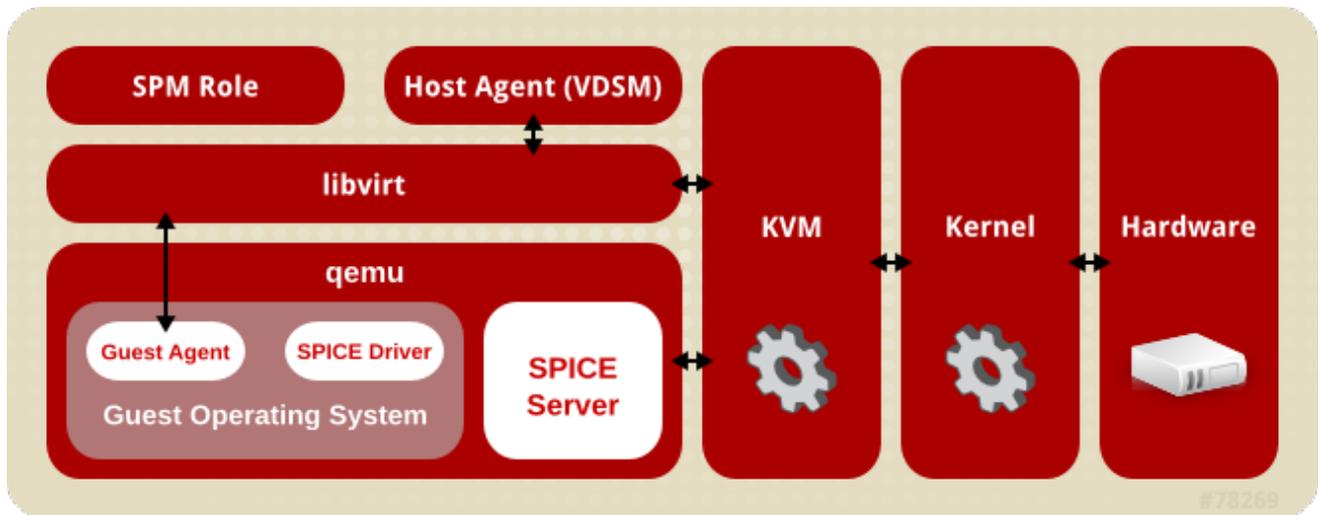


図1.2 ホストのアーキテクチャー

Kernel-based Virtual Machine (KVM)

Kernel-based Virtual Machine (KVM) は、Intel VT または AMD-V ハードウェア拡張機能を使用して完全仮想化を提供するロード可能なカーネルモジュールです。KVM 自体はカーネルスペース内で稼働しますが、その上で実行されるゲストは、ユーザースペース内の個別の QEMU プロセスとして稼働します。KVM によりホストは仮想マシンに対して物理ハードウェアを提供することができます。

QEMU

QEMU は、完全なシステムエミュレーションを提供するマルチプラットフォームのエミュレーターです。QEMU は、完全なシステム (例: 1 基または複数のプロセッサを搭載した周辺機器付きの PC) をエミュレーションします。QEMU は異なるオペレーティングシステムの起動やシステムコードのデバッグに使用することができます。QEMU は、KVM および適切な仮想化拡張機能付きのプロセッサと連動して、完全なハードウェア仮想化支援機能を提供します。

VDSM - Red Hat Virtualization Manager ホストエージェント

Red Hat Virtualization では、VDSM が仮想マシンおよびストレージ上のアクションを開始し、ホスト間の通信も円滑化します。VDSM は、メモリーやストレージ、ネットワークなどのホストのリソースをモニタリングします。また、VDSM は、仮想マシンの作成、統計の集計、ログの収集などのタスクにも対応します。VDSM インスタンスは各ホスト上で実行され、再設定可能なポート **54321** を使用して Red Hat Virtualization Manager から管理コマンドを受信します。

VDSM-REG

VDSM は、VDSM-REG を使用して各ホストを Red Hat Virtualization Manager に登録します。VDSM-REG は、ポート **80** またはポート **443** を使用して VDSM-REG 自体の情報とそのホストに関する情報を提供します。

libvirt

libvirt は、仮想マシンおよびそれらに関連付けられた仮想デバイスの管理を容易にします。Red Hat Virtualization Manager が仮想マシンのライフサイクルコマンド (起動/停止/再起動) を開始する際には、VDSM が適切なホストマシン上に libvirt を呼び出してそれらのコマンドを実行します。

SPM - Storage Pool Manager

Storage Pool Manager (SPM) はデータセンター内の 1 台のホストに割り当てられるロールです。SPM ホストには、データセンターの全ストレージドメインの構造メタデータを変更する単独の権限が付与されます。これには、仮想ディスクイメージ、スナップショット、テンプレートの作成/削除/操作が含まれます。また、Storage Area Network (SAN) 上のスパーブロックデバイスの割り当て

も含まれます。SPM のルールは、同じデータセンター内の任意のホストに移行することが可能です。このため、ホストはすべて、データセンターで定義された全ストレージドメインにアクセスできる必要があります。

Red Hat Virtualization Manager は SPM を常に稼働状態に保ちます。ストレージ接続エラーが発生した場合には Manager が SPM ロールを別のホストに再割り当てします。

ゲストオペレーティングシステム

ゲストオペレーティングシステムは、修正を加えずに Red Hat Virtualization 環境内の仮想マシンにインストールすることができます。ゲストオペレーティングシステムおよびゲスト上のすべてのアプリケーションは、仮想化環境であることを意識せずに通常どおりに稼働します。

Red Hat は、仮想化デバイスへのアクセスを高速化/効率化することができる拡張デバイスドライバーを提供しています。Red Hat Virtualization Guest Agent をゲストにインストールして、管理コンソールに詳細なゲスト情報を提供することも可能です。

1.3. MANAGER にアクセスするためのインターフェース

ユーザーポータル

デスクトップ仮想化により、パーソナルコンピュータのデスクトップ環境と同様のデスクトップ環境がユーザーに提供されます。ユーザーポータルは、仮想デスクトップインフラストラクチャーをユーザーに提供するためのインターフェースです。ユーザーは Web ブラウザーを使用してユーザーポータルにアクセスし、自分に割り当てられた仮想デスクトップを表示してアクセスします。ユーザーポータルでユーザーが使用できるアクションは、システム管理者によって設定されます。標準ユーザーは、システム管理者により割り当てられたデスクトップを起動、停止、使用することができます。パワーユーザーは、一部の管理操作を行うことができます。いずれのタイプのユーザーも、同じ URL からユーザーポータルにアクセスします。ユーザーポータルには、ログイン時のパーミッションレベルに応じて、適切なオプションが表示されます。

● 標準ユーザーによるアクセス

標準ユーザーは、ユーザーポータルを介して仮想デスクトップの電源を投入/切断したり、接続したりすることができます。Simple Protocol for Independent Computing Environments (SPICE) または Virtual Network Computing (VNC) クライアントを使用して、仮想マシンへ容易に直接接続することができます。いずれのプロトコルも、ローカルにインストールしたデスクトップ環境と同様の環境をユーザーに提供します。仮想マシンの接続に使用するプロトコルは、仮想マシンの作成時に管理者が指定します。

ユーザーポータルで利用可能な操作やサポートされているブラウザーおよびクライアントに関する詳しい情報は、『[ユーザーポータルの概要](#)』を参照してください。

● パワーユーザーによるアクセス

Red Hat Virtualization ユーザーポータルは、仮想リソースの作成、使用、モニタリングを行うためのグラフィカルユーザーインターフェースをパワーユーザーに提供します。システム管理者は、ユーザーにパワーユーザーのアクセスを付与することによって、一部の管理タスクを委任することができます。また、パワーユーザーは、標準ユーザーが実行できるタスクに加えて、以下のような操作を行うことができます。

- 仮想マシンの作成/編集/削除
- 仮想ディスクおよびネットワークインターフェースの管理
- 仮想マシンに対するユーザーパーミッションの割り当て

- 仮想マシンを迅速にデプロイするためのテンプレートの作成/使用
- リソースの使用状況および重大度が高いイベントのモニタリング
- 仮想マシンを以前の状態に復元するためのスナップショットの作成/使用

パワーユーザーは、委任された仮想マシンの管理タスクを実行することができます。データセンターおよびクラスターレベルの管理タスクは、その環境の管理者に確保されます。

管理ポータル

管理ポータルは、Red Hat Virtualization Manager サーバーを管理するためのグラフィカルインターフェースです。これにより管理者は、Web ブラウザーを使用して、仮想環境内の全要素をモニタリング/作成/維持管理することができます。管理ポータルでは以下のようなタスクを実行することができます。

- 仮想インフラストラクチャー (ネットワーク、ストレージドメイン) の作成と管理
- ホストのインストールと管理
- 論理エンティティ (データセンター、クラスター) の作成と管理
- 仮想マシンの作成と管理
- Red Hat Virtualization のユーザーおよびパーミッションの管理

管理ポータルは、JavaScript を使用して表示されます。

管理ポータルの機能は『[Red Hat Virtualization 管理ガイド](#)』で詳しく説明されています。管理ポータルがサポートするブラウザーおよびプラットフォームに関する情報は『[Red Hat Virtualization インストールガイド](#)』を参照してください。

Representational State Transfer (REST) API

Red Hat Virtualization REST API は、Red Hat Virtualization 環境における問い合わせおよび制御を行うためのソフトウェアインターフェースを提供します。REST API は、HTTP アクションをサポートする任意のプログラミング言語で使用することができます。

REST API により、開発者および管理者は以下のような作業を行うことができます。

- エンタープライズ IT システムとの統合
- サードパーティー製の仮想化ソフトウェアとの統合
- 自動メンテナンスおよびエラーチェックタスクの実行
- Red Hat Virtualization 環境内の反復タスクを自動化するためのスクリプトの使用

API の仕様や使用例は『[Red Hat Virtualization REST API ガイド](#)』を参照してください。

1.4. MANAGER をサポートするコンポーネント

Red Hat JBoss Enterprise Application Platform

Red Hat JBoss Enterprise Application Platform は Java アプリケーションサーバーです。クロスプラットフォームの Java アプリケーションの効率的な開発とデリバリーをサポートするフレームワークを提供します。Red Hat Virtualization Manager のデリバリーには Red Hat JBoss Enterprise

Application Platform を使用します。



重要

Red Hat Virtualization Manager に同梱されている Red Hat JBoss Enterprise Application Platform のバージョンは、Red Hat Virtualization Manager にサービスを提供する目的でカスタマイズされているため、その他のアプリケーションへのサービス提供には使用できません。Manager に同梱されている Red Hat JBoss Enterprise Application Platform を別の目的に使用すると、Red Hat Virtualization 環境へのサービス提供機能に悪影響を及ぼします。

レポートおよび履歴データの収集

Red Hat Virtualization Manager には、ホスト/仮想マシン/ストレージのモニタリングデータを収集する Data Warehouse が含まれており、数多くの事前定義済みレポートを利用することができます。ユーザーは SQL をサポートする任意のクエリーツールを使用して環境を分析したり、レポートを作成したりすることができます。

Red Hat Virtualization Manager のインストールプロセスでは 2 つのデータベースが作成されます。これらのデータベースは、インストール中に選択した Postgres インスタンス上に作成されます。

- engine データベースは、Red Hat Virtualization Manager が使用するプライマリデータストアです。仮想環境の状態、設定、パフォーマンスについての情報は、このデータベースに保管されます。
- ovirt_engine_history データベースには、engine オペレーションデータベースから経時的に収集した設定情報および統計メトリックが格納されます。engine データベース内の設定データは、毎分チェックされ、変更内容は ovirt_engine_history データベースに複製されます。データベースに加えられた変更内容をトラッキングすることによって、データベース内のオブジェクトに関する情報が提供されます。この情報により、Red Hat Virtualization 環境のパフォーマンスを分析して強化し、問題を解決することができます。

ovirt_engine_history データベースをベースとしたレポート生成に関する詳しい説明は『Red Hat Virtualization Data Warehouse ガイド』の「[履歴データベース](#)」を参照してください。



重要

ovirt_engine_history データベースへのデータ複製は、**RHEVM History Service** (ovirt-engine-dwhd) によって実行されます。

ディレクトリーサービス

ディレクトリーサービスは、ユーザーおよび組織の情報を保管するための一元化されたネットワークベースのストレージを提供します。保管される情報の種類には、アプリケーションの設定、ユーザープロファイル、グループデータ、ポリシー、アクセス制御などが含まれます。Red Hat Virtualization Manager は Active Directory、Identity Management (IdM)、OpenLDAP、および Red Hat Directory Server 9 をサポートしています。また管理目的専用のローカルの内部ドメインもあります。この内部ドメインのユーザーは admin ユーザーの 1 つのみです。

1.5. ストレージ

Red Hat Virtualization は、仮想ディスクイメージ、テンプレート、スナップショット、ISO ファイルなどの格納に一元化されたストレージシステムを使用します。ストレージとは論理的にグループ化されて

ストレージプールの中で、このストレージプールはストレージドメインで構成されます。ストレージドメインとは、イメージを格納するストレージ容量とストレージの内部構造を記述するメタデータの組み合わせです。ストレージドメインには、データ、エクスポート、ISO の 3 タイプがあります。

データストレージドメインは、各データセンターが必要とする唯一のストレージドメインです。データストレージドメインは単一のデータセンター専用ですが、エクスポートドメインおよび ISO ドメインはオプションです。ストレージドメインは共有リソースなので、データセンター内の全ホストがアクセス可能である必要があります。

ストレージネットワークは Network File System (NFS)、Internet Small Computer System Interface (iSCSI)、GlusterFS、Fibre Channel Protocol (FCP) または POSIX 準拠のネットワークファイルシステムのいずれかを使用して実装することができます。

NFS (およびその他の POSIX 準拠のファイルシステム) では、仮想ディスク、テンプレート、スナップショットはすべて単純なファイルです。

SAN (iSCSI/FCP) ドメインでは、論理ボリュームマネージャー (LVM) によってブロックデバイスが 1 つのボリュームグループ (VG) に集約されます。各仮想ディスク、テンプレート、スナップショットは、VG 上の論理ボリューム (LV) です。LVM についての詳しい情報は、『[Red Hat Enterprise Linux 論理ボリュームマネージャーの管理ガイド](#)』を参照してください。

データストレージドメイン

データドメインは、環境内で稼働中の全仮想マシンの仮想ハードディスクイメージを格納します。仮想マシンのテンプレートおよびスナップショットもデータドメインに保管されます。データドメインは異なるデータセンター間では共有できません。

エクスポートストレージドメイン

エクスポートドメインは、データセンターと Red Hat Virtualization 環境の間でイメージをコピーしたり移動したりするのに使用する一時的なストレージリポジトリです。エクスポートドメインは仮想マシンとテンプレートのバックアップにも使用できます。エクスポートドメインは、データセンター間で移動することが可能ですが、一度に 1 つのデータセンターでしかアクティブにできません。

ISO ストレージドメイン

ISO ドメインは、ISO ファイル (仮想マシンにオペレーティングシステムやアプリケーションをインストールするのに使用する論理 CD-ROM) を格納します。ISO ドメインは、物理 CD-ROM/DVD のライブラリーの代わりとなる論理エンティティとして、データセンターでの物理メディアの必要性を排除します。ISO ドメインは異なるデータセンター間で共有することができます。

1.6. ネットワーク

Red Hat Virtualization ネットワークアーキテクチャーは、Red Hat Virtualization 環境の異なる要素間の接続を円滑化します。ネットワークアーキテクチャーはネットワーク接続をサポートするだけでなく、ネットワークの分離を可能にします。

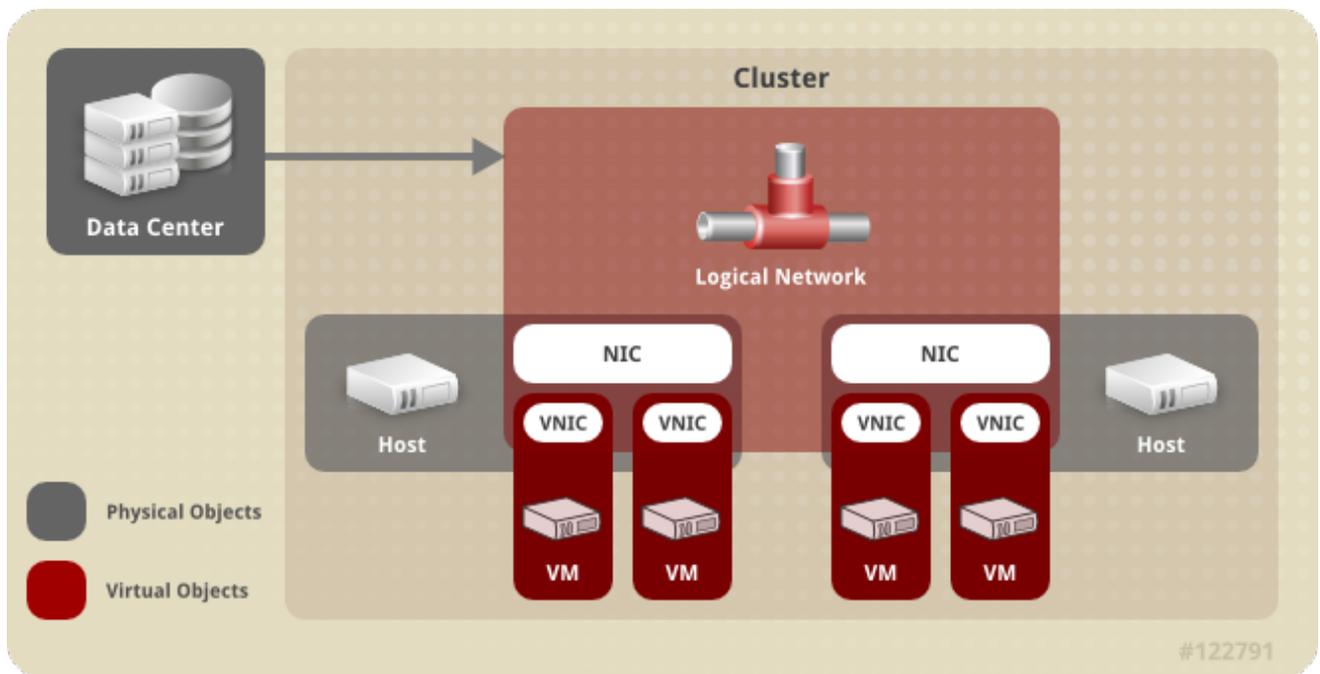


図1.3 ネットワークアーキテクチャー

Red Hat Virtualization では、ネットワークは複数の層に定義されます。配下の物理ネットワークインフラストラクチャーは、Red Hat Virtualization 環境のハードウェアと論理コンポーネントの間での接続を可能にするように配置/設定される必要があります。

ネットワークインフラストラクチャー層

Red Hat Virtualization のネットワークアーキテクチャーは、複数の共通ハードウェア/ソフトウェアデバイスに依存します。

- *Network Interface Controller* (NIC): ホストをネットワークに接続する物理ネットワークインターフェースデバイス。
- *仮想 NIC* (VNIC): ホストの物理 NIC を使用して稼働する論理 NIC。仮想マシンへのネットワーク接続を提供します。
- *ボンディング*: 複数の NIC を単一のインターフェースにバインドします。
- *ブリッジ*: パケット交換網のパケット転送技術です。仮想論理ネットワークの基礎を形成します。

論理ネットワーク

論理ネットワークでは、環境要件に基づいてネットワークトラフィックを分けることができます。論理ネットワークの種類は次のとおりです。

- 仮想マシンネットワークトラフィックを伝送する論理ネットワーク
- 仮想マシンネットワークトラフィックを伝送しない論理ネットワーク
- 任意の論理ネットワーク
- 必須ネットワーク

すべての論理ネットワークは、必須または任意のいずれかです。

仮想マシンネットワークトラフィックを伝送する論理ネットワークは、ソフトウェアブリッジデバイスとしてホストレベルで実装されます。デフォルトでは、Red Hat Virtualization Manager のインストール中に1つの論理ネットワーク (**ovirtmgmt** 管理ネットワーク) が定義されます。

管理者が追加できる他の論理ネットワークは、ストレージ専用の論理ネットワークとディスプレイ専用の論理ネットワークです。仮想マシントラフィックを伝送しない論理ネットワークでは、ホストに関連付けられたブリッジデバイスが存在せず、ホストのネットワークインターフェースに直接関連付けられます。

Red Hat Virtualization は、管理ネットワークのトラフィックから移行ネットワークのトラフィックを分離しています。これにより、ライブマイグレーション専用ネットワークの (ルーティングなしで) 使用が可能となり、またマイグレーションの実行中に、管理ネットワーク (**ovirtmgmt**) からハイパーバイザーへの接続が中断されないようになります。

異なる層の論理ネットワーク

論理ネットワークは、仮想環境の各層に異なる影響を及ぼします。

データセンター層

論理ネットワークはデータセンターレベルで定義されます。各データセンターには、デフォルトで **ovirtmgmt** 管理ネットワークがあります。これ以外の論理ネットワークはオプションですが、作成することをお勧めします。**仮想マシンのネットワーク** とカスタムの MTU はデータセンターレベルで設定できます。データセンターに定義された論理ネットワークは、その論理ネットワークを使用するクラスターにも追加する必要があります。

クラスター層

論理ネットワークは、データセンターから提供され、そのネットワークを使用するクラスターに追加する必要があります。デフォルトでは、各クラスターは管理ネットワークに接続されます。オプションとして、クラスターの親データセンターに定義されたクラスター論理ネットワークに追加することもできます。必須論理ネットワークがクラスターに追加された場合は、クラスター内の各ホストに対して必須論理ネットワークを実装する必要があります。必要に応じて、任意の論理ネットワークをホストに追加できます。

ホスト層

仮想マシン論理ネットワークは、該当するネットワークインターフェースに関連付けられたソフトウェアブリッジデバイスとしてクラスター内の各ホストに実装されます。仮想マシンの論理ネットワーク以外のネットワークでは、関連付けられたブリッジは存在せず、ホストのネットワークインターフェースに直接関連付けられます。Red Hat Virtualization 環境にホストを追加すると、そのホストのネットワークデバイスの1つを使用して管理ネットワークがブリッジとして実装されます。クラスターに追加されたその他の必須論理ネットワークは、クラスターに対して動作するように各ホスト上のネットワークインターフェースに関連付ける必要があります。

