

物理マシンから Redhat OpenStack & KVM 上への移行検証



物理マシンから Redhat OpenStack & KVM 上への移行検証

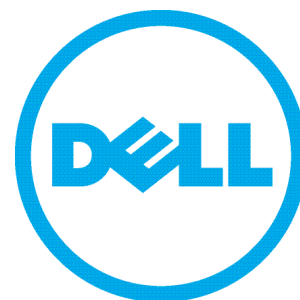
Version 1.0

2015 Mar.

Toshio Morimoto (TIS Inc.)

Yoshimi Tominaga (TIS Inc.)

Tsukasa Sato (Abelsoft Inc.)



物理マシンから Redhat OpenStack & KVM 上への移行検証

目次

1	はじめに.....	- 5 -
1.1	概要	- 5 -
1.2	目的	- 5 -
2	検証環境.....	- 6 -
2.1	検証に使用した機材	- 6 -
2.2	検証に使用した OS およびアプリケーション	- 6 -
2.3	IP アドレス	- 6 -
2.4	ネットワーク構成.....	- 7 -
3	検証方法.....	- 8 -
3.1	物理マシン→仮想マシンへの変換(P2V).....	- 8 -
3.2	Cloud Cloning	- 8 -
3.2.1	Live CD の選定.....	- 8 -
3.2.2	Cold Cloning を行うために必要な環境.....	- 9 -
3.2.3	Cold Cloning 作業の流れ.....	- 10 -
3.2.4	Cold Cloning 作業の注意点	- 12 -
3.2.5	Hot Cloning を行うために必要な環境.....	- 15 -
3.2.6	Hot Cloning 作業の流れ	- 16 -
3.2.7	Hot Cloning 作業の注意点.....	- 18 -
4	OpenStack の仮想マシンイメージ (Cloud image)	- 21 -
4.1	Cloud image の特徴.....	- 21 -
4.2	Linux の Cloud image	- 22 -
4.2.1	物理マシン上のディスク構成.....	- 22 -
4.2.2	Cloud image のディスク構成	- 22 -
4.3	Windows の Cloud image の特徴.....	- 23 -
4.3.1	物理マシン上のディスク構成.....	- 23 -

物理マシンから Redhat OpenStack & KVM 上への移行検証

4.3.2	Cloud image のディスク構成	23 -
4.4	物理マシンイメージから Cloud image への変換手順 (案)	24 -
4.6	Windows の場合	27 -
4.7	物理マシンから OpenStack への移行の限界.....	28 -
5	仮想マシンイメージから Cloud image への変換.....	29 -
5.1	Glance で登録可能な Cloud image のファイル形式.....	29 -
5.2	OpenStack で利用可能な Compute node	30 -
5.3	QEMU ディスクイメージ管理ツール (qemu-img)	30 -
5.4	その他準備が必要な環境	30 -
5.4.1	Linux KVM イメージの場合	31 -
5.4.2	Windows KVM イメージの場合.....	33 -
5.4.3	Linux VMware イメージの場合.....	35 -
5.4.4	Windows VMware イメージの場合.....	38 -
5.5	仮想マシンからの Cloud image 作成作業の注意点	41 -
5.5.1	ディスク容量による発生する問題と対策	41 -
5.5.2	Cloud image の形式の選択.....	41 -
5.5.3	VMware の場合の移行元仮想マシンイメージの形式.....	41 -
5.5.4	シック・シンプロビジョニング形式イメージのメリット・デメリット	42 -
5.5.5	仮想マシンイメージのカスタマイズ	43 -
5.6	仮想マシンからの Cloud image 作成の未検証部分	47 -
6	OpenStack 上でのインスタンス更新の運用.....	48 -
6.1	OpenStack 上の Snapshot を利用する	48 -
6.2	仮想マシンイメージのカスタマイズ環境を利用する	48 -
6.3	OpenStack 上の SnapShot を利用する場合の運用フロー	49 -
6.4	仮想マシンイメージのカスタマイズ環境を利用する倍の運用フロー	50 -
7	総括.....	51 -

物理マシンから Redhat OpenStack & KVM 上への移行検証

7.2	仮想マシンイメージを取得するデバイスの選択.....	- 51 -
7.3	移行検証・カスタマイズ用サーバの必要性.....	- 52 -
7.4	物理サーバを Cloud image に移行するのは有効なのか？.....	- 52 -
7.5	事前検証は必ず実施する	- 53 -
8	謝辞.....	- 53 -