仮想マシン層

論理ネットワークは、物理マシンに対してネットワークを提供するのと同じ方法で、仮想マシンに提供することができます。仮想マシンの仮想 NIC は、仮想マシンを実行しているホストに実装された任意の仮想マシン論理ネットワークに接続できます。仮想マシンは、接続された論理ネットワークで利用可能な任意の他のデバイスまたは接続先に接続できます。

例1.1 管理ネットワーク

Red Hat Virtualization Manager インストール時には、**ovirtmgmt** という管理用論理ネットワークが自動的に作成されます。**ovirtmgmt** ネットワークは Red Hat Virtualization Manager とホスト間の管理トラフィック専用です。この他に特殊目的のブリッジが設定されていない場合は、**ovirtmgmt** が全トラフィックのデフォルトブリッジになります。

■

1.7. データセンター

Red Hat Virtualization において最も抽象度が高いのはデータセンターです。データセンターは次の 3 タイプのサブコンテナで構成されるコンテナです。

- ストレージコンテナは、ストレージドメインの接続情報など、ストレージタイプおよびストレージドメインに関する情報を格納します。ストレージはデータセンターに対して定義され、そのデータセンター内の全クラスターが使用可能です。1 つのデータセンター内のホストクラスターはすべて同じストレージドメインにアクセスできます。
- ネットワークコンテナは、データセンターの論理ネットワークに関する情報を格納します。これには、ネットワークアドレス、VLAN タグ、STP サポートなどの情報が含まれます。論理ネットワークはデータセンターに対して定義され、オプションとしてクラスターレベルで実装されます。
- クラスターコンテナは、クラスターを格納します。クラスターとは、AMD または Intel のプロセッサの互換性があるプロセッサコアを搭載したホストのグループのことです。クラスターはマイグレーションドメインであり、クラスター内の任意のホストに仮想マシンをライブマイグレーションすることができますが、別のクラスターのホストには移行できません。単一のデータセンターに複数のクラスターを格納し、各クラスターを複数のホストで構成することができます。

第2章 ストレージ

2.1. ストレージドメインの概要

ストレージドメインとは、共通のストレージドインターフェースを使用するイメージの集合体です。ストレージドメインには、テンプレートおよび仮想マシン (スナップショットを含む) の完全なイメージ、ISO ファイル、およびそれら自体についてのメタデータが格納されます。ストレージドメインには、ブロックデバイス (SAN: iSCSI または FCP) またはファイルシステム (NAS: NFS、GlusterFS またはその他の POSIX 準拠ファイルシステム) を使用することができます。

NAS では、仮想ディスク、テンプレート、スナップショットはすべてファイルです。

SAN (iSCSI/FCP) では、仮想ディスク、テンプレート、スナップショットはそれぞれが 1 つの論理ボリュームです。ブロックデバイスは、ボリュームグループと呼ばれる単一の論理エンティティに集約された後に、仮想ハードディスクとして使用するように、LVM (論理ボリュームマネージャー) によって分割されます。LVM に関する詳しい情報は『[Red Hat Enterprise Linux 論理ボリュームマネージャーの管理ガイド](#)』を参照してください。

仮想ディスクには 2 つの形式 (QCOW2 または RAW) のいずれかを使用することができます。ストレージのタイプは、スパースまたは事前割り当て済みに指定することができます。スナップショットは常にスパースですが、いずれの形式のディスクのスナップショットも作成することができます。

同じストレージドメインを共有する仮想マシンは、同じクラスターに属するホスト間で移行することができます。

2.2. ストレージドメインをバックアップするストレージのタイプ

ストレージドメインは、ブロックベースおよびファイルベースのストレージを使用して実装することができます。

ファイルベースのストレージ

Red Hat Virtualization がサポートするファイルベースのストレージタイプには、NFS、GlusterFS、その他の POSIX 準拠のファイルシステム、ホストのローカルストレージがあります。

ファイルベースのストレージは、Red Hat Virtualization 環境の外部で管理されます。

NFS ストレージは、Red Hat Enterprise Linux NFS サーバーまたはその他のサードパーティー製 NAS サーバーで管理されます。

ホストは、独自のローカルストレージファイルシステムを管理することができます。

ブロックベースのストレージ

ブロックストレージは、未フォーマットのブロックデバイスを使用します。ブロックデバイスは LVM (論理ボリュームマネージャー) により **ボリュームグループ** に集約されます。LVM のインスタンスは全ホストで実行され、各 LVM インスタンスは他のホストで実行している LVM インスタンスを認識しません。VDSM は、ボリュームグループの変更をスキャンすることにより、クラスター化のロジックを LVM 上に追加します。変更が検出されると、VDSM はボリュームグループ情報をリフレッシュするよう各ホストに指示して、それらのホストを更新します。ホストは、ボリュームグループを論理ボリュームに分割し、論理ボリュームのメタデータをディスクに書き込みます。既存のストレージドメインにストレージ容量が追加された場合には、Red Hat Virtualization Manager は各ホストの VDSM を使用してボリュームグループ情報を最新の状態に更新します。

論理ユニット番号 (LUN) は個別のブロックデバイスです。LUN への接続には、サポートされているブロックストレージプロトコル (iSCSI、FCoE、SAS のいずれか) を使用することができます。Red

Hat Virtualization Manager は、LUN へのソフトウェアの iSCSI 接続を管理します。その他すべてのブロックストレージ接続は、Red Hat Virtualization 環境外で管理されます。論理ボリュームの作成、拡張、削除や新規 LUN の追加など、ブロックベースのストレージ環境における変更は、専用に選択された Storage Pool Manager と呼ばれるホストの LVM によって処理されます。この変更は、VDSM によって同期され、クラスター内の全ホストにわたって更新されます。

2.3. ストレージドメインタイプ

Red Hat Virtualization でサポートされる 3 種類のストレージドメイン、および各ストレージドメインでサポートされるストレージのタイプは、以下のとおりです。

- データストレージドメインには、Red Hat Virtualization 環境の全仮想マシンのハードディスクイメージを格納します。ディスクイメージには、インストールされているオペレーティングシステム、仮想マシンに保管されているデータ、仮想マシンによって生成されたデータなどが含まれる場合があります。データストレージドメインは、NFS、iSCSI、FCP、GlusterFS、POSIX 準拠のストレージをサポートしています。データドメインは、複数のデータセンター間では共有できません。
- エクスポートストレージドメインは、データセンター間で移動するハードディスクイメージや仮想マシンテンプレート用の一時的なストレージを提供します。また、エクスポートストレージドメインは、仮想マシンのバックアップコピーを格納します。エクスポートストレージドメインは NFS ストレージをサポートしています。単一のエクスポートストレージドメインに複数のデータセンターがアクセスすることが可能ですが、一度に使用できるのは 1 つのデータセンターのみです。
- ISO ストレージドメインは、イメージとも呼ばれる ISO ファイルを格納します。ISO ファイルは物理 CD/DVD の代わりとなります。Red Hat Virtualization 環境で一般的な ISO ファイルのタイプには、オペレーティングシステムのインストールディスク、アプリケーションのインストールディスク、ゲストエージェントのインストールディスクなどがあります。物理ディスクをディスクドライブに挿入して起動するのと同じように、これらのイメージを仮想マシンにアタッチして起動することができます。ISO ストレージドメインにより、データセンター内の全ホストが ISO を共有できるため、物理的な光学メディアを用意する必要がなくなります。

2.4. 仮想ディスクイメージのストレージ形式

QCOW2 形式の仮想マシンストレージ

QCOW2 は、仮想ディスクイメージ用のストレージフォーマットです。QCOW は *QEMU Copy On Write* の略です。QCOW2 フォーマットは、論理/物理ブロック間のマッピングを追加することにより、仮想層から物理ストレージ層を分離します。各論理ブロックはその物理オフセットにマッピングされます。このマッピングによりストレージのオーバーコミットと仮想マシンのスナップショットが可能となり、各 QCOW ボリュームが配下のディスクイメージに加えられた変更のみを表示します。

最初のマッピングは全論理ブロックからバックアップファイルまたはボリューム内のオフセットをポイントします。仮想マシンがスナップショットの後にデータを QCOW2 ボリュームに書き込むと、対象のブロックはバックアップボリュームから読み込まれ、新たな情報で修正されてから、新しいスナップショット QCOW2 ボリュームに書き込まれます。この後にマッピングが新たな場所をポイントするように更新されます。

RAW

RAW ストレージ形式は、QCOW2 よりもパフォーマンス面で優れており、RAW 形式で保管されている仮想ディスクイメージにはフォーマットが適用されません。RAW 形式で保管されているディスクイメージ上の仮想マシンデータの操作には、ホストからの追加処理は必要ありません。仮

想マシンが仮想ディスク内の特定のオフセットにデータを書き込むと、その I/O は、バッキングファイルまたは論理ボリュームの同じオフセットに書き込まれます。

外部で管理されている、ストレージレイからのシンプロビジョニングされた LUN を使用していない限り、RAW 形式では、定義されたイメージの全領域が事前割り当て済みである必要があります。

2.5. 仮想ディスクイメージ用ストレージの割り当てポリシー

事前割り当てストレージ

仮想ディスクイメージに必要なストレージはすべて、その仮想マシンの作成前に割り当てられます。仮想マシン用に 20 GB のディスクイメージが作成された場合は、そのディスクイメージは 20 GB のストレージドメイン容量を使用します。事前に割り当てられたディスクイメージは、拡張できません。ストレージを事前に割り当てておくと、ランタイム中にはストレージ割り当てが行われないため、書き込み時間が短縮されますが、柔軟性が犠牲になります。このようなストレージ割り当ての場合は、Red Hat Virtualization Manager がストレージをオーバーコミットする能力が低くなります。事前割り当てストレージは、ストレージのレイテンシーに対応しにくい I/O タスクを集中的に処理する仮想マシンに推奨されます。通常、サーバー仮想マシンがこのような条件に該当します。



注記

ストレージバックエンドによって提供されるシンプロビジョニング機能を使用している場合でも、仮想マシン用にストレージをプロビジョニングする際に管理ポータルから事前割り当てストレージを選択する必要があります。

スパース割り当てストレージ

仮想ディスクイメージの上限サイズは、仮想マシンの作成時に設定されます。最初は、ディスクイメージはストレージドメインの容量を使用しませんが、仮想マシンがデータをディスクに書き込むと、上限に達するまで使用量は増えていきます。ディスクイメージ内のデータが削除されても、ストレージドメインに容量は戻りません。スパース割り当てのストレージは、ストレージの遅延に対してある程度の耐性があり、集中度が低/中程度の I/O タスクに使用する仮想マシンに適しています。通常、デスクトップ仮想マシンがこの条件に該当します。



注記

シンプロビジョニング機能がストレージバックエンドによって提供される場合には、このオプションを推奨のシンプロビジョニング実装として使用すべきです。ストレージはグラフィカルユーザーインターフェースから事前割り当てとしてプロビジョニングし、バックエンドソリューションにより提供されるシンプロビジョニング機能をそのまま活用します。

2.6. RED HAT VIRTUALIZATION におけるストレージメタデータバージョン

Red Hat Virtualization では、ストレージドメインに関する情報はストレージドメイン自体にメタデータとして保管されます。ストレージメタデータの実装は、Red Hat Virtualization のメジャーリリースのたびに改善されています。

- V1 メタデータ (Red Hat Virtualization 2.x シリーズ)

各ストレージドメインには、そのストレージドメイン自体の構造と、仮想ディスクイメージをバックアップしている全物理ボリュームの名前を記述したメタデータが格納されています。

マスタードメインには追加で、ストレージプール内の全ドメインと物理ボリューム名のメタデータが格納されています。このメタデータの合計サイズの上限は 2 KB で、プール内に格納できるストレージドメイン数が制限されています。

テンプレートおよび仮想マシンベースイメージは読み取り専用です。

V1 メタデータは、NFS、iSCSI、FC ストレージドメインに適用可能です。

- **V2 メタデータ (Red Hat Enterprise Virtualization 3.0)**

ストレージドメインとプールのメタデータはすべて、論理ボリュームに書き込まれるのではなく、論理ボリュームタグとして保管されます。仮想ディスクのボリュームに関するメタデータは、ドメイン上の論理ボリュームに保管されます。

物理ボリューム名は、メタデータには含まれなくなりました。

テンプレートおよび仮想マシンベースイメージは読み取り専用です。

V2 メタデータは、iSCSI および FC ストレージドメインに適用可能です。

- **V3 metadata (Red Hat Enterprise Virtualization 3.1 以降)**

ストレージドメインとプールのメタデータはすべて、論理ボリュームに書き込まれるのではなく、論理ボリュームタグとして保管されます。仮想ディスクのボリュームに関するメタデータは、ドメイン上の論理ボリュームに保管されます。

仮想マシンおよびテンプレートのベースイメージは、読み取り専用ではなくなりました。この変更により、ライブスナップショット、ライブストレージマイグレーション、スナップショットからのクローン作成が可能となりました。

英語以外の言語で記述されたボリューム名に対する Unicode メタデータのサポートが追加されました。

V3 メタデータは、NFS、GlusterFS、POSIX、iSCSI、FC ストレージドメインに適用可能です。

2.7. RED HAT VIRTUALIZATION におけるストレージドメインの自動リカバリ

Red Hat Virtualization 環境内のホストは、各ドメインからのメタデータを読み取ることにより、データセンター内のストレージドメインをモニタリングします。データセンター内の全ホストが、ストレージドメインにアクセスできないことを報告すると、そのストレージドメインは非アクティブな状態となります。

Manager は、アクティブでないストレージドメインの接続を解除する代わりに、ストレージドメインが一時的なネットワークの停止などを理由に一時的にアクティブでない状態になっていると仮定し、5 分ごとにアクティブでないストレージドメインの再アクティブ化を試みるようになりました。

管理者による操作は、ストレージ接続の中断の原因を修復するために必要となる場合がありますが、接続が復元されてからのストレージドメインの再アクティブ化は Manager が処理します。

2.8. STORAGE POOL MANAGER

Red Hat Virtualization では、ストレージドメインの内部構造の記述にメタデータを使用します。構造メタデータは各ストレージドメインのセグメントに書き込まれ、ホストは単一ライター/複数リーダーの構成をベースに、ストレージドメインのメタデータと連携します。ストレージドメインの構造メタデータは、イメージおよびスナップショットの作成/削除およびボリュームとドメインの拡張をトラッキングします。

データドメインの構造を変更することができるホストは、Storage Pool Manager (SPM) として知られています。SPM は、ディスクイメージの作成/削除、スナップショットの作成/マージ、ストレージドメイン間でのイメージのコピー、テンプレートの作成、ブロックデバイス用のストレージ割り当てなど、データセンター内におけるすべてのメタデータ変更を調整します。SPM の役割を担うホストは、1 データセンターにつき 1 台です。SPM 以外のホストはすべて、ストレージドメインの構造メタデータを読み取ることしかできません。

ホストは、手動で SPM に選択するか、Red Hat Virtualization Manager により SPM に割り当てることができます。Manager は、SPM ホストの候補に *ストレージセントリックリース* の引き継ぎを試行させることにより SPM ロールを割り当てます。このリースにより、SPM ホストはストレージメタデータの書き込みが可能になります。ストレージセントリックと呼ばれる理由は、Manager またはホストによりトラッキングされるのではなく、ストレージドメインに書き込まれるためです。ストレージセントリックリースは、マスターストレージドメインの **leases** と呼ばれる特殊な論理ボリュームに書き込まれます。データドメインの構造に関するメタデータは、**metadata** と呼ばれる特殊な論理ボリュームに書き込まれます。**leases** 論理ボリュームは、**metadata** 論理ボリュームへの変更を防ぎます。

Manager は VDSM を使用して、ホストに **spmStart** コマンドを実行します。これにより、そのホスト上の VDSM はストレージセントリックリースの引き継ぎを試みます。ホストは、引き継ぎに成功すると SPM となり、Red Hat Virtualization Manager が別のホストに SPM ロールを引き継ぐように要求するまで、ストレージセントリックリースを維持します。

Manager は、次のような場合に SPM ロールを別のホストに移行します。

- SPM ホストがマスターストレージドメインにはアクセスできるが、全ストレージドメインにアクセスできない場合。
- SPM ホストがストレージへの接続を失ったために、リースを更新できない場合、またはリース容量が満杯なために書き込み操作を実行できない場合。
- SPM ホストがクラッシュした場合。

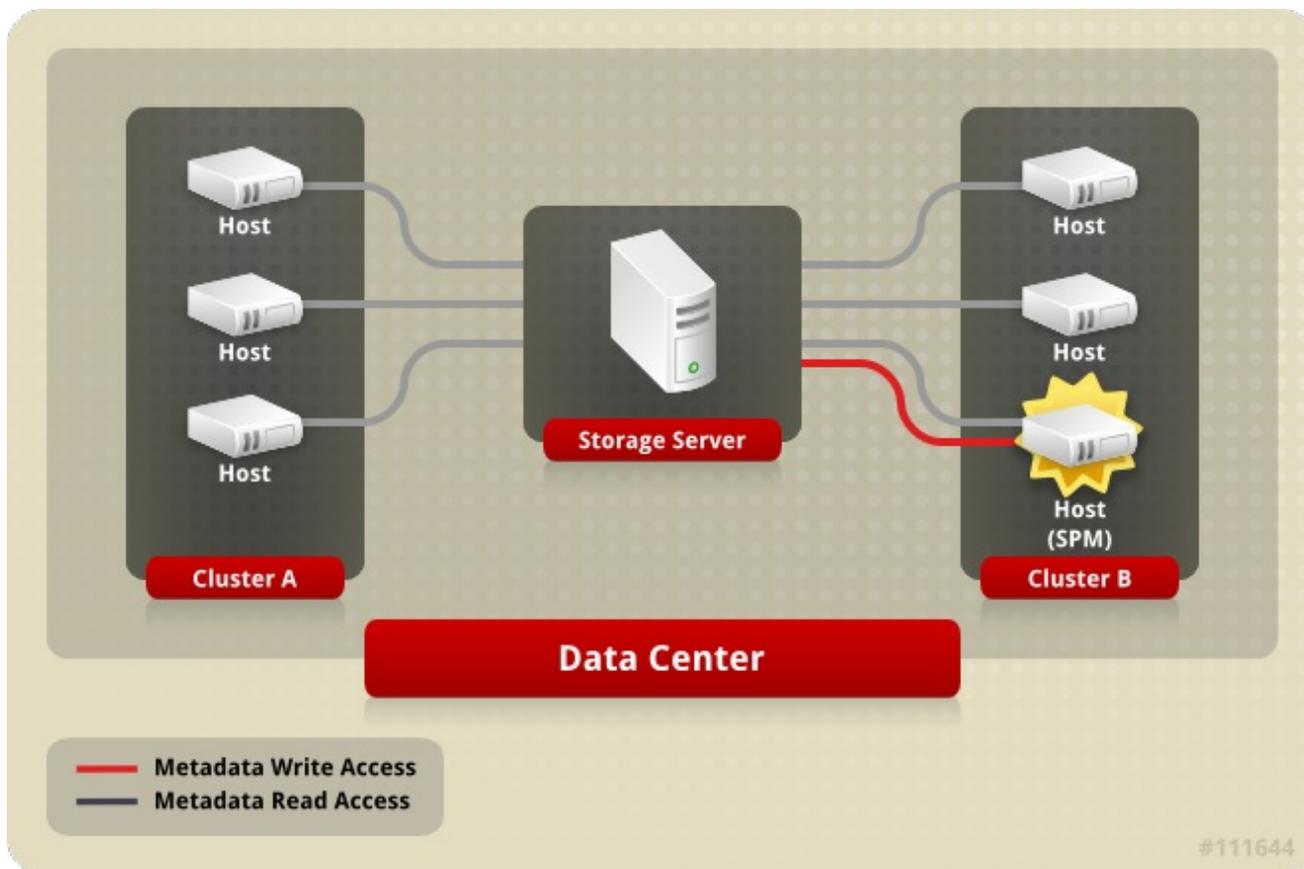


図2.1 Storage Pool Manager による構造メタデータの排他的書き込み

2.9. STORAGE POOL MANAGER の選択プロセス

Storage Pool Manager (SPM) ロールがホストに手動で割り当てられていない場合には、Red Hat Virtualization Manager が SPM 選択プロセスを開始して管理します。

Red Hat Virtualization Manager は最初にストレージセントリックリースを持つホストを確認するよう VDSM に要求します。

Red Hat Virtualization Manager はストレージドメインの初回作成以降の SPM 割り当て履歴をトラッキングします。SPM ロールの稼働状況は以下の 3 つの方法で確認します。

- 「getSPMstatus」コマンド: Manager が VDSM を使用して、SPM のステータスが最後に割り当てられたホストをチェックすると、「SPM」、「Contending」、「Free」のいずれかの値が返されます。
- ストレージドメインのメタデータボリュームには、SPM ステータスを最後に割り当てられたホストが記載されています。
- ストレージドメインのメタデータボリュームには、SPM ステータスが最後に割り当てられたホストのバージョンの情報が含まれています。

稼働中かつ応答可能なホストがストレージセントリックリースを維持している場合には、Red Hat Virtualization Manager は管理ポータルでそのホストを SPM として表示してそれ以上は何も行いません。

SPM ホストが応答しない場合は、そのホストは到達不可とみなされます。ホストに電源管理が設定されている場合は、ホストが自動的にフェンスされます。電源管理が設定されていない場合には、手動でフェンスする必要があります。SPM のロールは、前の SPM がフェンスされるまで、新しいホストに割り当てることはできません。

SPM のロールとストレージセントリックリースが使用可能な場合に、Red Hat Virtualization Manager はデータセンター内で無作為に選択された稼働中のホストに割り当てます。

新規ホストへの SPM ロールの割り当てに失敗した場合には、Red Hat Virtualization Manager は、操作に失敗したホストが含まれる一覧にそのホストを追加し、これらのホストから SPM ロールの資格をなくします。このリストは、次に SPM を選択するプロセスのはじめに消去されて、もう 1 度すべてのホストに SPM ロールを割り当てる資格を与えます。

Red Hat Virtualization Manager は、SPM 選択が成功するまで、操作に失敗したホストの一覧に含まれていない、無作為に選択されたホストが SPM およびストレージセントリックリースを引き継ぐように要求を継続します。

現行の SPM が応答なしの状態となったり、SPM の責任を遂行できなくなった場合には毎回、Red Hat Virtualization Manager が SPM の選択プロセスを開始します。

2.10. RED HAT VIRTUALIZATION の排他的なリソースおよび SANLOCK

Red Hat Virtualization 環境の特定のリソースには排他的にアクセスする必要があります。

SPM ロールは、そのようなリソースの 1 つです。複数のホストが SPM になると、同じデータが 2 つの場所から同時に変更される可能性があるため、データが破損する危険性があります。

Red Hat Enterprise Virtualization 3.1 より以前のバージョンでは、**safelease** という VDSM 機能を使用して SPM が排他的に維持および追跡されていました。このリースは、データセンター内の全ストレージドメインにある特別な領域に書き込まれ、環境内のすべてのホストが、ネットワークに依存しない方法で SPM ステータスを追跡することが可能でした。VDSM のセーフリースでは、1 つのリソース (SPM ロール) の排他性のみが維持されました。

Sanlock は同じ機能を提供しますが、SPM ロールを、ロックできるリソースの 1 つとして扱います。追加リソースをロックできるため、Sanlock にはより大きな柔軟性があります。

リソースのロックが必要なアプリケーションは、Sanlock で登録できます。登録されたアプリケーションは、Sanlock がアプリケーションの代わりにリソースをロックするように要求して、他のアプリケーションがそのリソースにアクセスできないようにすることが可能です。たとえば、VDSM は SPM ステータスをロックする代わりに、Sanlock にその処理を行うように要求します。

ロックは、**lockspace** のディスクで追跡されます。各ストレージドメインには 1 つの lockspace があります。SPM リソースのロックの場合、各ホストがライブ状態であるかどうか (ライブネス) は、ホストがストレージに接続したときに Manager から受け取った `hostid` を更新できるかどうか、ホストがタイムスタンプを定期的に lockspace に書き込むことができるかどうかにより、lockspace で追跡されます。**ids** 論理ボリュームは、各ホストの一意 ID を追跡し、ホストが `hostid` を更新するたびに更新されます。SPM リソースは、ライブ状態のホストのみが保持できます。

リソースは、リース論理ボリュームのディスクで追跡されます。リソースは、リソースを取得したプロセスの一意な ID でディスク上のその表現が更新された場合に、**取得された**と見なされます。SPM ロールの場合、SPM リソースは SPM リソースを取得した `hostid` で更新されます。

各ホストの Sanlock プロセスは、リソースが取得されたことを確認するためにリソースを一度だけチェックする必要があります。最初のチェック後に、Sanlock は、ロックされたリソースがあるホストのタイムスタンプが古くなるまで lockspace を監視できます。

Sanlock は、リソースを使用するアプリケーションを監視します。たとえば、VDSM では SPM ステータスと `hostid` が監視されます。ホストが Manager から `hostid` を更新できない場合は、`hostid` が lockspace のすべてのリソースで排他的に失われます。Sanlock は、リソースが取得されていないことを示すようリソースを更新します。

SPM ホストが、一定の期間に、ストレージドメインの lockspace にタイムスタンプを書き込むことができない場合、Sanlock のホストインスタンスは VDSM プロセスがそのリソースを解放するよう要求します。VDSM プロセスが応答する場合、そのリリースは解放され、別のホストが lockspace の SPM リソースを取得できます。

SPM ホスト上の VDSM がリソースを解放する要求に応答しない場合、ホスト上の Sanlock は VDSM プロセスを終了します。kill コマンドが失敗した場合、Sanlock は sigkill を使用して VDSM を終了しようとします。sigkill が失敗した場合、Sanlock は **watchdog** デーモンを使用してホストを再起動します。

ホスト上の VDSM が hostid を更新し、タイムスタンプを lockspace に書き込むたびに、watchdog デーモンは **pet** を受け取ります。VDSM がそのような処理を行えない場合、watchdog デーモンは pet を受け取らなくなります。watchdog デーモンが一定時間の間 pet を受け取らないと、ホストが再起動されます。この最終的な手段によって、確実に SPM リソースが解放され、別のホストが SPM リソースを取得できるようになります。

2.11. シンプロビジョニングとストレージのオーバーコミット

Red Hat Virtualization Manager は、仮想環境内でのストレージ使用を最適化するプロビジョニングポリシーを提供します。シンプロビジョニングポリシーにより、ストレージリソースのオーバーコミットや、仮想環境の実際のストレージ使用状況に応じてストレージのプロビジョニングを行うことができます。

ストレージのオーバーコミットとは、ストレージプール内で物理的に利用可能な容量を超えるストレージを仮想マシンに割り当てることです。通常、仮想マシンが使用するストレージは、割り当てられた容量を下回ります。シンプロビジョニングにより、仮想マシンは、定義されたストレージが完全に割り当てられているかのように稼働することができますが、実際に割り当てられているストレージ容量はごくわずかです。



注記

Red Hat Virtualization Manager は独自のシンプロビジョニング機能を提供していますが、ご使用のストレージバックエンドでシンプロビジョニング機能が提供されている場合には、その機能を使用することをお勧めします。

ストレージのオーバーコミットをサポートするには、論理ストレージの割り当てを実際のストレージ使用量と比較する閾値を VDSM で定義します。この閾値を使用して、ディスクイメージに書き込まれるデータが、そのディスクイメージをバックアップする論理ボリュームの容量を下回るようにします。QEMU は論理ボリューム内で書き込まれた最も高いオフセットを特定します。これは、ストレージの最大使用ポイントを指します。VDSM は QEMU がマークした最も高いオフセットをモニタリングして、使用量が所定の閾値を超過しないようにします。最高オフセットが閾値を下回っていると VDSM が示し続ける限り、Red Hat Virtualization Manager は、対象の論理ボリュームに稼働を続けるのに十分なストレージ容量が残っているものと認識します。

使用量が閾値を超えつつあることを QEMU が示すと、VDSM はディスクイメージが間もなく論理ボリュームのサイズに達することを Manager に伝達します。Red Hat Virtualization Manager は SPM ホストに論理ボリュームを拡張するように要求します。このプロセスは、そのデータセンターのデータストレージドメインに空き容量が残っている限り繰り返すことが可能です。データストレージドメインの空き容量がなくなった場合には、手動でストレージ容量を拡張する必要があります。

2.12. 論理ボリュームの拡張

Red Hat Virtualization Manager はシンプロビジョニングを使用して、ストレージプール内の使用可能なストレージをオーバーコミットし、物理的に使用可能な容量を超えるストレージを割り当てます。仮想

マシンは、動作に応じてデータを書き込みます。シンプロビジョニングされたディスクイメージを使用する仮想マシンによって書き込まれるデータは、いずれはディスクイメージをバックアップする論理ボリュームが格納できる容量を超えてしまいます。その際には、論理ボリュームが拡張されて追加のストレージが提供され、その仮想マシンは稼働状態を継続することができます。

Red Hat Virtualization では、LVM でシンプロビジョニングのメカニズムを提供します。QCOW2 フォーマットのストレージを使用する場合には、Red Hat Virtualization はホストシステムのプロセス `qemu-kvm` に依存して、逐次的方法でディスク上のストレージブロックを論理ブロックにマッピングします。これにより、たとえば 1 GB の論理ボリュームで 100 GB の論理ディスクの定義が可能になります。qemu-kvm が VDSM によって設定された使用量の閾値を超えると、ローカルの VDSM インスタンスは論理ボリュームを 1 ギガバイト拡張するように SPM に要求します。ボリューム拡張を必要とする仮想マシンを実行しているホスト上の VDSM は、追加領域が必要なことを SPM の VDSM に通知します。SPM が論理ボリュームを拡張し、SPM の VDSM インスタンスは、ホストの VDSM を使用してボリュームグループ情報をリフレッシュし、拡張操作が完了したことを認識します。これでホストは稼働を継続することができます。

論理ボリュームを拡張する際には、対象のホストは、他のどのホストが SPM であるかを知る必要はありません。そのホスト自体が SPM である可能性もあります。ストレージ拡張の伝達は、データストレージドメイン内の専用論理ボリュームであるストレージメールボックスを介して行われます。SPM による論理ボリューム拡張を必要とするホストは、ストレージメールボックス内にあるそのホスト用の所定箇所にメッセージを書き込みます。SPM は定期的に受信メールを読み、要求された論理ボリューム拡張を実行して、返信内容を送信メールに書き込みます。要求を送信後、ホストは 2 秒ごとに受信メールをチェックして応答を確認します。論理ボリューム拡張要求が成功したという返信を受け取ると、ホストはデバイスマッパー内の論理ボリュームマップをリフレッシュし、新たに割り当てられたストレージを認識します。

ストレージプールに提供されている物理ストレージがほぼ使い果たされると、複数のイメージで使用可能なストレージが不足してしまい、リソースの補充を行う手段がなくなります。ストレージプールがストレージを使い果たすと、QEMU は デバイスに使用可能なストレージが残っていないことを示す **enospc error** を返します。この時点で、実行中の仮想マシンは自動的に一時停止状態となるので、管理者が介入して、新規 LUN をボリュームグループに追加する必要があります。

ボリュームグループに新しい LUN が追加されると、SPM は追加ストレージを必要としている論理ボリュームにその LUN を自動的に分配します。追加リソースの自動割り当てにより、関連する仮想マシンは中断せずに稼働を継続し、また停止状態であった場合は稼働が自動的に再開されます。

第3章 ネットワーク

3.1. ネットワークアーキテクチャー

Red Hat Virtualization のネットワークについては、ネットワークの基礎用語、クラスター内のネットワーク、ホストのネットワーク設定に分けて説明します。ネットワークの基礎用語のセクションでは、ネットワークに使用する基本的なハードウェアおよびソフトウェアについて解説します。クラスター内のネットワークのセクションでは、ホスト、論理ネットワーク、仮想マシンなどのクラスターレベルのオブジェクト間でのネットワークの対話について記載しています。ホストのネットワーク設定のセクションでは、ホスト内のネットワークにサポートされている設定について説明します。

ネットワークを適切に設計/構築することにより、高帯域幅のタスクに適切な帯域幅が提供され、ユーザーインタラクションの機能を損なうような遅延は発生せず、仮想マシンを移行ドメイン内で問題なく移行できるようになります。ネットワークが適切に構築されていない場合には、許容できないような遅延が発生することや、ネットワークフラディングにより移行やクローン作成が失敗する可能性があります。

3.2. ネットワークの基礎用語

Red Hat Virtualization は、以下の機能を活用して仮想マシン、仮想ホスト、およびより広範なネットワーク間のネットワーク機能を提供します。

- ネットワークインターフェースコントローラー (NIC)
- ブリッジ
- ボンディング
- 仮想 NIC (VNIC)
- 仮想 LAN (VLAN)

NIC、ブリッジ、仮想 NIC を使用することにより、ホスト、仮想マシン、ローカルエリアネットワーク、インターネットの間のネットワーク通信が可能となります。ボンディングおよび VLAN をオプションで実装すると、セキュリティ、耐障害性、ネットワークキャパシティが強化されます。

3.3. ネットワークインターフェースコントローラー

NIC (ネットワークインターフェースコントローラー) とは、コンピューターをコンピューターネットワークに接続するネットワークアダプターまたは LAN アダプターのことです。NIC はマシンの物理層およびデータリンク層の両方で稼働し、ネットワーク接続を可能にします。Red Hat Virtualization 環境内の全仮想ホストは、NIC を少なくとも 1 枚搭載していますが、2 枚以上の NIC を搭載している方がより一般的です。

1 枚の物理 NIC には、複数の仮想 NIC (VNIC) を論理的に接続することが可能です。仮想 NIC は、仮想マシンの物理ネットワークインターフェースとして機能します。仮想 NIC とそれをサポートする NIC とを区別するため、Red Hat Virtualization Manager は各仮想 NIC に一意の MAC アドレスを割り当てます。

3.4. ブリッジ

ブリッジとは、パケット交換ネットワークでパケット転送を使用するソフトウェアデバイスです。ブリッジングにより、1 枚の NIC の接続を複数のネットワークインターフェースデバイスが共有し、ネットワーク上で個別の物理デバイスのように表示することができます。ブリッジは、パケットのソースア

ドレスを確認して、適切なターゲットアドレスを決定します。ターゲットアドレスが決定されると、ブリッジはその場所を後で参照できるようにテーブルに追加します。これによりホストは、ブリッジのメンバーである、仮想マシンに関連付けされた仮想 NIC にネットワークトラフィックをリダイレクトすることが可能となります。

Red Hat Virtualization では、論理ネットワークはブリッジを使用して実装されます。IP アドレスを受け取るホスト上の物理的なインターフェースではなく、ブリッジを使用します。このブリッジに関連付けられる IP アドレスは、接続のためにブリッジを使用する仮想マシンと同じサブネット内になくても構いません。ブリッジを使用する仮想マシンと同じサブネット上にある IP アドレスをそのブリッジに割り当てると、ホストは論理ネットワーク内で仮想マシンによるアドレス指定が可能になります。原則として、Red Hat Virtualization ホストでは、ネットワークに公開されるサービスを実行することを推奨しません。ゲストはゲストの仮想 NIC を使って論理ネットワークに接続され、ホストはホストの NIC を使って論理ネットワークのリモート要素に接続されます。各ゲストには、DHCP を使用して、または静的に、仮想 NIC の IP アドレスを個別に設定することができます。ブリッジは、ホストの外部にあるオブジェクトに接続することはできますが、そのような接続は必須ではありません。

カスタムプロパティーは、ブリッジおよびイーサネット接続の両方に定義することができます。VDSM はネットワークの定義とカスタムプロパティーを設定ネットワークフックスクリプトに渡します。

3.5. ボンディング

ボンディングとは、複数のネットワークインターフェースをソフトウェアで定義したデバイス 1 つに集約することです。ボンディングされたネットワークインターフェースは、ボンディングに含まれているネットワークインターフェースカード (NIC) の伝送機能を統合して、1 つのネットワークインターフェースとして機能するため、単一の NIC よりも伝送速度が早くなります。また、ボンディング内の NIC すべてに障害が発生しない限り、ボンディング自体には障害が発生しないため、ボンディングすることでフォールトトレランスが向上します。ただし、1 点制約があり、ボンディング内のすべてのネットワークインターフェースカードが同じオプションやモードをサポートするように、ネットワークインターフェースをボンディングする NIC は、必ず同じメーカーおよびモデルでなければなりません。

ボンディングの packets 分散アルゴリズムは、使用するボンディングモードによって決定されます。



重要

モード 1、2、3、4 は、仮想マシン (ブリッジ) および物理マシン (ブリッジなし) のネットワークタイプをサポートします。モード 0、5、6 は、物理マシン (ブリッジなし) のネットワークのみをサポートします。

ボンディングモード

Red Hat Virtualization は、デフォルトでモード 4 を使用しますが、以下にあげる一般的なボンディングモードに対応しています。

モード 0 (*round-robin* ポリシー)

このモードは、ネットワークインターフェースカードを順番に使用してパケットを送信します。パケットの送信は、ボンディングで最初に利用可能なネットワークインターフェースカードから、最後に利用可能なネットワークインターフェースカードまでループで使用をくり返します。それ以降のループでもすべて、最初に利用可能なネットワークインターフェースカードから使用されます。モード 0 では、ネットワークに対して耐障害性や負荷分散が提供されていますが、ブリッジと併用できないため、仮想マシンの論理ネットワークとの互換性はありません。

モード 1 (*active-backup* ポリシー)

このモードは、すべてのネットワークインターフェースカードをバックアップ状態に設定して、1 つだけアクティブなカードを残します。アクティブなネットワークインターフェースカードで障害が発生すると、バックアップに設定されていたネットワークインターフェースカードの 1 つが、障害

の発生したインターフェースに代わって、ボンディング内で唯一のアクティブインターフェースになります。1つ以上のポートでアドレスが表示されていると、有効なネットワークインターフェースカードの MAC アドレスを反映するためにボンディングの MAC アドレスが変更された場合に混乱が生じる可能性があり、このような混乱を避ける目的で、モード 1 のボンディングの MAC アドレスは、1つのポートだけで表示されます。モード 1 は耐障害性を提供し、Red Hat Virtualization でサポートされています。

モード 2 (XOR ポリシー)

このモードは、送信元と送信先の MAC アドレスの XOR (排他的理論和) をネットワークインターフェースカードのスレーブ数で除算した剰余に基づいて、パッケージの送信先のネットワークインターフェースカードを選択します。この計算により、各送信先の MAC アドレスに必ず同じネットワークインターフェースカードが選択されるようにします。モード 2 は耐障害性と負荷分散を提供し、Red Hat Virtualization でサポートされています。

モード 3 (broadcast ポリシー)

このモードは、全パッケージを全ネットワークインターフェースカードに送信します。モード 3 は耐障害性を提供し、Red Hat Virtualization でサポートされています。

モード 4 (IEEE 802.3ad ポリシー)

このモードは、任意の集約グループを作成し、このグループ内のインターフェースが速度およびデュプレックスの設定を共有します。モード 4 は、IEEE 802.3ad 仕様に従ってアクティブな集約グループ内のネットワークインターフェースカードをすべて使用します。このモードも、Red Hat Virtualization でサポートされています。

モード 5 (adaptive transmit load balancing ポリシー)

このモードは、ボンディング内の各ネットワークインターフェースカードの負荷に応じて発信トラフィックが分散され、現在のネットワークインターフェースカードが全着信トラフィックを受信するようにします。トラフィックの受信に割り当てられているネットワークインターフェースカードに障害が発生した場合には、着信トラフィックの受信ロールは別のネットワークインターフェースカードに割り当てられます。モード 5 はブリッジと併用できないため、仮想マシンの論理ネットワークとの互換性はありません。

モード 6 (adaptive load balancing ポリシー)

このモードは、モード 5 (adaptive transmit load balancing ポリシー) に IPv4 トラフィックの受信負荷分散を組み合わせたポリシーで、特別なスイッチ要件はありません。ARP ネゴシエーションを使用して受信負荷の分散を行います。モード 6 はブリッジと併用できないため、仮想マシンの論理ネットワークとの互換性はありません。

3.6. ボンディング用のスイッチ設定

スイッチ設定は、ハードウェアの要件により異なります。お使いのオペレーティングシステムにあったデプロイメントおよびネットワーク設定ガイドを参照してください。



重要

いずれのスイッチタイプの場合も、Cisco *Port Aggregation Protocol* (PAgP) プロトコルではなく、*Link Aggregation Control Protocol* (LACP) プロトコルでスイッチボンディングを設定することが重要です。

3.7. 仮想ネットワークインターフェースカード

仮想ネットワークインターフェースカードは、ホストの物理ネットワークインターフェースカードをベースとした仮想ネットワークインターフェースです。各ホストには、複数のネットワークインターフェースカードが搭載されていますが、各ネットワークインターフェースカードを、複数の仮想ネットワークインターフェースカードのベースとして設定することができます。

仮想ネットワークインターフェースカードを仮想マシンにアタッチすると、Red Hat Virtualization Manager により、仮想ネットワークインターフェースカードのアタッチ先の仮想マシン、仮想ネットワークインターフェースカード自体、仮想ネットワークインターフェースカードのベースとなる物理ホストのネットワークインターフェースカードの間で複数の関連付けが作成されます。具体的には、仮想ネットワークインターフェースカードが仮想マシンにアタッチされると、仮想ネットワークインターフェースカードがベースとする物理ホストのネットワークインターフェースカード上で、新しい仮想ネットワークインターフェースカードと MAC アドレスが作成されます。そして、仮想ネットワークインターフェースカードのアタッチ後、初めて仮想マシンを起動すると、**libvirt** により、仮想ネットワークインターフェースカードに PCI アドレスが割り当てられます。次に、この MAC アドレスと PCI アドレスを使用して、仮想マシンの仮想ネットワークインターフェースカードの名前 (例: **eth0**) が取得されます。

テンプレートやスナップショットをベースに仮想マシンを作成する場合は、MAC アドレスを割り当てるプロセス、およびこれらの MAC アドレスと PCI アドレスを関連付けるプロセスが若干異なります。テンプレートやスナップショット用に PCI アドレスがすでに作成されている場合は、そのテンプレートやスナップショットをベースに作成した仮想マシン上の仮想ネットワークインターフェースカードは PCI アドレスの順に整理され、MAC アドレスがこの順に割り当てられます。一方、テンプレート用に PCI アドレスが作成されていない場合は、そのテンプレートをベースに作成した仮想マシン上の仮想ネットワークインターフェースカードには、その名前順に MAC アドレスが割り当てられます。スナップショット用に PCI アドレスが作成されていない場合は、そのスナップショットをベースに作成した仮想マシン上の仮想ネットワークインターフェースカードには、Red Hat Virtualization Manager が新しい MAC アドレスを割り当てます。

作成が済むと、ネットワークインターフェースカードがネットワークブリッジデバイスに追加されます。ネットワークブリッジデバイスは、仮想マシンを仮想マシン論理ネットワークに接続する手段です。

仮想化ホスト上で **ip addr show** を実行すると、そのホスト上の仮想マシンに関連付けられた仮想ネットワークインターフェースカードすべてが表示されます。また、論理ネットワークを強化するために作成されたネットワークブリッジや、ホストで使用されるネットワークインターフェースカードなどが表示されます。

```
[root@rhev-host-01 ~]# ip addr show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 16436 qdisc noqueue state UNKNOWN
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP qlen 1000
    link/ether 00:21:86:a2:85:cd brd ff:ff:ff:ff:ff:ff
    inet6 fe80::221:86ff:fea2:85cd/64 scope link
        valid_lft forever preferred_lft forever
3: wlan0: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc mq state DOWN qlen 1000
    link/ether 00:21:6b:cc:14:6c brd ff:ff:ff:ff:ff:ff
5: ;vdsmdummy;; <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN
    link/ether 4a:d5:52:c2:7f:4b brd ff:ff:ff:ff:ff:ff
6: bond0: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
    link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
7: bond4: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
```

```

link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
8: bond1: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
9: bond2: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
10: bond3: <BROADCAST,MULTICAST,MASTER> mtu 1500 qdisc noop state DOWN
link/ether 00:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
11: ovirtmgmt: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue
state UNKNOWN
link/ether 00:21:86:a2:85:cd brd ff:ff:ff:ff:ff:ff
inet 10.64.32.134/23 brd 10.64.33.255 scope global ovirtmgmt
inet6 fe80::221:86ff:fea2:85cd/64 scope link
valid_lft forever preferred_lft forever

```

このコマンドでは、複数のデバイス (ループバックデバイス (**lo**) 1 つ、Ethernet デバイス 1 つ (**eth0**)、ワイヤレスデバイス 1 つ (**wlan0**)、VDSM ダミーデバイス 1 つ (**;vdsmdummy;**)、ボンディングデバイス 5 つ (**bond0**、**bond4**、**bond1**、**bond2**、**bond3**)、ネットワークブリッジ (**ovirtmgmt**) 1 つ) がコンソールに出力されます。

仮想ネットワークインターフェースカードはすべて、論理ネットワークのネットワークブリッジデバイスのメンバーです。ブリッジのメンバーシップは **brctl show** コマンドで表示することができます。

```

[root@rhev-host-01 ~]# brctl show
bridge name bridge id STP enabled interfaces
ovirtmgmt 8000.e41f13b7fdd4 no vnet002
vnet001
vnet000
eth0

```

brctl show コマンドのコンソール出力から、**virtio** 仮想ネットワークインターフェースカードが **ovirtmgmt** ブリッジのメンバーであることがわかります。仮想ネットワークインターフェースカードが関連付けられている仮想マシンはすべて、**ovirtmgmt** 論理ネットワークに接続されています。**eth0** のネットワークインターフェースカードは **ovirtmgmt** ブリッジのメンバーでもあります。**eth0** デバイスは、ホスト外部への接続を提供するスイッチに接続されています。

3.8. 仮想 LAN (VLAN)

VLAN (仮想 LAN) とは、ネットワークパケットに適用できる属性です。ネットワークパケットは、番号が付いた VLAN にタグ付けすることができます。VLAN は、スイッチレベルでネットワークトラフィックを完全に分離するのに使用するセキュリティー機能です。VLAN は完全に切り離されており、相互排他的です。Red Hat Virtualization Manager は VLAN に対応しており、VLAN トラフィックをタグ付けしてリダイレクトすることができます。ただし、VLAN の実装には VLAN をサポートするスイッチが必要となります。

スイッチレベルで、ポートに VLAN 指定が割り当てられます。スイッチは、特定のポートを起点とするトラフィックに VLAN タグを付けてそのトラフィックを VLAN の一部としてマークし、応答にも同じ VLAN タグが付けられるようにします。VLAN は複数のスイッチ全体に拡張することができます。スイッチ上の VLAN タグ付きネットワークトラフィックは、適正な VLAN に指定されたポートに接続されたマシン以外には完全に検出不可能です。任意のポートに複数の VLAN タグを付けることが可能です。これにより、複数の VLAN からのトラフィックを単一のポートに送信し、トラフィックを受信するマシン上でソフトウェアを使用して解読することができるようになります。

3.9. ネットワークラベル

ネットワークラベルを使用すると、論理ネットワークの作成/管理や、物理ホストネットワークインターフェース/ボンディングへの関連付けに伴う複数の管理タスクを大幅に簡素化することができます。

ネットワークラベルは、プレーンテキスト形式の人間が判読可能なラベルで、論理ネットワークまたは物理ホストネットワークインターフェースにアタッチすることができます。ラベルの長さには制限はありませんが、アルファベットの大文字/小文字、アンダースコア、およびハイフンを組み合わせる必要があります。スペースや特殊文字を使用することはできません。

論理ネットワークまたは物理ホストネットワークインターフェースにラベルをアタッチすると、以下のように、同じラベルがアタッチされた他の論理ネットワークや物理ホストネットワークインターフェースと関連付けされます。

ネットワークラベルの関連付け

- ラベルを論理ネットワークにアタッチすると、その論理ネットワークは、その特定のラベルが付いた物理ホストネットワークインターフェースに自動的に関連付けられます。
- ラベルを物理ホストネットワークインターフェースにアタッチすると、その特定のラベルが付いた論理ネットワークは、その物理ホストネットワークインターフェースに自動的に関連付けられます。
- 論理ネットワークまたは物理ホストネットワークインターフェースにアタッチされたラベルを変更すると、ラベルを削除して新規追加したのと同じように機能し、対象の論理ネットワークまたは物理ホストネットワークインターフェースの間の関連付けが更新されます。

ネットワークラベルとクラスター

- ラベル付きの論理ネットワークがクラスターに追加され、同じラベルが付いた物理ホストネットワークインターフェースがそのクラスター内にある場合には、論理ネットワークはその物理ホストネットワークインターフェースに自動的に追加されます。
- ラベル付きの論理ネットワークがクラスターからデタッチされ、同じラベルが付いた物理ホストネットワークインターフェースがそのクラスター内にある場合には、論理ネットワークはその物理ホストネットワークインターフェースから自動的に削除されます。

ネットワークラベルとロール付き論理ネットワーク

- ラベル付き論理ネットワークがディスプレイネットワークまたは移行ネットワークとして機能するよう割り当てられている場合には、その論理ネットワークは、IP アドレスを割り当てることができるように、物理ホストネットワークインターフェース上で DHCP を使用して設定されます。

ロールネットワーク (例: 「移行ネットワーク」や「ディスプレイネットワーク」など) にラベルを設定すると、そのネットワークが全ホストに一括でデプロイされます。このようなネットワークの一括追加は、DHCP を使用して処理されます。この方法による一括デプロイは、静的アドレスを入力する方法よりも優先されます。これは、多数の静的 IP アドレスを入力する作業が性質上スケラブルでないことが理由です。

3.10. クラスターネットワーク

クラスターレベルのネットワークオブジェクトには、以下が含まれます。

- クラスター
- 論理ネットワーク

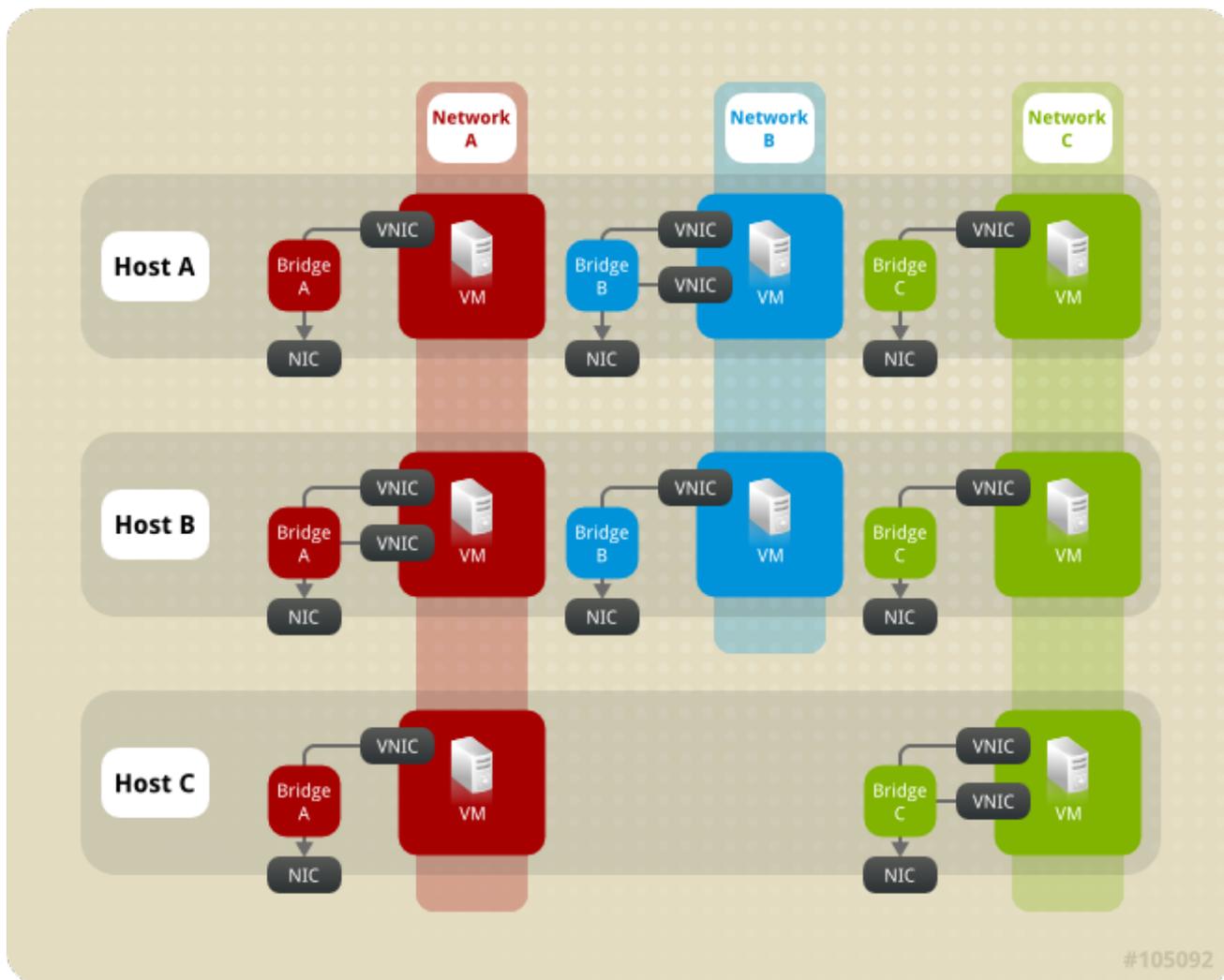


図3.1 クラスタ内のネットワーク

データセンターとは、複数のクラスタの論理グループで、各クラスタは複数のホストの論理グループです。図3.1「クラスタ内のネットワーク」は、単一のクラスタに含まれるリソースを示しています。

クラスタ内のホストはすべて同じストレージドメインにアクセスすることができます。クラスタ内のホストには、クラスタレベルで論理ネットワークが適用されます。仮想マシンの論理ネットワークを稼働させて仮想マシンに使用するには、Red Hat Virtualization Manager を使用してクラスタ内の各ホストでネットワークを定義、実装しておく必要があります。その他のタイプの論理ネットワークは、そのネットワークを使用するホスト上のみを実装することができます。

マルチホストネットワーク設定により、そのネットワークが割り当てられたデータセンター内の全ホストに、更新されたネットワーク設定が自動的に適用されます。

3.11. 論理ネットワーク

論理ネットワークにより、Red Hat Virtualization 環境はネットワークトラフィックをタイプ別に分離することが可能となります。たとえば、`ovirtmgmt` ネットワークは、Manager とホストの間で管理を目的とした通信に使用するために Red Hat Virtualization のインストール中にデフォルトで作成されます。論理ネットワークは、要件が同じようなネットワークトラフィックをグループ化し、まとめて使用するのが一般的な用途です。管理者は多くの場合には、ストレージネットワークとディスプレイネットワークを作成して、各タイプのトラフィックを分離することによって、最適化やトラブルシューティングを行います。

論理ネットワークの種類は以下のとおりです。

- 仮想マシンネットワークトラフィックを伝送する論理ネットワーク
- 仮想マシンネットワークトラフィックを伝送しない論理ネットワーク
- 任意の論理ネットワーク
- 必須ネットワーク

すべての論理ネットワークは、必須または任意のいずれかに指定することができます。

論理ネットワークは、データセンターレベルで定義され、ホストに追加されます。必須論理ネットワークが動作するには、該当するクラスターの各ホストに対して論理ネットワークを実装する必要があります。

Red Hat Virtualization 環境内の各仮想マシン論理ネットワークは、ホストのネットワークブリッジデバイスによってサポートされるので、新しい仮想マシン論理ネットワークをクラスターに対して定義する場合は、論理ネットワークを仮想マシンで使用する前に、クラスター内の各ホストで適切なブリッジデバイスを作成する必要があります。Red Hat Virtualization Manager では、仮想マシン論理ネットワークに対して必要なブリッジが自動的に作成されます。

仮想マシン論理ネットワークデバイスをバックアップするために Red Hat Virtualization Manager によって作成されたブリッジデバイスは、ホストネットワークインターフェースに関連付けられます。ブリッジに含まれるホストネットワークインターフェースがネットワークに接続されている場合には、それ以降にブリッジに追加されるネットワークインターフェースはすべて、そのブリッジのネットワーク接続を共有します。仮想マシンを作成して特定の論理ネットワーク上に配置すると、仮想マシンの仮想ネットワークカードは、その論理ネットワークのブリッジに追加されます。これらの仮想マシンはお互いに通信を行ったり、そのブリッジに接続されている他のオブジェクトと通信を行ったりすることができます。

仮想マシンネットワークトラフィックに使用されない論理ネットワークは、ホストネットワークインターフェースに直接関連付けられません。

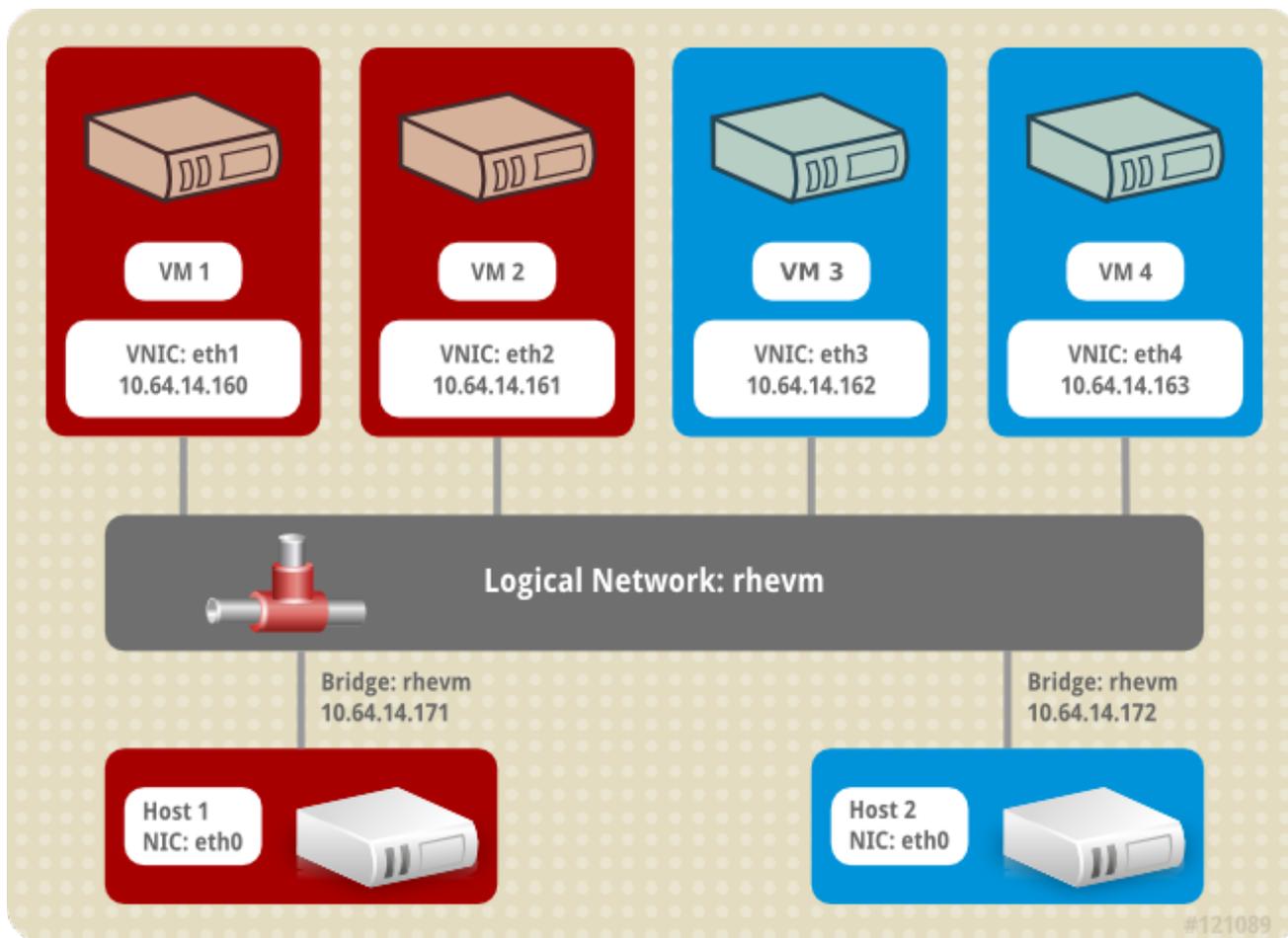


図3.2 ovirtmgmt 論理ネットワーク

例3.1 論理ネットワークの使用例

Purple データセンター内の Pink クラスタに Red と White という 2 つのホストがあります。Red と White はいずれも、デフォルトの論理ネットワーク **ovirtmgmt** をすべてのネットワーク機能に使用していましたが、Pink 担当のシステム管理者は、Web サーバーのテスト用にネットワークを分離するため Web サーバーと一部のクライアント仮想マシンを別の論理ネットワーク (**network_testing**) に配置することにしました。

管理者は、まず Purple データセンターに論理ネットワークを定義し、次にその論理ネットワークを Pink クラスタに適用します。論理ネットワークは、ホストがメンテナンスモードに入っている状態で実装する必要があります。このため、管理者はまず実行中の仮想マシンをすべて Red に移動してから、White をメンテナンスモードに切り替えた上で、ブリッジに追加する物理ネットワークインターフェースに関連付けられている **仮想 NIC** を編集します。選択した物理ネットワークインターフェースの **リンクステータス** が **Down** から **Non-Operational** に変わります。Non-Operational のステータスになるのは、Pink クラスタ内の各ホスト上の物理ネットワークインターフェースを **network_testing** ネットワークに追加して、クラスタ内の全ホストで対象のブリッジを設定する必要があります。次に管理者は、White をアクティブ化して実行中の仮想マシンを Red から移行し、同じプロセスを Red で再度実行します。

物理ネットワークインターフェースにブリッジされた **network_testing** 論理ネットワークが White と Red の両方に適用されると、**network_testing** 論理ネットワークは **Operational** に変わります。これで仮想マシンがこの論理ネットワークを使用する準備が整ったこととなります。

3.12. 必須ネットワーク、任意ネットワーク、仮想マシンネットワーク

必須ネットワークとは、クラスター内の全ホストで利用できる必要のある論理ネットワークのことです。ホストの必須ネットワークが非稼働状態になった場合には、そのホストで実行されている仮想マシンは別のホストに移行されます。移行の範囲は選択したスケジューリングポリシーによって異なります。この機能は、仮想マシンがミッションクリティカルなワークロードを実行している場合に役立ちます。

任意ネットワークとは、**必須** ネットワークと明示的に宣言されていない論理ネットワークのことです。任意ネットワークは、ネットワークを使用するホストのみに実装されます。このネットワークの有無によって、ホストの **Operational** のステータスが変わるわけではありません。任意ネットワークが非稼働状態になった場合には、そのネットワークで実行中の仮想マシンは別のホストに移行されません。これは、大量に仮想マシンを移行することで発生する不必要な I/O の過負荷を防止します。論理ネットワークを作成してクラスターに追加する際に、**必須** ボックスがデフォルトで選択されている点にご注意ください。

ネットワークを **必須** に指定するには、管理ポータルからそのネットワークを選択して、**クラスター タブ** をクリックし、**管理ネットワーク** ボタンをクリックします。

仮想マシンネットワーク (ユーザーインターフェースでは **仮想マシンのネットワーク** と呼ばれる) は、仮想マシンのネットワークトラフィックのみを伝送するよう指定された論理ネットワークです。仮想マシンのネットワークは、必須または任意に指定することができます。任意の仮想マシンネットワークを使用する仮想マシンは、そのネットワークを使用するホストでのみ起動します。

3.13. 仮想マシンの接続性

Red Hat Virtualization では、仮想マシンの作成時に、その仮想マシンの NIC が論理ネットワークに配置されます。その時点から、仮想マシンは同じネットワーク上のその他の任意のターゲットと通信可能となります。

ホストの観点から見ると、仮想マシンが論理ネットワークに配置された時点で、仮想マシンの NIC をバックアップする仮想 NIC がメンバーとしてその論理ネットワークのブリッジデバイスに追加されます。たとえば、仮想マシンが **ovirtmgmt** 論理ネットワーク上にある場合には、その仮想 NIC は仮想マシンを実行しているホストの **ovirtmgmt** ブリッジのメンバーとして追加されます。

3.14. ポートミラーリング

ポートミラーリングは、任意の論理ネットワークおよびホスト上のレイヤー 3 ネットワークトラフィックを仮想マシン上の仮想インターフェースにコピーします。この仮想マシンは、ネットワークのデバッグとチューニング、侵入検出、同一のホストおよび論理ネットワーク上にある仮想マシンの動作のモニタリングに使用することができます。

コピーされるトラフィックは単一のホスト上の単一の論理ネットワークの内部トラフィックのみです。ホストの外部のネットワーク上のトラフィックは増加しませんが、ポートミラーリングを有効化した仮想マシンは他の仮想マシンよりもホストの CPU および RAM の使用率が高くなります。

ポートミラーリングは、論理ネットワークの仮想 NIC で有効化/無効化されます。ただし、以下の制約があります。

- ポートミラーリングが有効になっているプロファイルを持つ仮想 NIC のホットプラグは、サポートされません。
- 仮想 NIC プロファイルが仮想マシンにアタッチされている場合は、ポートミラーリングは変更できません。

上記の制約をもとに、専用の仮想 NIC プロファイルを追加して、そのプロファイルに対してポートミラーリングを有効化するように推奨します。



重要

ポートミラーリングを有効化すると、他のネットワークユーザーのプライバシーレベルが低くなる点に注意してください。

3.15. ホストのネットワーク構成

Red Hat Virtualization ホストの一般的なネットワーク構成タイプには、以下のような構成が含まれます。

- ブリッジと NIC の構成
- ブリッジ、VLAN、NIC の構成
- ブリッジ、ボンディング、VLAN の構成
- 複数のブリッジ、複数の VLAN、NIC の構成

3.16. ブリッジの構成

Red Hat Virtualization で最もシンプルなホスト構成はブリッジと NIC の構成です。図3.3「ブリッジと NIC の構成」に示したように、この構成では、ブリッジを使用して単一または複数の仮想マシン (またはゲスト) をホストの NIC に接続します。

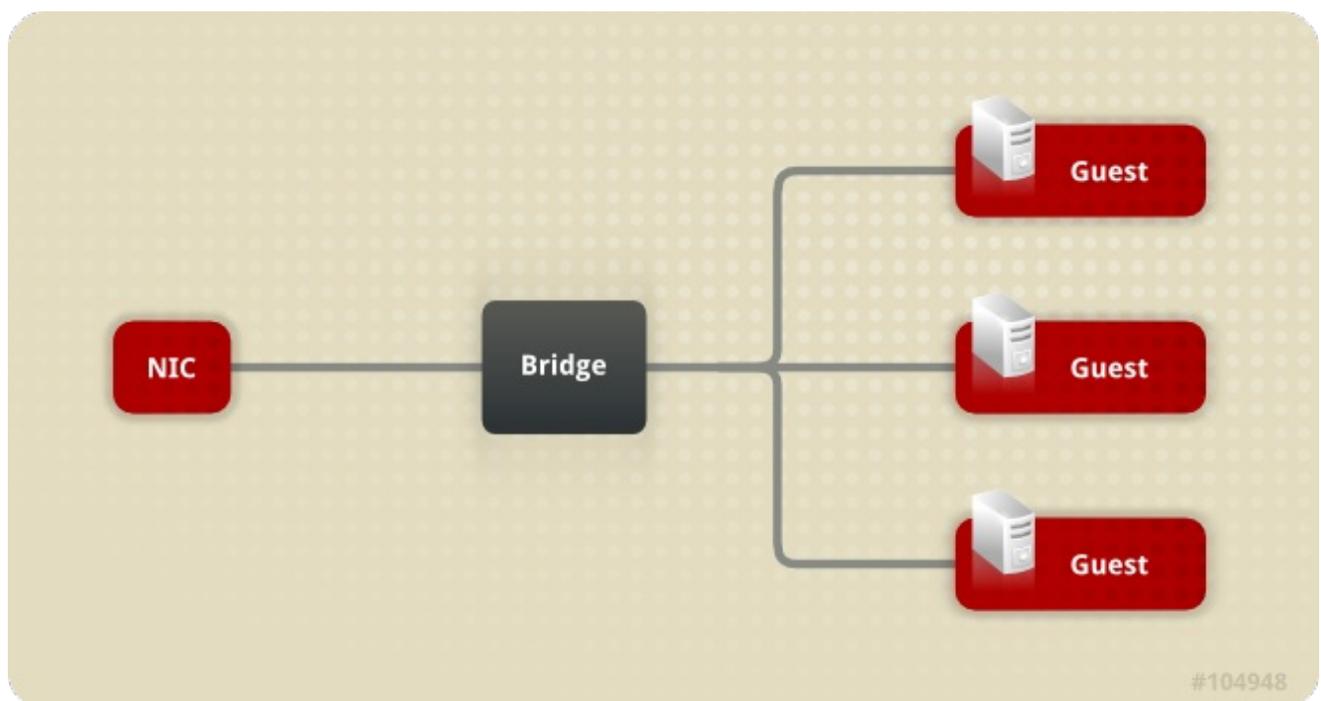


図3.3 ブリッジと NIC の構成

Red Hat Virtualization Manager インストール時に自動作成される **ovirtmgmt** ブリッジは、この構成の一例です。インストール時には、Red Hat Virtualization Manager が **VDSM** をホスト上にインストールします。**VDSM** のインストールプロセスで **ovirtmgmt** ブリッジが作成されます。その後 **ovirtmgmt** ブリッジがホストの **IP** アドレスを取得して、そのホストで管理のための通信が可能となります。

3.17. VLAN の構成

図3.4「ブリッジ、VLAN、NICの構成」は、ホストのNICとブリッジの接続に仮想LAN (VLAN) を追加したもう1つの構成を示しています。

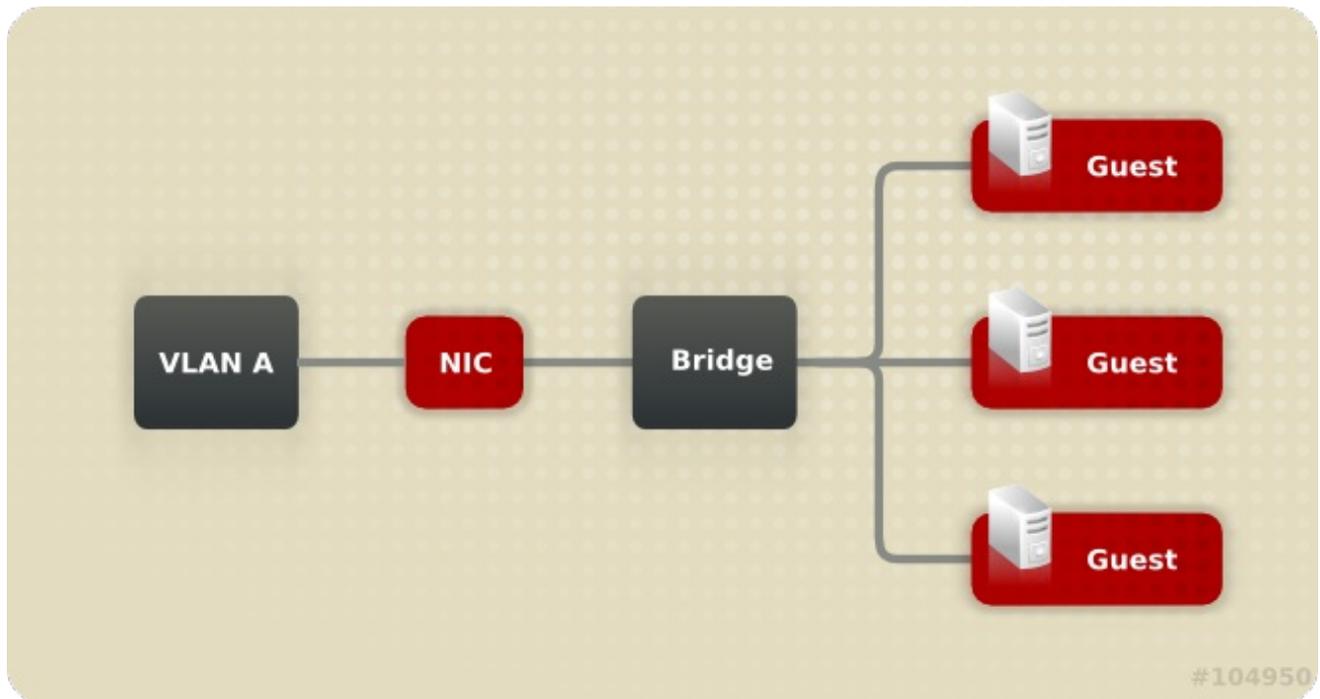


図3.4 ブリッジ、VLAN、NICの構成

VLAN を追加することにより、このネットワーク上でのデータ転送用にセキュアなチャンネルが提供され、複数の VLAN を使用する単一の NIC に複数のブリッジを接続するオプションがサポートされます。

3.18. ブリッジとボンディングの構成

図3.5「ブリッジ、ボンディング、NICの構成」は、ボンディングを追加して複数のホストNICを同じブリッジおよびネットワークに接続する構成を示しています。

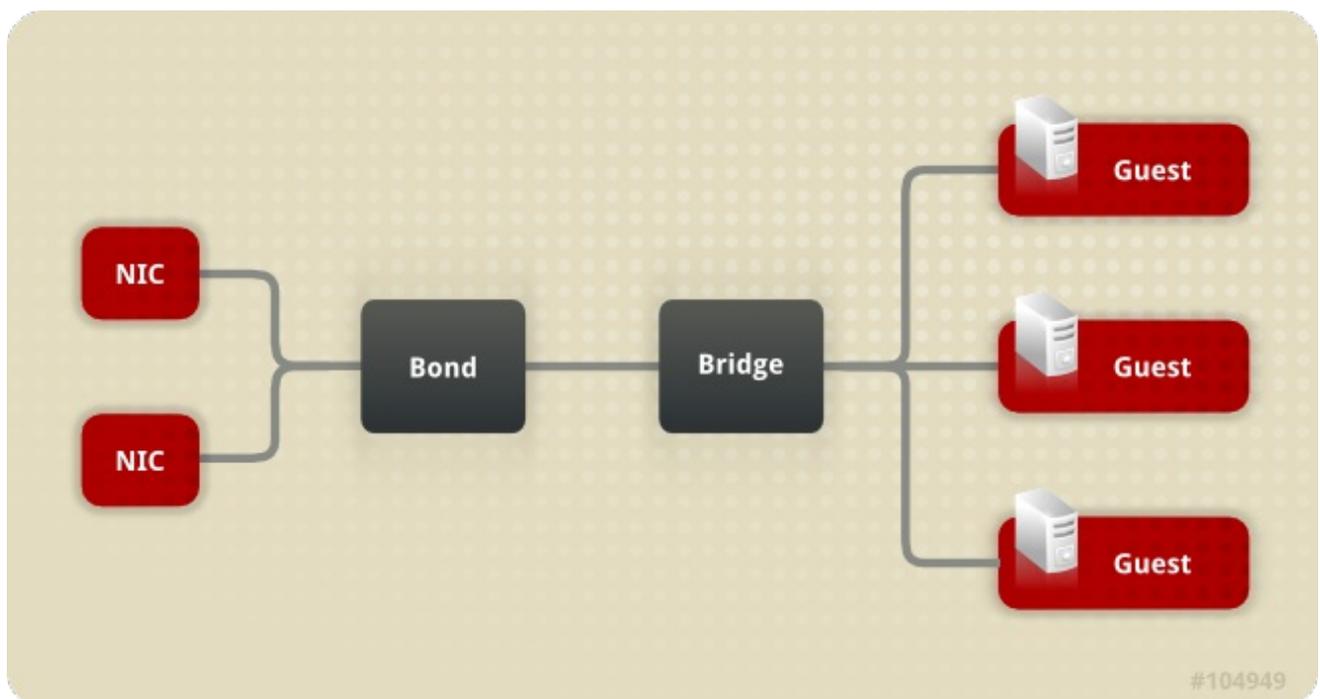


図3.5 ブリッジ、ボンディング、NICの構成

ボンディングを追加することにより、2つ (またはそれ以上) の物理イーサネットリンクを結合した1つの論理リンクが作成されます。これにより、ボンディングモードに応じて NIC の耐障害性向上や潜在的な帯域幅の拡張などのメリットが得られます。

3.19. 複数のブリッジ、複数の VLAN、NIC の構成

図3.6「複数のブリッジ、複数の VLAN、NIC の構成」は、1つの NIC を2つの VLAN に接続する構成例を示しています。この例では、2つの VLAN のいずれかにタグ付けされているネットワークトラフィックをホスト上の NIC 1つに渡すようにネットワークスイッチが設定されていることを前提としています。ホストは2つの仮想 NIC を使用して VLAN 別に VLAN トラフィックを分離します。適切な仮想 NIC をブリッジメンバーとして追加することにより、いずれか一方の VLAN にタグ付けされたトラフィックは、個別のブリッジに接続します。各ブリッジには、複数の仮想マシンが順に接続されます。

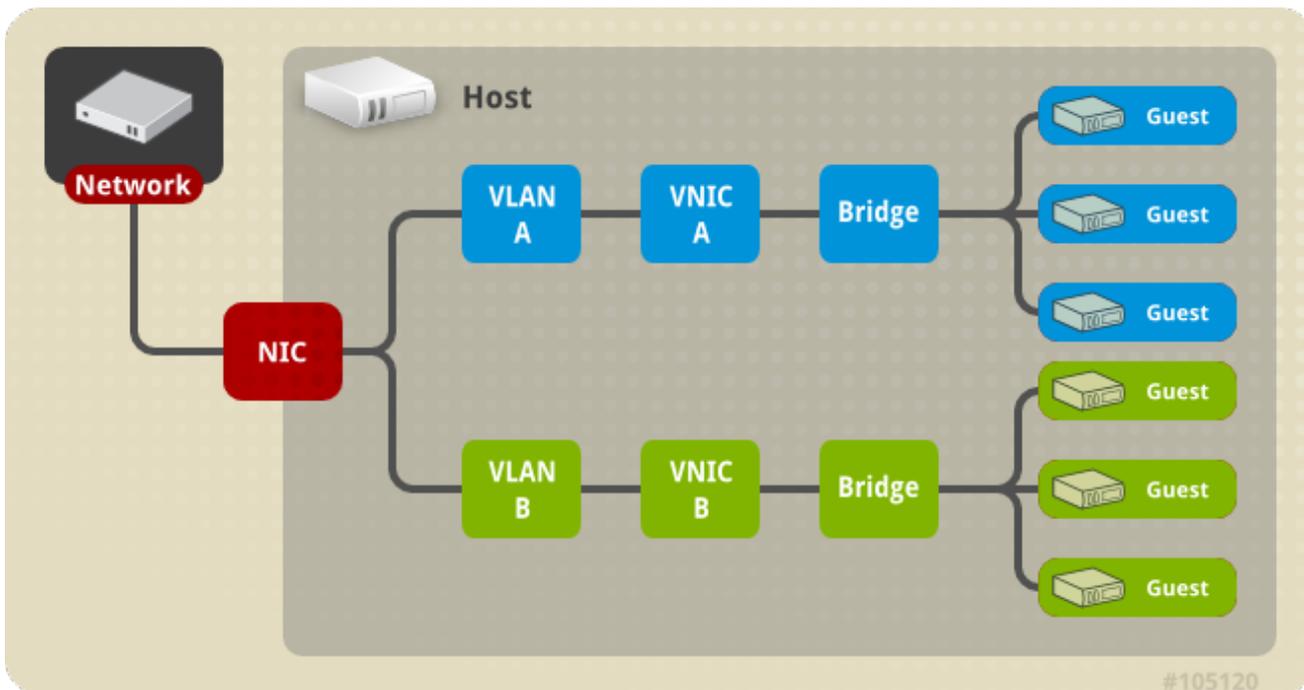


図3.6 複数のブリッジ、複数の VLAN、NIC の構成

3.20. 複数のブリッジ、複数の VLAN、ボンディングの構成

図3.7「ボンディング接続を使用した複数のブリッジ、複数の VLAN、複数の NIC」は、複数の NIC をボンディングして複数の VLAN との接続を円滑化する構成を示します。

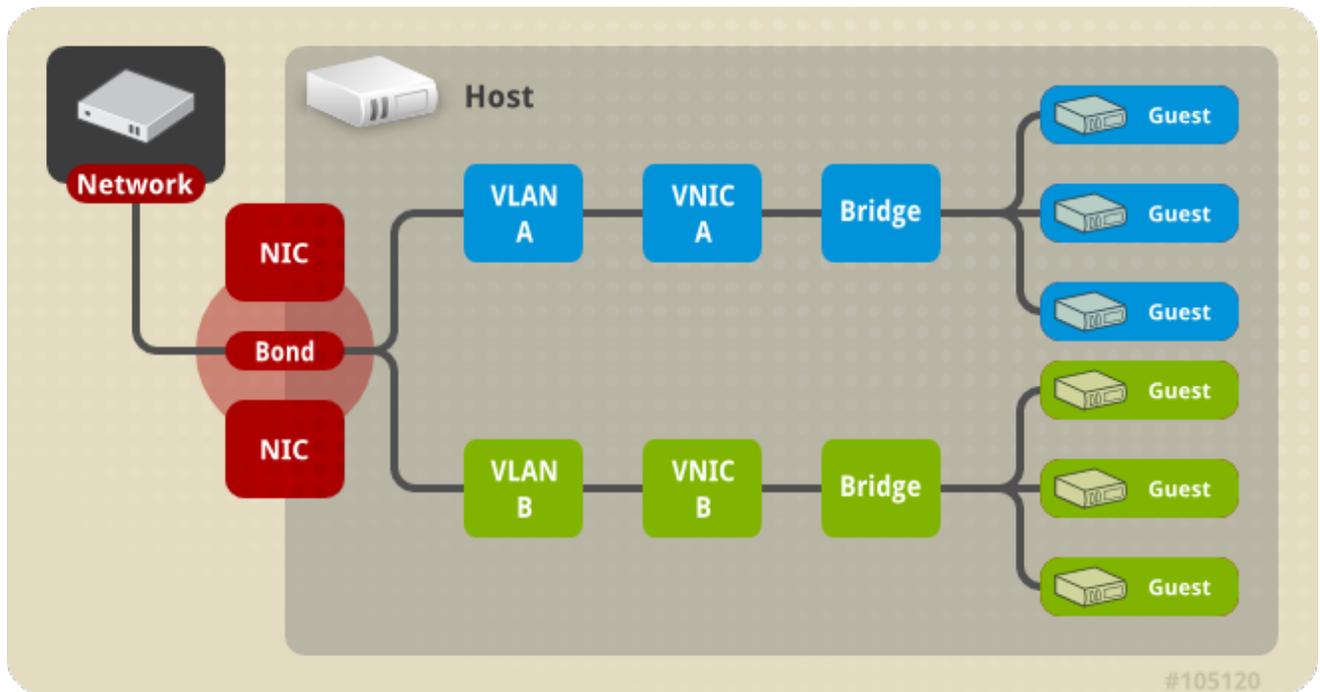


図3.7 ボンディング接続を使用した複数のブリッジ、複数の VLAN、複数の NIC

この構成では、それぞれの VLAN が複数の NIC で構成されるボンディングで定義されます。各 VLAN には個別のブリッジが接続され、各ブリッジには単一または複数のゲストが接続されます。

第4章 電源管理

4.1. 電源管理とフェンシングについて

Red Hat Virtualization 環境は、電源管理とフェンシングを設定することで柔軟性および回復力が最も高くなります。電源管理を使用すると、Red Hat Virtualization Manager はホストの電源サイクルの操作を制御できるのに加えて、最も重要なのは問題が検出されたホストをリブートできるようになることです。フェンシングは、問題のあるホストをリブートして、稼働中の Red Hat Virtualization 環境から分離して、パフォーマンスの低下を防ぐのに使用します。フェンスされたホストは、管理者の操作によって応答可能な状態に戻した後に、再度環境に復帰させることができます。

電源管理とフェンシングには、ホストのオペレーティングシステムには依存せずにホストを再起動するための特殊な専用ハードウェアを使用します。Red Hat Virtualization Manager は、ネットワーク IP アドレスまたはホスト名を使用して電源管理デバイスに接続します。Red Hat Virtualization では、電源管理デバイスとフェンスデバイスは同じデバイスを意味します。

4.2. RED HAT VIRTUALIZATION におけるプロキシを使用した電源管理

Red Hat Virtualization Manager はフェンスエージェントとは直接通信を行いません。その代わりに、Manager はプロキシを使用して電源管理のコマンドをホストの電源管理デバイスに送ります。Manager は VDSM を利用して電源管理デバイスの操作を実行し、環境内の別のホストがフェンシングプロキシとして使用されます。

以下のいずれかを選択することができます。

- フェンシングが必要なホストと同じクラスター内にある任意のホスト
- フェンシングが必要なホストと同じデータセンター内にある任意のホスト

有効なフェンシングプロキシホストのステータスは **UP** または **Maintenance** です。

4.3. 電源管理

Red Hat Virtualization Manager は、非稼働状態または応答なしの状態となったホストを再起動したり、節電のために稼働率の低いホストの電源をオフにする準備をしたりすることができます。この機能は、電源管理デバイスが正しく設定されているかどうかによって左右されます。Red Hat Virtualization 環境は以下のような電源管理デバイスをサポートしています。

- *American Power Conversion (apc)*
- *Bladecenter*
- *Cisco Unified Computing System (cisco_ucs)*
- *Dell Remote Access Card 5 (drac5)*
- *Dell Remote Access Card 7 (drac7)*
- *Electronic Power Switch (eps)*
- *HP BladeSystem (hpblade)*
- *Integrated Lights Out (ilo、ilo2、ilo3、ilo4)*
- *Intelligent Platform Management Interface (ipmilan)*

- *Remote Supervisor Adapter*(rsa)
- rsb
- *Western Telematic, Inc* (wti)



注記

apc フェンスエージェントは、APC 5.x 電源管理デバイスをサポートしていません。代わりに **apc_snmp** フェンスエージェントを使用してください。

上記の電源管理デバイスと通信を行うために、Red Hat Virtualization Manager は フェンスエージェントを使用します。Red Hat Virtualization Manager では、管理者は環境内の電源管理デバイスに対して、そのデバイスが受け入れ、応答するパラメーターを指定してフェンスエージェントを設定することができます。基本設定オプションは、グラフィカルユーザーインターフェースを使用して設定することができます。特殊な設定オプションも入力できます。これは、未解析でフェンスデバイスに渡されます。特殊な設定オプションは所定のフェンスデバイス固有ですが、基本設定オプションはサポートされている全電源管理デバイスによって提供される機能を対象としています。全電源管理デバイスによって提供される基本的な機能は以下のとおりです。

- **Status:** ホストのステータスを確認します。
- **Start:** ホストの電源を投入します。
- **Stop:** ホストの電源を切断します。
- **Restart:** ホストを再起動します。実際には、stop、wait、status、start、wait、status として実装されます。

電源管理設定のテストは、初期設定完了時に1回、それ以降も機能が適切に稼働を継続するよう時々実行するのがベストプラクティスです。

環境内の全ホストで電源管理デバイスを適切に設定することにより、耐障害性が提供されます。フェンスエージェントにより、Red Hat Virtualization Manager は問題のあるホストのオペレーティングシステムを通さずホストの電源管理デバイスと通信し、そのホストをリブートすることにより、環境内のその他のリソースからホストを切り離すことができます。次に Manager は、問題の発生したホストが SPM ロールを保持している場合には、そのロールを再度割り当てて、他のホスト上の高可用性仮想マシンを安全に再起動することができます。

4.4. フェンシング

Red Hat Virtualization 環境におけるフェンシングとは、フェンスエージェントを使用して Manager により開始されて、電源管理デバイスにより実行されるホストの再起動のことです。フェンシングにより、クラスターはホストの予期せぬ障害に対応したり、節電や負荷分散、仮想マシンの可用性のポリシーを施行したりすることができます。

フェンシングは、Storage Pool Manager (SPM) のロールが常に稼働中のホストに割り当てられるようにします。フェンスされたホストが SPM の場合には、その SPM ロールが解除されて、応答可能なホストに再割り当てされます。SPM ロールが割り当てられたホストは、データドメイン構造のメタデータの書き込みができる唯一のホストであるため、SPM ホストが応答なしの状態になり、フェンスされていないと、その環境は仮想ディスクの作成や破棄、スナップショットの作成、論理ボリュームの拡張、およびデータドメイン構造のメタデータの変更が必要なその他すべてのアクションを実行する機能を失ってしまいます。

ホストが応答なしの状態となった場合には、そのホストで実行中の仮想マシンもすべて応答なしの状態

になる可能性があります。応答なしのホストは、実行している仮想マシンの仮想マシンハードディスクイメージのロックを保持しています。第 2 のホストで仮想マシンを起動し、仮想マシンハードディスクイメージの書き込み権限を 2 番目のホストに割り当てようとするとデータが破損する可能性があります。

フェンシングにより、Red Hat Virtualization Manager は仮想マシンハードディスクイメージのロックが解除されていると想定することができます。これにより、Manager がフェンスエージェントを使用して問題のあるホストがリブートされていることを確認できます。この確認を受信すると、Red Hat Virtualization Manager は、データ破損のリスクを冒すことなく、問題のあるホストの仮想マシンを別のホストで起動することができます。フェンシングは高可用性仮想マシンの基盤です。高可用性に指定された仮想マシンは、データが破損しないことが確信できなければ、別のホストで安全に起動することはできません。

Red Hat Virtualization Manager は、ホストが応答なしの状態になった場合に、処置を取るまでに 30 秒の猶予期間を設けて、ホストが一時的なエラーから回復できるようにします。猶予期間を経過してもホストが応答なしの場合には、Manager は応答なしのホストのもたらず悪影響の緩和を自動的に開始します。Manager はホストの電源管理カードにフェンスエージェントを使用して、まずホストを停止し、停止したことを確認したら、ホストを起動して、起動の確認を行います。ホストのブートが完了すると、フェンスされる前のクラスターに再度参加させるように試みます。ホストが応答なしの状態となった原因が再起動により解決している場合には、ホストは自動的に **Up** のステータスに設定され、仮想マシンの起動およびホスティングが再び可能となります。

4.5. ホストのソフトフェンシング

ホストは、予期しない問題が原因となって応答なしの状態になる場合があります。VDSM は要求に応答できませんが、VDSM に依存している仮想マシンは稼働を続け、アクセス可能な状態のままとなります。このような状況が発生した場合には、VDSM を再起動すると、VDSM が応答可能な状態に戻り、問題は解決します。

「SSH を介したソフトフェンシング」は、Manager が SSH を使用して、応答しない状態のホストで VDSM の再起動を試みるプロセスです。Manager が SSH を使用した VDSM の再起動に失敗した場合には、フェンシングは外部のフェンスエージェントの責任となります (外部のフェンスエージェントが設定されている場合)。

SSH ソフトフェンシングが機能するためには、ホストでフェンシングが設定および有効化されており、かつ有効なプロキシホスト (同じデータセンター内にある、ステータスが Up の第 2 のホスト) が存在する必要があります。Manager とホスト間の接続がタイムアウトになると、次のような状態となります。

1. 初回のネットワーク障害発生時には、ホストのステータスが「connecting」に変わります。
2. Manager は次に VDSM に対してステータス確認を 3 回試みるか、ホストの負荷によって決定される時間が経過するのを待ちます。この時間は、 $[\text{TimeoutToResetVdsInSeconds (デフォルトは 60 秒)}] + [\text{DelayResetPerVmInSeconds (デフォルトは 0.5 秒)}] * [\text{ホスト上で実行中の仮想マシン数}] + [\text{DelayResetForSpmlnInSeconds (デフォルトは 20 秒)}] * [1 (\text{ホストが SPM として稼働している場合}) \text{ または } 0 (\text{ホストが SPM としては稼働していない場合})]$ の計算式で決定されます。VDSM が応答する時間を最大限にするために、Manager は上記のオプション (VDSM のステータス確認を 3 回試みる、または上記の計算式で決定される時間の経過を待つ) でいずれか長い方を選択します。
3. この時間が経過してもホストが応答しない場合には、SSH を介して **vdsml restart** が実行されます。
4. **vdsml restart** を実行しても、ホストと Manager 間の接続が再度確立されない場合には、ホストのステータスが **Non Responsive** に変わります。電源管理が設定されている場合には、フェンシングは外部のフェンスエージェントによって引き継がれます。



注記

SSH を介したソフトフェンシングは、電源管理を設定していないホストに対しても実行することが可能です。これは、「フェンシング」とは異なります。フェンシングは、電源管理が設定されたホストでしか実行することはできません。

4.6. 複数の電源管理フェンスエージェントの使用

単一のエージェントは、プライマリーエージェントとして扱われます。フェンスエージェントが2つある場合には、セカンダリーエージェントが有効になります (例: デュアルパワーのホストで、各電源スイッチに、同じ電源スイッチに接続されたエージェントが2つある場合)。エージェントは、同じタイプまたは異なるタイプを使用することができます。

1 台のホストに複数のフェンスエージェントがあると、フェンシングプロシージャの信頼性が高くなります。たとえば、ホストに単一のフェンスエージェントしかない場合には、そのフェンスエージェントに障害が発生すると、ホストは手動でリブートするまで非稼働状態のままとなります。そのホストで実行されていた仮想マシンは一時停止され、元のホストが手動でフェンスされてからでないと、クラスター内の別のホストにフェイルオーバーされません。複数のエージェントがある場合には、第1のエージェントに障害が発生しても、第2のエージェントを呼び出すことができます。

1 台のホストで2つのフェンスエージェントが定義されている場合には、*同時*または*順次*のフローを使用するように設定することができます。

- **同時:** ホストが停止するには、プライマリーエージェントとセカンダリーエージェントの両方が停止のコマンドに応答する必要があります。一方のエージェントが起動のコマンドに応答すると、ホストが起動します。
- **順次:** ホストの停止または起動には、プライマリーエージェントが最初に使用され、それが失敗した場合にセカンダリーエージェントが使用されます。

第5章 負荷分散、スケジューリング、移行

5.1. 負荷分散、スケジューリング、移行

個々のホストのハードウェアリソースには限りがあり、障害の影響を受けやすい状態です。このため、複数のホストをクラスターにグループ化して、実質的に共有リソースを1つにまとめ、障害の発生やリソース消費を軽減します。Red Hat Virtualization 環境では、負荷分散ポリシー、スケジューリング、移行などを使用して、ホストリソースの需要の変化に対応します。Manager により、クラスター内の全仮想マシンの負荷が1台のホストだけにかからないようにできます。また逆に、Manager が稼働率の低いホストを認識して、そのホストから仮想マシンを別のホストに移行して、管理者がそのホストをシャットダウンして節電できるようにすることも可能です。

使用可能なリソースは、以下の3つのイベントでチェックされます。

- 仮想マシンの起動: リソースをチェックして、仮想マシンを起動するホストを決定します。
- 仮想マシンの移行: リソースをチェックして、移行先となる適切なホストを決定します。
- 経過時間: 一定の間隔でリソースをチェックし、各ホストの負荷が、クラスター負荷分散ポリシーに適合しているかどうかを確認します。

Manager は、クラスターを対象とする負荷分散ポリシーを使用して、クラスター内のホスト間における仮想マシン移行をスケジュールすることにより、使用可能なリソースの変化に対応します。負荷分散ポリシー、スケジューリング、仮想マシン移行の間の関係については、以下のセクションで説明します。

5.2. 負荷分散ポリシー

負荷分散ポリシーは、1つのクラスターに対して設定します。クラスターには単一または複数のホストが含まれ、各ホストのハードウェアパラメーターや使用可能なメモリ容量が異なる場合があります。Red Hat Virtualization Manager は負荷分散ポリシーを使用してクラスター内のどのホストで仮想マシンを起動するかを決定します。また、負荷分散ポリシーにより、Manager は仮想マシンを使用率の高いホストから使用率の低いホストにいつ移動するかを決定することもできます。

負荷分散のプロセスは、データセンター内の各クラスターに対して毎分実行されます。このプロセスにより、使用率が高いホスト、使用率が低いホスト、および仮想マシンの有効な移行先を判断します。これは、管理者が特定のクラスターに対して設定する負荷分散ポリシーに基づいて決定されます。負荷分散ポリシーには、**VM_Evenly_Distributed**、**Evenly_Distributed**、**Power_Saving**、**None** のオプションがあります。

5.3. 負荷分散ポリシー: VM_EVENLY_DISTRIBUTED

仮想マシン均等配分負荷分散ポリシー (VM_Evenly_Distributed) では、仮想マシン数をベースに、ホスト間で仮想マシンが均等に配分されます。HighVmCount は各ホストで実行可能な最大仮想マシン数のことで、この値を超えるとホストが過負荷の状態であるとみなされます。このポリシーでは、管理者はホストで実行することができる仮想マシンの最大数を設定することができます。また、稼働率の最も高いホストと最も低いホストの間での仮想マシン数の差異の最大値 (この値を含む) も設定することができます。クラスター内の全ホストで仮想マシン数がこの移行閾値 (MigrationThreshold) 内に収まる場合は、このクラスターはバランスが取れた状態となります。また、管理者は、SPM ホスト上で仮想マシン用に確保されるスロット数に関する設定を行うこともできます。SPM ホストの負荷が他のホストよりも低くなるように、この変数で SPM ホストが他のホストよりもどれだけ少ない数の仮想マシンを実行するかを定義します。ホストが HighVMCount (最大仮想マシン数) を超える数の仮想マシンを実行しており、仮想マシン数が MigrationThreshold の範囲外となるホストが少なくとも1台ある場合には、仮想マシンは、CPU 使用率が一番少ないクラスター内のホストに1台ずつ移行されます。クラスター内の全ホストの仮想マシン数が移行閾値内に収まるまで、仮想マシンは1度に1台ずつ移行されます。

5.4. 負荷分散ポリシー: `EVENLY_DISTRIBUTED`

均等配分負荷分散ポリシー (Evenly_Distributed) では、CPU やメモリー使用率の低い順に、新規仮想マシン用のホストが選択されます。上限閾値は、クラスター内のホストに許容される最大の CPU およびメモリー使用率で、この値を超えると、環境のパフォーマンスが低下します。このポリシーにより、管理者は仮想マシン実行に関する上限閾値を設定することができます。また、Red Hat Virtualization Manager が介入するまでホストがこの上限閾値で稼働を持続できる時間も管理者が設定します。ホストが上限閾値に達し、その状態が所定の時間続くと、そのホスト上の仮想マシンは同じクラスター内で CPU またはメモリー使用率が最も低いホストに 1 台ずつ移行されます。ホストリソースは毎分チェックされ、ホストの CPU およびメモリー使用率が上限閾値を下回るまで、仮想マシンが 1 回に 1 台ずつ移行されます。

5.5. 負荷分散ポリシー: `POWER_SAVING`

省電力負荷分散ポリシー (Power_Saving) では、CPU およびメモリー使用率の低い順に、新規仮想マシン用のホストが選択されます。上限閾値は、クラスター内のホストに許容される最大の CPU およびメモリー使用率で、この値を超えると、環境のパフォーマンスが低下します。下限閾値は、許容される最小の CPU およびメモリー使用率で、この範囲を外れると、引き続きホストを運転することが電力の非効率な使用とみなされます。このポリシーにより、管理者は仮想マシン実行に関する上限および下限閾値を設定することができます。また、Red Hat Virtualization Manager が介入するまでホストがこの上限および下限閾値で稼働を持続できる時間も管理者が設定します。ホストが上限閾値に達し、その状態が所定の時間続くと、そのホスト上の仮想マシンは CPU およびメモリー使用率が最も低いホストに 1 台ずつ移行されます。このプロセスは、ホストの CPU およびメモリー使用率が上限閾値を下回るまで続行されます。ホストの CPU およびメモリー使用率が下限閾値を下回ると、仮想マシンはクラスター内の他のホストに移行されます。ただし、移行先のホストの上限閾値で許容される必要があります。使用率の低いホストの仮想マシンがすべて移行されると、Manager が自動的にホストマシンをシャットダウンして、負荷分散で必要になった場合やクラスター内に未使用のホストが不足した場合には再起動します。

5.6. 負荷分散ポリシー: `NONE`

負荷分散ポリシーを選択していない場合、仮想マシンはクラスター内で CPU 使用率とメモリー使用率の最も低いホストで起動します。CPU 使用状況を判断するには、仮想 CPU 数と CPU 使用率を考慮に入れた複合メトリックを使用します。このアプローチは、ホストを唯一選択できるタイミングが新規仮想マシンの起動時のみであるため、可変性が最も低い方法です。ホストに対する需要の増加を反映して、仮想マシンが自動的に移行されることはありません。

管理者は、仮想マシンの適切な移行先となるホストを決定する必要があります。仮想マシンは、ピンングを使用して特定のホストに関連付けすることも可能です。ピンングは、仮想マシンが他のホストに自動的に移行されるのを防ぎます。リソースの消費率が高い環境では、手動の移行が最適の方法です。

5.7. 高可用性仮想マシンの確保

高可用性 (HA) 仮想マシンの確保ポリシーでは、高可用性仮想マシン用のクラスターの容量を Red Hat Virtualization Manager によりモニタリングすることができます。Manager は、個別の仮想マシンに高可用性としてフラグを立てることができるため、ホストで障害が発生した場合、これらの仮想マシンは別のホストで再起動されます。このポリシーは、クラスター内のホスト全体で、高可用性の仮想マシンの負荷を分散し、クラスター内のホストに問題が発生した場合は、残りのホストがクラスターのパフォーマンスに影響を与えることなく、高可用性の仮想マシンの負荷の移行をサポートします。高可用性仮想マシン予約が有効な場合には、既存のホストで予期しないエラーが発生した際に Manager により高可用性仮想マシンの移行に適した容量がクラスター内で確保されます。

5.8. スケジューリング

Red Hat Virtualization においてスケジューリングとは、Red Hat Virtualization Manager が新規または移行対象の仮想マシンのターゲットとしてクラスター内のホストを選択する方法のことを意味します。

ホストが仮想マシンの起動先もしくは他のホストから移行される仮想マシンの受け入れ先の対象となるには、十分なメモリー空き容量があり、かつ CPU が仮想マシンの起動先/移行先となるための要件を満たしている必要があります。複数のホストが対象のターゲットである場合には、そのクラスターの負荷分散ポリシーに基づいて 1 台が選択されます。たとえば、Evenly_Distributed ポリシーが採用されている場合には、Manager は CPU 使用率が最も低いホストを選択します。Power_Saving ポリシーが採用されている場合には、上限閾値と下限閾値の間で CPU 使用率が最も低いホストが選択されます。そのホストの Storage Pool Manager (SPM) のステータスも、仮想マシンの起動先/移行先となる適格性に影響を及ぼします。SPM でないホストの方が優先されます。たとえば、クラスター内で最初に起動される仮想マシンは、そのクラスター内のホストが SPM ロールを保持している場合は、SPM ホストでは実行されません。

5.9. 移行

Red Hat Virtualization Manager は移行を使用してクラスターの負荷分散ポリシーを有効にします。仮想マシンの移行は、クラスターの負荷分散ポリシーとクラスター内のホストに対する現在の需要に応じて実行されます。移行は、ホストがフェンスされたり、メンテナンスモードに切り替えられたりした時に自動的に実行されるように設定することができます。Red Hat Virtualization Manager は最初に、CPU 使用率が最も低い仮想マシンを移行します。これはパーセンテージで算出され、RAM の使用状況や I/O 操作については考慮されません。ただし、CPU 使用率に影響を及ぼす I/O 操作は例外となります。CPU 使用率の同じ仮想マシンが複数ある場合には、Red Hat Virtualization Manager が仮想マシンの CPU 使用率を確認するために実行するデータベースクエリーで最初に返された仮想マシンが 1 番目に移行されます。

デフォルトで、仮想マシンの移行には以下の制限があります。

- 52 MiBps の帯域幅制限が各仮想マシンの移行に課せられます。
- 移行は、仮想マシンメモリー 1 GB につき 64 秒後にタイムアウトされます。
- 240 秒間進捗がない場合には、移行は中断されます。
- 同時移行を行う場合の移行元のホスト数は、各ホストの CPU コア 1 つにつき 1 台、もしくは 2 台のいずれか小さい値の台数に制限されます。

移行の設定調整に関する詳しい情報は、<https://access.redhat.com/ja/solutions/1579463> を参照してください。

第6章 ディレクトリーサービス

6.1. ディレクトリーサービス

Red Hat Virtualization プラットフォームでは、ユーザーの認証および承認についてはディレクトリーサービスに依存します。ユーザーポータル、管理ポータル、REST API を含む全 Manager インターフェースとの対話は、認証済み/承認済みのユーザーのみに限定されます。Red Hat Virtualization 環境内の仮想マシンは、同じディレクトリーサービスを使用して認証/承認を行うことができますが、認証/承認ができるように設定しておく必要があります。Red Hat Virtualization Manager での使用がサポートされているディレクトリーサービスのプロバイダーには、*Identity Management (IdM)*、*Red Hat Directory Server 9 (RHDS)*、*Active Directory (AD)*、*OpenLDAP* があります。Red Hat Virtualization Manager は、以下のような目的でディレクトリーサーバーに接続します。

- ポータルへのログイン (ユーザー、パワーユーザー、管理者、REST API)
- クエリーによるユーザー情報の表示
- ドメインへの Manager の追加

認証とは、データの生成者とその生成されたデータの整合性を検証/識別することです。プリンシパルとはアイデンティティーの検証を受ける側で、検証者とはプリンシパルのアイデンティティーの確認/保証を要求する側です。Red Hat Virtualization の場合は、Manager が検証者で、ユーザーがプリンシパルとなります。データの整合性とは、受信したデータがプリンシパルによって生成されたデータと同じであると保証することです。

機密性と承認は、認証と密接な関係にあります。機密性とは、目的の受信者以外にデータが開示されないように保護することです。強固な認証メソッドでは、オプションで機密性に対応します。承認により、プリンシパルが操作を実行可能かどうか判断されます。Red Hat Virtualization では、ディレクトリーサービスを使用してユーザーとロールを関連付けし、それに応じて承認を行います。承認は通常、プリンシパルが認証された後に行われ、承認の際にベースとする情報は、検証者にとってローカルの場合もリモートの場合もあります。

インストール中には、Red Hat Virtualization 環境の管理用にローカルの内部ドメインが自動的に設定されます。インストールの完了後には、ドメインの追加が可能になります。

6.2. ローカル認証: 内部ドメイン

Red Hat Virtualization Manager は、インストール中に、制限付きの内部管理ドメインを作成します。このドメインは、ディレクトリーサーバー上のディレクトリーサービスユーザーとしてではなく、Red Hat Virtualization PostgreSQL データベース内のキーに基づいて存在するので、AD や IdM ドメインとは異なります。また内部ドメインには、**admin@internal** の 1 ユーザーしかいない点も外部ドメインとは異なります。このアプローチを使用して、初回の認証を行うことにより、完全に機能するディレクトリーサーバーなしに Red Hat Virtualization を評価することができるのに加えて、外部ディレクトリーサービスに伴う問題のトラブルシューティングに管理アカウントを確実に使用できるようになります。

admin@internal ユーザーは、環境の初期設定を行うユーザーです。これにはホストのインストールや受け入れ、外部の AD/IdM 認証ドメインの追加、外部ドメインのユーザーに対するパーミッションの付与などが含まれます。

6.3. GSSAPI を使用したリモート認証

Red Hat Virtualization におけるリモート認証とは、Red Hat Virtualization Manager の外部で処理される認証のことを意味します。リモート認証は、AD、IdM、または RHDS ドメイン内部から Manager に接続するユーザーまたは API を対象に使用されます。ドメインに参加するには、管理者が **engine-**

manage-domains ツールを使用して Red Hat Virtualization Manager の設定を行う必要があります。これには、システムをドメインに参加させるのに十分な権限を使用して、Manager にディレクトリー内のアカウントの認証情報を提供する必要があります。ドメインに追加した後は、Red Hat Virtualization Manager がディレクトリーサーバーに対してパスワードを使用してドメインユーザーの認証を行うことが可能となります。Manager は *Simple Authentication and Security Layer (SASL)* と呼ばれるフレームワークを使用します。このフレームワークは *Generic Security Services Application Program Interface (GSSAPI)* を使用してユーザーのアイデンティティーをセキュアに検証し、そのユーザーに提供されている承認レベルを確認します。

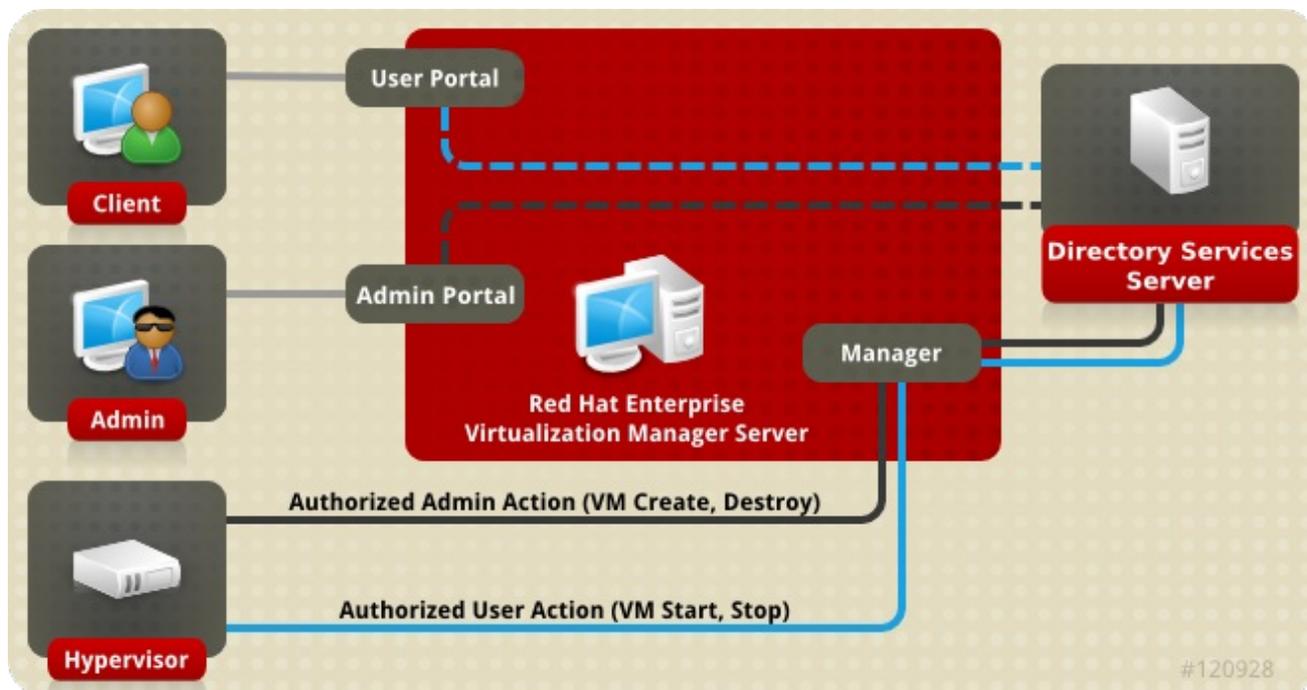


図6.1 GSSAPI 認証

第7章 テンプレートとプール

7.1. テンプレートとプール

Red Hat Virtualization 環境は、テンプレートとプールなど、ユーザーへの仮想マシンのプロビジョニングを簡略化する管理者向けツールを提供しています。テンプレートとは、既存の事前設定済み仮想マシンをベースに管理者が新規仮想マシンを迅速に作成できるようにするショートカットで、オペレーティングシステムのインストールや設定作業を省略できます。特に、テンプレートはアプライアンスのように使用する仮想マシン (例: Web サーバー仮想マシン) に便利です。特定の Web サーバーのインスタンスを多数使用している組織の場合には、管理者がテンプレートとして使用する仮想マシンを作成し、オペレーティングシステム、Web サーバー、補助パッケージをインストールし、独自の設定変更を適用します。これで管理者は実際に使用できる仮想マシンをベースにテンプレートを作成し、そのテンプレートを使用して全く同じ仮想マシンを必要に応じて新規作成できるようになります。

仮想マシンプールは、任意のテンプレートをベースにした仮想マシンのグループです。これにより、ユーザーへのプロビジョニングを迅速に行うことができます。プール内の仮想マシンを使用するためのパーミッションは、プールレベルで付与され、プールを使用するパーミッションを付与されたユーザーには、そのプールから仮想マシンが割り当てられます。仮想マシンプールには、そのプール内の仮想マシンの一時的な特性が継承されます。ユーザーが以前プール内のどの仮想マシンを使用したかには関係なく仮想マシンが割り当てられるので、プールは、データの永続性を必要とする目的には適していません。仮想マシンプールは、ユーザーデータが中央ロケーションに保管されていて、そのデータにアクセスして使用する一手段として仮想マシンを使用する場合や、データの永続性が重要でない場合に最も適しています。プールを作成すると、仮想マシンも作成され、停止した状態でそのプールの中に追加されます。これらの仮想マシンは、ユーザーの要求で起動します。

7.2. テンプレート

テンプレートを作成するには、仮想マシンを作成してカスタマイズします。任意のパッケージをインストールして、カスタマイズされた設定を適用し、使用目的に応じて準備をすることにより、デプロイ後に必要となる変更を最小限に抑えます。オプションとして、仮想マシンからテンプレートを作成する前には一般化することを推奨しています。一般化はシステムのユーザー名やパスワード、タイムゾーンなど、デプロイ時に変更される情報を削除するのに使用します。一般化はカスタマイズされた設定には影響を及ぼしません。Red Hat Virtualization 環境における Windows および Linux ゲストの一般化については、『仮想マシン管理ガイド』の「[テンプレート](#)」で詳しく説明しています。Red Hat Enterprise Linux ゲストの一般化には **sys-unconfig**、Windows ゲストの一般化には **sys-prep** を使用します。

テンプレートのベースとなる仮想マシンを適切に設定し、また必要な応じて汎用化を行い、停止した後に、管理者はその仮想マシンからテンプレートを作成することができます。仮想マシンからテンプレートを作成すると、特別に設定された仮想ディスクイメージの読み取り専用コピーが作成されます。読み取り専用イメージは、そのテンプレートをベースに以降作成される全仮想マシンの補助イメージを形成します。つまり、テンプレートは、基本的には、仮想ハードウェア設定に関連付けされた、カスタマイズ済みの読み取り専用ディスクイメージということになります。テンプレートから作成した仮想マシン内のハードウェアは変更可能です。たとえば、1 ギガバイトの RAM を搭載したテンプレートから作成した仮想マシンの RAM を 2 ギガバイトにすることが可能です。ただし、テンプレートのディスクイメージを変更することはできません。そのテンプレートをベースにした全仮想マシンが変更されてしまうことになるためです。

作成されたテンプレートは、複数の仮想マシンのベースとして使用することができます。テンプレートから仮想マシンを作成する際には、シンプロビジョニングメソッドまたはクローンプロビジョニングメソッドのいずれかを使用します。テンプレートからクローン作成される仮想マシンは、テンプレートベースイメージの完全な書き込み可能コピーを取得します。この場合、シンプロビジョニング作成メソッドによるスペース節約は犠牲となりますが、仮想マシンはテンプレートの存在に依存しなくなります。シンプロビジョニングメソッドを使用してテンプレートから作成した仮想マシンは、テンプレートの読み取り専用イメージをベースイメージとして使用するので、そのテンプレートおよびそのテンプレートから作成された全仮想マシンを同じストレージドメインに保管する必要があります。データへの

変更および新たに生成されたデータは、Copy On Write イメージに保管されます。テンプレートをベースとする各仮想マシンは、同じ読み取り専用ベースイメージと、その仮想マシン固有の Copy On Write イメージを使用します。これにより、全く同じデータがストレージに保管される回数が制限されるので、ストレージの節約となります。また、読み取り専用イメージを頻繁に使用すると、アクセスされるデータがキャッシュされ、ネットパフォーマンスが向上します。

7.3. プール

仮想マシンプールにより、多数の同じ仮想マシンをデスクトップとして迅速にユーザーにプロビジョニングすることができます。プールから仮想マシンにアクセスするパーミッションを付与されているユーザーには、要求キュー内の位置に応じて使用可能な仮想マシンが提供されます。プール内の仮想マシンはデータを永続化させることはできません。プールから仮想マシンが割り当てられる際は毎回、ベース状態で割り当てられます。これは、ユーザーのデータが中央で保管されている場合に適しています。

仮想マシンプールは、テンプレートから作成されます。プール内の各仮想マシンは、同じ読み取り専用のバックイメージを使用し、一時的な Copy On Write イメージで変更されたデータおよび新規生成されたデータを保管します。プール内の仮想マシンは他の仮想マシンとは異なるため、ユーザーが生成/変更したデータを格納する Copy On Write 層はシャットダウン時に失われます。これは、仮想マシンプールには、プールをバックアップするテンプレートと、使用中に生成/変更されたデータ用のスペース用の容量を越えるストレージは必要ないことを意味します。仮想マシンプールは、各ユーザーに専用の仮想デスクトップを提供する場合のようにストレージコストをかけずに、タスクを処理するための演算能力をユーザーに提供する効率的な方法です。

例7.1 プールの使用例

技術サポートサービスを提供する某企業では、ヘルプデスクスタッフを 10 人雇用していますが、常時勤務しているのは 5 人のみです。各ヘルプデスクスタッフ 1 名につき 1 台、合計 10 台の仮想マシンを作成する代わりに、仮想マシン 5 台で構成されるプールを 1 つ作成することができます。ヘルプデスクスタッフは、シフト勤務の開始時に自分で仮想マシンを 1 台割り当て、終了時にはその仮想マシンをプールに戻します。

第8章 仮想マシンのスナップショット

8.1. スナップショット

スナップショットは、管理者が特定の時点の仮想マシン/アプリケーション/データの復元ポイントを作成することができるストレージ機能です。スナップショットでは、仮想マシンのハードディスクイメージに現在存在するデータが COW ボリュームとして保存され、スナップショット取得時に存在していたデータに復元できます。スナップショットにより、現在の層の上に新規 COW 層が作成されます。スナップショットの取得後に実行された書き込み操作はすべて新規 COW 層に対して行われます。

仮想マシンのハードディスクイメージは単一または複数のボリュームからなるチェーンであることを理解しておくことが重要です。仮想マシンから見ると、これらのボリュームは単一のディスクイメージに見えます。仮想マシンは、そのディスクが実際には複数のボリュームから構成されていることは認識しません。

COW ボリュームと COW 層という用語は同じ意味で使われていますが、層の方がスナップショットの一次的な性質を明確に認識します。スナップショットを作成することにより、管理者がそのスナップショットの**作成後に**、データに加えた変更が満足できなかった場合は、その変更の破棄が可能となります。スナップショットにより、多くのワードプロセッサで使用されている **元に戻す** 機能と同様の機能が提供されます。



注記

共有可能 と指定した仮想マシンハードディスクと **直接 LUN** 接続をベースとする仮想マシンハードディスクのスナップショットは、ライブでもそれ以外でもサポートされません。

スナップショットは、以下のような 3 つの主要な操作があります。

- 作成: 仮想マシンの初回のスナップショットを作成する操作
- プレビュー: スナップショットが作成された時点までシステムデータを復元するかどうかを判断するためにスナップショットをプレビューする操作
- 削除: 必要がなくなった復元ポイントを削除する操作

スナップショットの操作に関するタスクベースの情報は、『Red Hat Virtualization 仮想マシン管理ガイド』の「[スナップショット](#)」を参照してください。

8.2. RED HAT VIRTUALIZATION でのライブスナップショット

共有可能 と指定した仮想マシンハードディスクと **直接 LUN** 接続をベースとする仮想マシンハードディスクのスナップショットは、ライブでもそれ以外でもサポートされません。

クローン作成中または移行中でないその他の仮想マシンでは、実行中、一時停止中、または停止中にスナップショットを作成することができます。

仮想マシンのライブスナップショットが開始されると、Manager は、仮想マシンが使用する新しいボリュームを作成するよう SPM ホストに要求します。新しいボリュームが利用可能な状態になると、Manager は VDSM を使用して、仮想マシン書き込み操作のために新しいボリュームを使用して起動する必要がある仮想マシンが実行されているホストの libvirt および qemu と通信します。仮想マシンが新しいボリュームに書き込むことができる場合には、スナップショット操作は成功と見なされ、仮想マシンは以前のボリュームへの書き込みを停止します。仮想マシンが新しいボリュームに書き込むことができない場合には、スナップショット操作は失敗と見なされ、新しいボリュームが削除されます。

ライブスナップショットの開始から新しいボリュームの準備後まで、仮想マシンは現在のボリュームと新しいボリュームの両方にアクセスできる必要があります (両方のボリュームが読み書きアクセスで開きます)。

休止をサポートするゲストエージェントがインストールされた仮想マシンでは、スナップショット作成の前後でファイルシステムの整合性を維持できます。登録された Red Hat Enterprise Linux ゲストには **qemu-guest-agent** をインストールして、スナップショット前の休止を有効にできます。

スナップショットの取得時に、休止に対応したゲストエージェントが仮想マシンに存在する場合には、VDSM は libvirt を使用してエージェントと通信し、スナップショットを作成します。未処理の書き込みアクションが完了し、スナップショットの取得前にファイルシステムが凍結されます。スナップショットが完了して、libvirt により仮想マシンがディスク書き込みアクションのために新しいボリュームに切り替えられると、ファイルシステムの解凍が解除され、ディスクへの書き込みが再開されます。

休止が有効な状態ですべてのライブスナップショットが試行されます。互換ゲストエージェントが存在しないため、スナップショットコマンドが失敗した場合には、スナップショットは休止の使用フラグなしで再び開始されます。仮想マシンが休止ファイルシステムでスナップショット前の状態に戻された場合には、仮想マシンはクリーンに起動し、ファイルシステムチェックは必要ありません。休止なしファイルシステムを使用して以前のスナップショットに戻すには、起動時にファイルシステムチェックを行う必要があります。

8.3. スナップショットの作成

Red Hat Virtualization では、仮想マシンの最初のスナップショットは、既存ボリュームの形式を引き継ぎ QCOW2 または RAW のいずれかとなります。これがそれ以降のスナップショットとは異なる点です。仮想マシンの最初のスナップショットでは、既存のボリュームがベースイメージとして使用されます。それ以降のスナップショットは COW 層として追加され、以前のスナップショット以降イメージに格納されたデータに加えられた変更を追跡します。

Red Hat Virtualization では、イメージがシンプロビジョニングされたイメージとして作成されない限り、またはユーザーが明示的に QCOW2 と指定しない限り、通常ゲスト仮想マシンは RAW ディスクイメージと対話します。図8.1「初回のスナップショット作成」に示したように、スナップショットを作成すると、仮想ディスクイメージを構成するボリュームはそれ以降のすべてのスナップショットのベースイメージとして機能します。

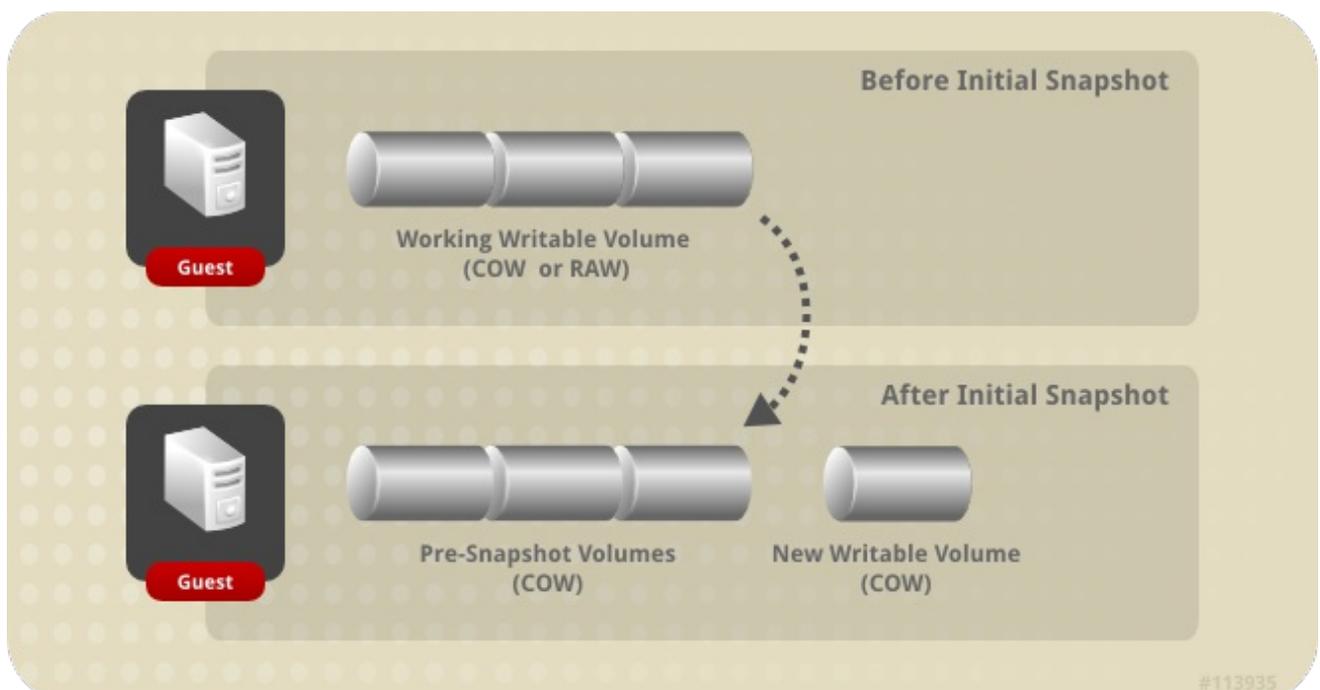


図8.1 初回のスナップショット作成

最初のスナップショット後にスナップショットを取得すると、新しい COW ボリュームが作成されます (このボリュームには、スナップショットの取得後に作成または変更されたデータが格納されます)。新しい COW 層の初期状態では、COW メタデータのみが格納されます。スナップショットの取得後、仮想マシンの使用や動作により作成されるデータは新しい COW 層に書き込まれます。仮想マシンを使用して以前の COW 層にあるデータを変更する場合には、データは以前の層から読み込まれてから最新の層に書き込まれます。仮想マシンは、最新の層から古い層へと順番に、その仮想マシンに対して透過的に各 COW 層をチェックしてデータを検索します。

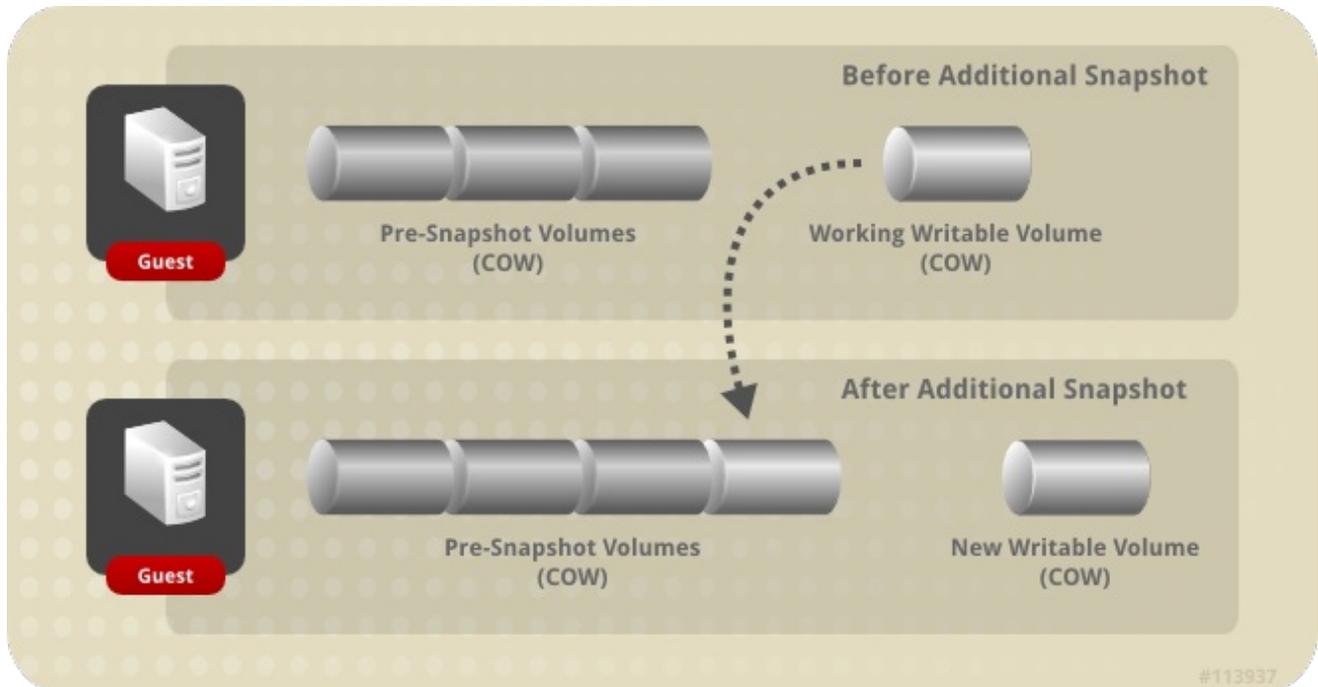


図8.2 追加スナップショットの作成

8.4. スナップショットのプレビュー

仮想ディスクイメージの復元ポイントとなるスナップショットを選択する際には、それまでに作成した全スナップショットをプレビューすることができます。

管理者は、ゲスト別に提供されているスナップショットから、スナップショットボリュームを選択してその内容をプレビューすることができます。図8.3「スナップショットのプレビュー」に示したように、各スナップショットは COW ボリュームとして保存され、プレビューされる際には、プレビューしているスナップショットから、新たなプレビュー層がコピーされます。ゲストは実際のスナップショットボリュームではなく、このプレビューと対話を行います。

選択したスナップショットをプレビューした後には、そのプレビューをコミットして、ゲストのデータをスナップショットにキャプチャーされている状態に復元することができます。管理者がプレビューをコミットすると、ゲストはそのプレビュー層にアタッチされます。

スナップショットをプレビュー後、**元に戻す**を選択して表示したスナップショットのプレビュー層を破棄することができます。プレビュー層は破棄されますが、スナップショット自体を含む層は維持されます。

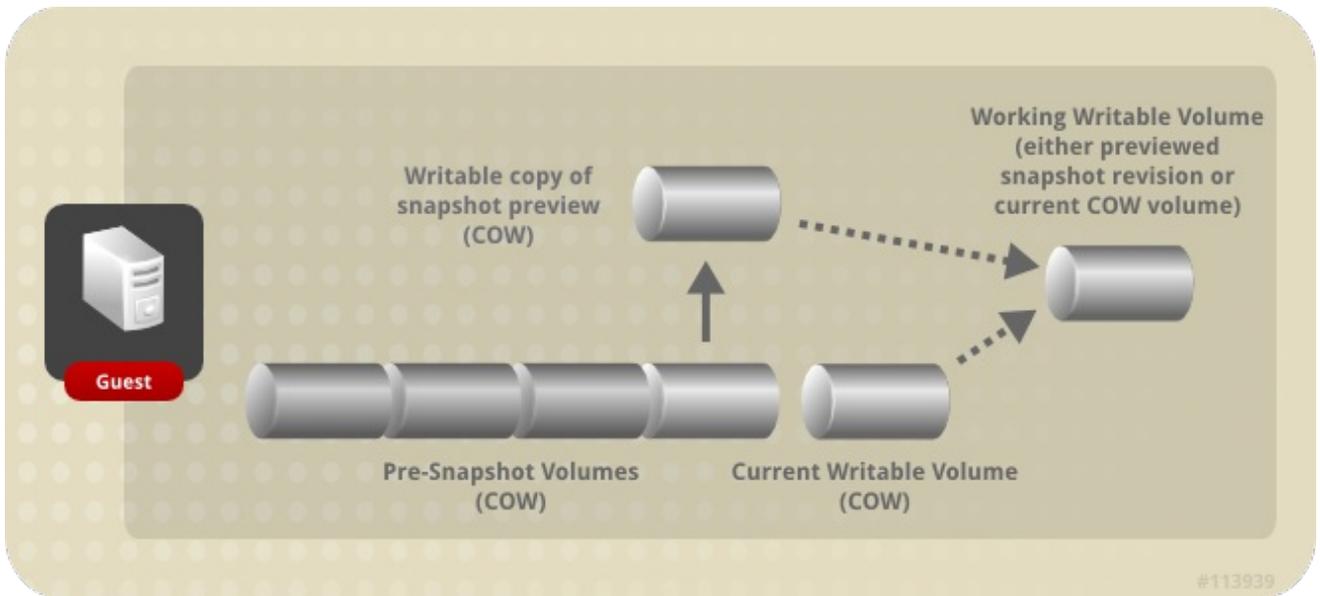


図8.3 スナップショットのプレビュー

8.5. スナップショットの削除

個別のスナップショットが必要なくなった場合には、スナップショットを削除することができます。スナップショットを削除すると、仮想ディスクイメージを特定の復元ポイントにリストアする能力が失われます。この操作によって、スナップショットが消費したディスク領域の再利用や、スナップショット内のデータの削除が必ずしも行われるわけではありません。このディスク領域は、削除したスナップショットのデータが後続のスナップショットにより上書きされた場合にのみ再利用されます。たとえば、5つあるスナップショットのうち3番目に作成したスナップショットを削除すると、4番目と5番目のスナップショットで使用できるように、3番目のスナップショットで変更のないデータはディスク上に確保する必要があります。しかし、4番目または5番目のスナップショットが3番目のスナップショットのデータを上書きした場合には、3番目のスナップショットは重複してしまうため、このディスク領域は再利用することができます。スナップショットを削除すると、ディスク領域の再利用が可能であるだけでなく、仮想マシンのパフォーマンスも向上できる場合もあります。

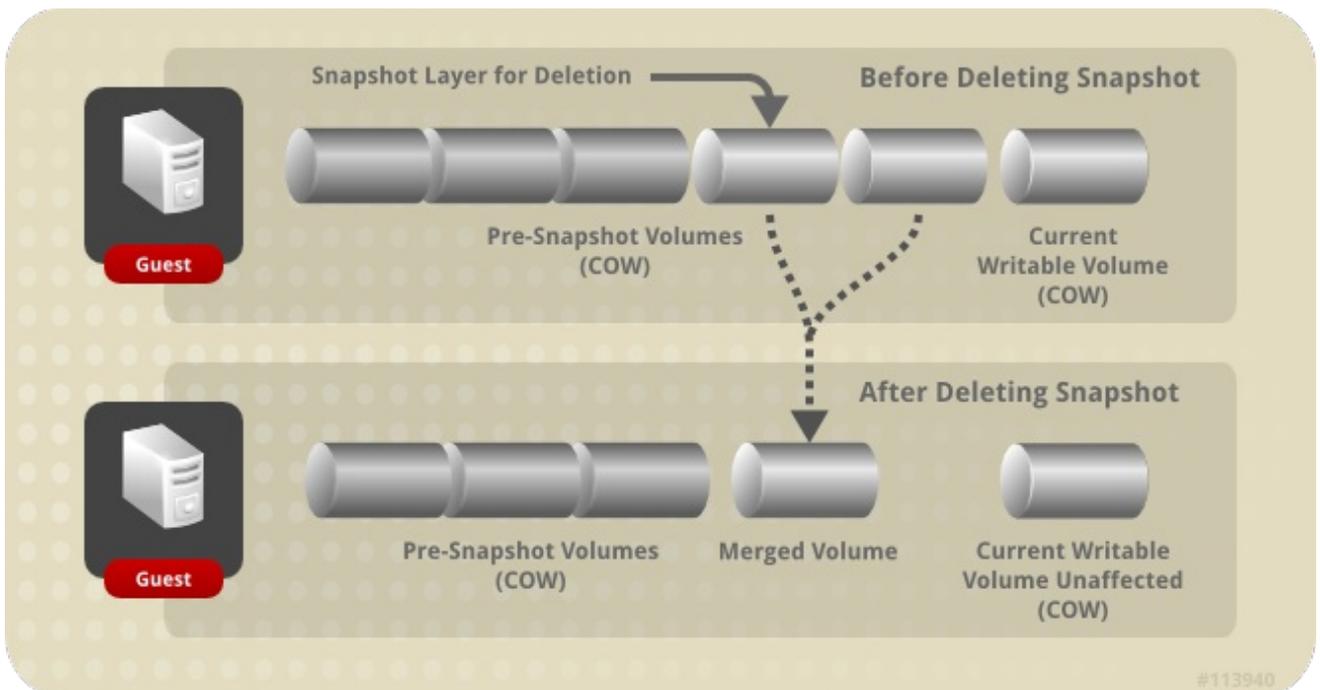


図8.4 スナップショットの削除

スナップショットの削除は、非同期ブロックジョブとして処理されます。この際、VDSM が再起動された場合や仮想マシンが操作中にシャットダウンされた場合でもジョブのトラッキングができるように、VDSM は、仮想マシンのリカバリーファイルに操作の記録を保持します。この操作が一旦開始すると、操作が失敗したり中断されたりしても、削除されるスナップショットをプレビューしたり、復元ポイントとして使用したりすることはできません。アクティブなレイヤーを親レイヤーとマージする操作では、データがアクティブなレイヤーから親レイヤーにコピーされるプロセスと、ディスクへの書き込みがアクティブなレイヤーおよび親レイヤーの両方にミラーリングされるプロセスの2段階に、操作が分割されます。最終的に、削除されるスナップショットのデータが親スナップショットとマージされ、VDSM によりイメージチェーン全体で変更が同期された時点で、このジョブは完了とみなされます。

第9章 ハードウェアのドライバーとデバイス

9.1. 仮想ハードウェア

Red Hat Virtualization は、3つの異なるタイプのシステムデバイスを仮想化ゲストに提示します。これらのハードウェアデバイスはすべて、仮想化ゲストに物理的に接続されたハードウェアデバイスのように表示されますが、デバイスドライバーの機能の仕方が異なります。

エミュレーションデバイス

エミュレーションデバイスは、*仮想デバイス*とも呼ばれ、完全にソフトウェア内に存在します。エミュレーションデバイスドライバーとは、ホスト上で実行しているオペレーティングシステム(ソースデバイスを管理)とゲストで実行しているオペレーティングシステム間の変換層です。デバイスレベルにおけるエミュレーションデバイスとの指示のやりとりは、ハイパーバイザーによってインターセプトされて、変換されます。Linux カーネルで認識される、同じタイプのエミュレーションデバイスはいずれも、エミュレーションドライバーのバックングソースデバイスとして使用可能です。

準仮想化デバイス

準仮想化デバイスには、ゲストオペレーティングシステムにデバイスドライバーをインストールして、ホストマシン上のハイパーバイザーと通信するためのインターフェースを提供する必要があります。このインターフェースを使用すると、ディスク I/O などの従来集中的なタスクを仮想化環境外で実行することができます。このような方法による仮想化固有のオーバーヘッド低減は、ゲストオペレーティングシステムのパフォーマンスを物理ハードウェア上で直接実行している場合のパフォーマンスに近づけることを目的としています。

物理共有デバイス

一部のハードウェアプラットフォームでは、仮想ゲストが直接ハードウェアデバイスやコンポーネントにアクセスすることができます。仮想化においてこのプロセスは、*パススルー*または *デバイス割り当て*として知られています。パススルーにより、デバイスはゲストオペレーティングシステムに物理的にアタッチされているように表示され、動作します。

9.2. RED HAT VIRTUALIZATION における不変のデバイスアドレス

仮想ハードウェアの PCI アドレスは ovirt-engine データベース内に永続的に割り当てられます。

PCI アドレスは **QEMU** により仮想マシンの作成時に割り当てられ、**libvirt** により **VDSM** に報告されます。それらのアドレスは **VDSM** によって Manager へ折り返し報告され、ovirt-engine データベースに保管されます。

仮想マシンが起動すると、Manager はデータベースに保管されているデバイスアドレスを **VDSM** に送ります。**VDSM** はそのアドレスを **libvirt** に渡すと、仮想マシンの初回実行時に割り当てられた PCI デバイスアドレスを使用して仮想マシンが実行されます。

デバイスが仮想マシンから削除されると、その仮想マシンへの全参照(不変の PCI アドレスを含む)も削除されます。削除されたデバイスの代わりにデバイスが追加される場合には、**QEMU** によりそのデバイスに割り当てられる PCI アドレスは、削除されたデバイスのアドレスとは異なる可能性が高くなります。

9.3. CPU (CENTRAL PROCESSING UNIT)

クラスター内の各ホストには、複数の *仮想 CPU* (vCPU) が搭載されています。この仮想 CPU は順にホスト上で実行しているゲストに対して公開されます。クラスター内でホストによって公開される仮想

CPU はすべて、Red Hat Virtualization Manager で最初にクラスターを作成した際に選択したタイプになります。1 クラスター内で異なる CPU タイプを混在させることはできません。

使用できる各仮想 CPU タイプは、同じ名前の物理 CPU に基づいた特性があります。ゲストオペレーティングシステムには、仮想 CPU と物理 CPU の区別はつきません。



注記

x2APIC: のサポート

Red Hat Enterprise Linux 7 ホストが提供する仮想 CPU モデルはすべて、x2APIC をサポートしています。これにより、ハードウェアの割り込み処理を向上させる *Advanced Programmable Interrupt Controller (APIC)* が提供されます。

9.4. システムデバイス

システムデバイスは、ゲストの稼働に極めて重要なので削除できません。また、ゲストにアタッチされるシステムデバイスには、空き PCI スロットも 1 つずつ必要となります。デフォルトのシステムデバイスは以下のとおりです。

- ホストのブリッジ
- ISA ブリッジおよび USB ブリッジ (USB ブリッジと ISA ブリッジは同じデバイス)
- グラフィックカード (Cirrus または qxl のいずれかのドライバーを使用)
- メモリーバルーンデバイス

9.5. ネットワークデバイス

Red Hat Virtualization では、ゲストに対して 3 つの異なるタイプのネットワークインターフェースコントローラーを公開することができます。ゲストに公開するネットワークインターフェースコントローラーのタイプは、ゲストの作成時に選択されますが、Red Hat Virtualization Manager から変更することも可能です。

- **e1000** ネットワークインターフェースコントローラーは、仮想 Intel PRO/1000 (e1000) をゲストに公開します。
- **virtio** ネットワークインターフェースコントローラーは、準仮想化ネットワークデバイスをゲストに公開します。
- **rtl8139** ネットワークインターフェースコントローラーは、仮想 **Realtek Semiconductor Corp RTL8139** をゲストに公開します。

1 ゲストにつき複数のネットワークインターフェースコントローラーが許可されています。追加したコントローラー 1 つにつき、ゲスト上の空き PCI スロットが 1 つ必要となります。

9.6. グラフィックデバイス

2 つのエミュレーショングラフィックデバイスが提供されています。これらのデバイスは SPICE プロトコルまたは VNC で接続することができます。

- **ac97** は **Cirrus CLGD 5446 PCI VGA** カードをエミュレーションします。

- **vga** は、**BochsVESA** 拡張機能 (ハードウェアレベルで、全非標準モードを含む) が付いたダミー VGA カードをエミュレーションします。

9.7. ストレージデバイス

ストレージデバイスとストレージプールは、ブロックデバイスドライバーを使用してストレージデバイスを仮想化ゲストにアタッチすることができます。ストレージドライバーはストレージデバイスではない点に注意してください。ドライバーは、バックングストレージデバイス、ファイル、ストレージプールボリュームなどを仮想化ゲストにアタッチするのに使用します。サポートされている任意のタイプのストレージデバイス、ファイル、ストレージプールボリュームをバックングストレージデバイスにすることができます。

- **IDE** ドライバーは、エミュレーションブロックデバイスをゲストに公開します。エミュレーション **IDE** ドライバーを使用して、仮想 **IDE** ハードディスクおよび仮想 **IDE** CD-ROM ドライブの任意の組み合わせ (最大 4 つ) を各仮想化ゲストにアタッチすることができます。エミュレーション **IDE** ドライバーは、仮想 DVD-ROM ドライブの提供にも使用します。
- **VirtIO** ドライバーは、準仮想化ブロックデバイスをゲストに公開します。準仮想化ブロックドライバーとは、仮想化ゲストにアタッチされたハイパーバイザーがサポートする全ストレージデバイス用のドライバーです (エミュレーションが必要なフロッピーディスクドライブは除く)。

9.8. サウンドデバイス

2つのエミュレーションサウンドデバイスが使用可能です。

- **ac97** は、**Intel 82801AA AC97 Audio** 対応のサウンドカードをエミュレーションします。
- **es1370** は、**ENSONIQ AudioPCI ES1370** サウンドカードをエミュレーションします。

9.9. シリアルドライバー

準仮想化シリアルドライバー (**virtio-serial**) は、バイトストリーム指向の文字ストリームドライバーです。準仮想化シリアルドライバーは、ネットワークが提供されていないもしくは使用できない場合に、ホストのユーザー領域とゲストのユーザー領域を連結するシンプルな通信インターフェースを提供します。

9.10. バルーンドライバー

バルーンドライバーにより、ゲストが必要とするメモリー容量をハイパーバイザーに示すことができます。また、ホストが効率的にメモリーをゲストに割り当てて、解放されたメモリーを他のゲストやプロセスに割り当てることが可能となります。

バルーンドライバーを使用しているゲストは、そのゲストの RAM のセクションを未使用としてマークすることができます (バルーン膨張)。ハイパーバイザーは、そのメモリーを解放して、他のホストのプロセスや同じホスト上の他のゲストに使用することができます。そのホストで再度メモリーの解放が必要となった際には、ハイパーバイザーはそのゲストに RAM を再割り当てすることができます (バルーン収縮)。

第10章 最小要件および技術的な制限事項

10.1. 最小要件およびサポート制限事項

Red Hat Virtualization 環境には物理的および論理的な制限事項が複数適用されます。以下に記載する制限事項に対応していない環境は、現在サポートされていません。

10.2. データセンターの制限事項

管理対象の仮想環境において、全リソースの最上位コンテナとなるのがデータセンターです。各データセンターに格納されるリソースには、複数の制限事項が適用されます。

表10.1 データセンターの制限事項

項目	制限
ストレージドメイン数	<ul style="list-style-type: none"> 推奨される最小のストレージドメイン数は1データセンターあたり2つ。1x データストレージドメインは必須、1x ISO ストレージドメインは推奨。
ホスト数	<ul style="list-style-type: none"> 1 データセンターにつき最大 200 ホストをサポート

10.3. クラスターの制限事項

クラスターとは、仮想マシンセットのリソースプールとして扱う物理ホストのセットです。クラスター内のホストは、同じネットワークインフラストラクチャーとストレージを共有します。クラスターは、移行ドメインです。クラスター内で、仮想マシンのホスト間移行を行うことができます。安定性を確保するために、各クラスターには複数の制限事項が適用されます。

- 管理対象のハイパーバイザーがすべてクラスター内にあること。
- クラスター内の全管理対象のハイパーバイザーの CPU タイプが同じであること。Intel CPU と AMD CPU は、同一のクラスター内では共存できません。



注記

クラスターに関する詳細情報は『管理ガイド』の「[クラスター](#)」を参照してください。

10.4. ストレージドメインの制限事項

ストレージドメインは、仮想ディスクイメージや ISO イメージ用のストレージおよび仮想マシンのインポート/エクスポート用の容量を提供します。任意のデータセンター内に多数のストレージドメインを作成することができますが、各ストレージドメインには複数の制限事項および推奨事項が適用されます。

表10.2 ストレージドメインの制限事項

項目	制限
ストレージタイプ	<p>サポートされているストレージタイプ</p> <ul style="list-style-type: none"> ● Fibre Channel Protocol (FCP) ● Internet Small Computer System Interface (iSCSI) ● Network File System (NFS) ● POSIX 準拠ファイルシステム (POSIX) ● Red Hat Gluster Storage (GlusterFS) <p>Red Hat Virtualization 4.1 では 新規 ISO およびエクスポートストレージドメインは、ファイルベースのストレージ (NFS、Posix または GlusterFS) で提供することができます。</p>
論理ユニット番号 (LUN)	<p>iSCSI または FCP で提供される場合は、各ストレージドメインに許可される LUN は 300 以下。</p>
論理ボリューム (LV)	<p>Red Hat Virtualization では、論理ボリュームは仮想マシン、テンプレート、および仮想マシンのスナップショット用の仮想ディスクを指します。</p> <p>iSCSI または FCP によって提供されるストレージドメインの論理ボリューム数は、1 ドメインにつき 350 以下とすることを推奨します。1 つのストレージドメインで論理ボリュームがこの数を上回る場合には、使用可能な記憶域を別々のストレージドメインに分割し、それぞれの論理ボリューム数が 350 以下となるようにすることをお勧めします。</p> <p>この制限の根本的な原因は、LVM メタデータのサイズです。論理ボリュームの数が増加すると、その論理ボリュームに関連付けられている LVM メタデータも増加します。このメタデータのサイズが 1 MB 超過すると、新規ディスクやスナップショットの作成などのプロビジョニング操作のパフォーマンスが低下し、QCOW ディスクの実行時に論理ボリュームをシンプロビジョニングする際の lvextend 操作の実行時間が長くなります。</p> <p>論理ボリュームに関する詳しい情報は https://access.redhat.com/solutions/441203 を参照してください。</p>



注記

ストレージドメインに関する詳細情報は『管理ガイド』の「[ストレージ](#)」を参照してください。

10.5. RED HAT VIRTUALIZATION MANAGER の制限事項

Red Hat Virtualization Manager サーバーは Red Hat Enterprise Linux 7 を実行する必要があります。また、追加のハードウェア要件を満たす必要もあります。

項目	サポート制限事項
CPU	<p>物理 CPU は最小で 1 つ必要です。Red Hat Virtualization はホストで以下の CPU モデルの使用をサポートしています。</p> <ul style="list-style-type: none"> ● AMD Opteron G1 ● AMD Opteron G2 ● AMD Opteron G3 ● AMD Opteron G4 ● AMD Opteron G5 ● Intel Conroe ● Intel Penryn ● Intel Nehalem ● Intel Westmere ● Intel Haswell ● Intel SandyBridge Family ● IBM POWER 8 <p>すべての CPU が Intel® 64 または AMD64 CPU の拡張機能をサポートし、AMD-V™ または Intel VT® のハードウェア仮想化拡張機能が有効化されている必要があります。No eXecute フラグ (NX) のサポートも必要です。</p>
RAM	<p>各仮想マシンに必要な RAM の容量は以下の条件によって異なります。</p> <ul style="list-style-type: none"> ● ゲストオペレーティングシステムの要件 ● ゲストアプリケーションの要件 ● 仮想マシンのメモリアクティビティと使用率 <p>また、KVM では、仮想マシンの物理 RAM のオーバーコミットが可能です。必要に応じてのみ仮想マシンに RAM を割り当て、使用率が低い仮想マシンの RAM を swap に移動することでオーバーコミットを行います。</p> <p>最大/最小サポートメモリーについては https://access.redhat.com/ja/articles/1271503 を参照してください。</p>

項目	サポート制限事項
ストレージ	<p>ホストの内部ストレージのサポート最小容量は、以下の一覧に示した容量の合計です。</p> <ul style="list-style-type: none">● root (/) パーティションには少なくとも 6 GB のストレージが必要です。● /boot パーティションには少なくとも 1 GB のストレージが必要です。● /var パーティションには少なくとも 15 GB のストレージが必要です。セルフホストエンジンのデプロイメントには、最低でも 60 GB が必要です。● swap パーティションには、最小で 8 MB のストレージが必要です。swap パーティションの推奨サイズは、ホストのインストール先のシステムや、その環境で予想されるオーバーコミットのレベルの両方により異なります。詳しい情報は、https://access.redhat.com/ja/solutions/108483を参照してください。 <p>この計算は、ホストインストールの最小限のストレージ要件である点に注意してください。これよりも大きなストレージ容量を用いた、デフォルトの割り当てを使用することを推奨します。</p>
PCI デバイス	ネットワークコントローラーが少なくとも 1 つ必要です。推奨される最小帯域幅は 1 Gbps です。

重要

Red Hat Virtualization Host の起動時に、次のようなメッセージが表示される場合があります。

```
Virtualization hardware is unavailable.  
(No virtualization hardware was detected on this system)
```

この警告は、仮想化拡張機能が無効になっているか、ご使用のプロセッサに拡張機能が搭載されていないことを示しています。CPU が上記の拡張機能をサポートし、システムの BIOS で有効になっていることを確認してください。

プロセッサに仮想化拡張機能が搭載されているかどうか、またそれらの機能が有効化されているかどうかを確認するには、以下の手順で行います。

- ホストの起動画面で任意のキーを押し、一覧から **Boot** または **Boot with serial console** のエントリを選択します。**Tab** を押して選択したオプションのカーネルパラメーターを編集します。最後のカーネルパラメーターの後ろに空白があることを確認した上で **rescue** パラメーターを追記します。
- **Enter** を押して、レスキューモードで起動します。
- プロンプトが表示されたら、次のコマンドを実行して、ご使用のプロセッサに仮想化拡張があることと、仮想化拡張機能が有効化されていることを確認します。

```
# grep -E 'svm|vmx' /proc/cpuinfo
```

何らかの出力が表示されれば、プロセッサはハードウェアの仮想化が可能です。出力が何も表示されない場合でも、ご使用のプロセッサがハードウェア仮想化に対応している可能性があります。場合によっては、メーカーが BIOS で仮想化拡張機能を無効にしていることがあります。これに該当すると思われる場合には、メーカーが提供しているシステムの BIOS とマザーボードに関するマニュアルを参照してください。

- 追加のチェック事項として、**kvm** モジュールがカーネルで読み込まれているかどうかを確認します。

```
# lsmod | grep kvm
```

出力に **kvm_intel** または **kvm_amd** が含まれている場合は、**kvm** ハードウェア仮想化モジュールが読み込まれており、システムが要件を満たしていることになります。

10.7. ゲストの要件とサポート制限事項

ゲストに適用される要件および制限に関する詳しい情報は、[「Red Hat Enterprise Linux technology capabilities and limits」](#) および [「Virtualization limits for Red Hat Enterprise Virtualization」](#) を参照してください。

10.8. SPICE の制限事項

SPICE が現在サポートしている最大解像度は 2560x1600 ピクセルです。

10.9. その他の参考資料

以下にあげる追加資料は、Red Hat Virtualization ドキュメントスイートには含まれていませんが、Red Hat Virtualization 環境を管理するシステム管理者に役立つ情報が記載されています。これらの資料は、<https://access.redhat.com/documentation/en/red-hat-enterprise-linux/> から入手することができます。

『Red Hat Enterprise Linux - システム管理者ガイド』

Red Hat Enterprise Linux のデプロイ、設定、管理に関するガイド。

『Red Hat Enterprise Linux DM マルチパス機能ガイド』

Red Hat Enterprise Linux のデバイスマッパーマルチパス機能の使用方法に関するガイド。

『Red Hat Enterprise Linux インストールガイド』

Red Hat Enterprise Linux のインストールに関するガイド。

『Red Hat Enterprise Linux ストレージ管理ガイド』

Red Hat Enterprise Linux のストレージデバイスおよびファイルシステムの管理に関するガイド。

『Red Hat Enterprise Linux - 仮想化導入および管理ガイド』

Red Hat Enterprise Linux の仮想化テクノロジーのインストール、設定、管理およびトラブルシューティングに関するガイド。