物理マシンから Redhat OpenStack & KVM 上への移行検証

略号の説明

略 号	意 味
VM	<u>V</u> irtual <u>M</u> achine(仮想マシン)
RHEL6.6	<u>R</u> ed <u>H</u> at <u>E</u> nterprise <u>L</u> inux ver <u>6.6</u>
RHEL7.0	<u>R</u> ed <u>H</u> at <u>E</u> nterprise <u>L</u> inux ver <u>7.0</u>
OS	Operating System
QCOW2	<u>Q</u> EMU <u>C</u> opy- <u>O</u> n- <u>W</u> rite image file format version <u>2</u>
VMDK	<u>V</u> irtual <u>M</u> achine <u>D</u> is <u>K</u>
P2V	<u>P</u> hysical machine to(2) <u>V</u> irtual machine
V2C	<u>V</u> irtual machine to(2) openstack Cloud image
KVM	<u>K</u> ernel-based <u>V</u> irtual <u>M</u> achine
iDRAC8	integrated Dell Remote Access Controller version 8
GUI	Graphical User Interface
NFS	Network File System
NAS	Network Attached Storage
MBR	Master Boot Record
FC	Fibre Channel
UAC	User Account Control

物理マシンから Redhat OpenStack & KVM 上への移行検証

1 はじめに

1.1 概要

本書は物理マシン上で稼働している OS と稼働しているアプリケーションおよびデータを Redhat OpenStack および Redhat KVM で構成されたクラウド（サーバ仮想化）プラットフォームへの移行の検証を行った結果を記載している。移行の実施にあたり、必要となる移行用のプラットフォームおよび作業手順、注意すべき点や問題点を記載し、実際の移行作業で利用可能な技術情報をまとめた資料となっている。

1.2 目的

昨今は、サーバ性能の向上によりクラウド環境の処理速度が非常に向上しており、処理性能が要求されるシステムをクラウド上で稼働させることも十分に可能となっている。クラウド環境ではサーバリソースや機材設定スペースの効率化や電源使用量削減の効果を実現でき、コストやエコロジーの観点からも利用が拡大している。さらにパブリッククラウドでは Amazon Web Services の cloud formation、プライベートクラウドは OpenStack などのクラウドコンピューティング製品が提供されており、運用の効率化によるコスト効果も実現している。

クラウド利用は今後、益々盛んになることが予測されるが、クラウド上に新規に構築されるシステムと既存の物理マシン上で稼働しているシステムとの連携が必須になる。ネットワークで接続することも可能だが、運用基盤が分散することは通信遅延によるパフォーマンスの低下や、運用負荷の増加につながる。その解決策として、新規開発を行わない既存システムを無改修でクラウド環境上の仮想マシンに移行する(P2V)を行うことが必須になると考えられる。

さらにクラウドコンピューティング製品を活用するためには仮想マシンをクラウドコンピューティング製品で管理可能な「テンプレート」に変換し、新規仮想マシンと同じ方法で利用できる状況を実現する必要が発生する。

本検証では、実際の物理マシンに Redhat Enterprise Linux 6.6 および Windows server 2012 Standard をインストールし、その OS イメージを複数の手段で仮想マシンイメージ化し Redhat OpenStack+Redhat KVM で構成したサーバ上に移行を行うことで、物理マシンから OpenStack テンプレートへの移行のフィジビリティの評価を行っている。

物理マシンから Redhat OpenStack & KVM 上への移行検証

2 検証環境

2.1 検証に使用した機材

No	名称	機材	用途
1	Windows 物理マシン	Dell PowerEdge R430	移行元となる Windows サーバ
2	Linux 物理マシン	Dell PowerEdge R430	移行元となる Linux サーバ
3	KVM サーバ(1)	Dell PowerEdge R430	移行およびカスタマイズ用の KVM サーバ
4	KVM サーバ(2)	Dell PowerEdge R430	同上
5	ESXi サーバ	Dell PowerEdge R430	移行およびカスタマイズ用の ESXi サーバ
6	OpenStack サーバ	Dell PowerEdge R430	移行先の OpenStack+KVM サーバ

2.2 検証に使用した OS およびアプリケーション

No	名称	OS	アプリケーション
1	Windows 物理マシン	Windows server 2012 STD x86_64 (評価版)	OTRS 3.3 IIS 8.0 SQLServer 2014 Express
2	Linux 物理マシン	Redhat Enterprise Linux 6.6 x86_64	AIPO 7 PostgreSQL 8.3
3	KVM サーバ(1)	Redhat Enterprise Linux 6.6 x86_64	Redhat KVM
4	KVM サーバ(2)	Redhat Enterprise Linux 6.6 x86_64	Redhat KVM
5	ESXi サーバ	VMware ESXi 5.5 (評価版)	
6	OpenStack サーバ	Redhat Enterprise Linux 7.0 x86_64	Redhat OpenStack (Icehouse) Redhat KVM

2.3 IP アドレス

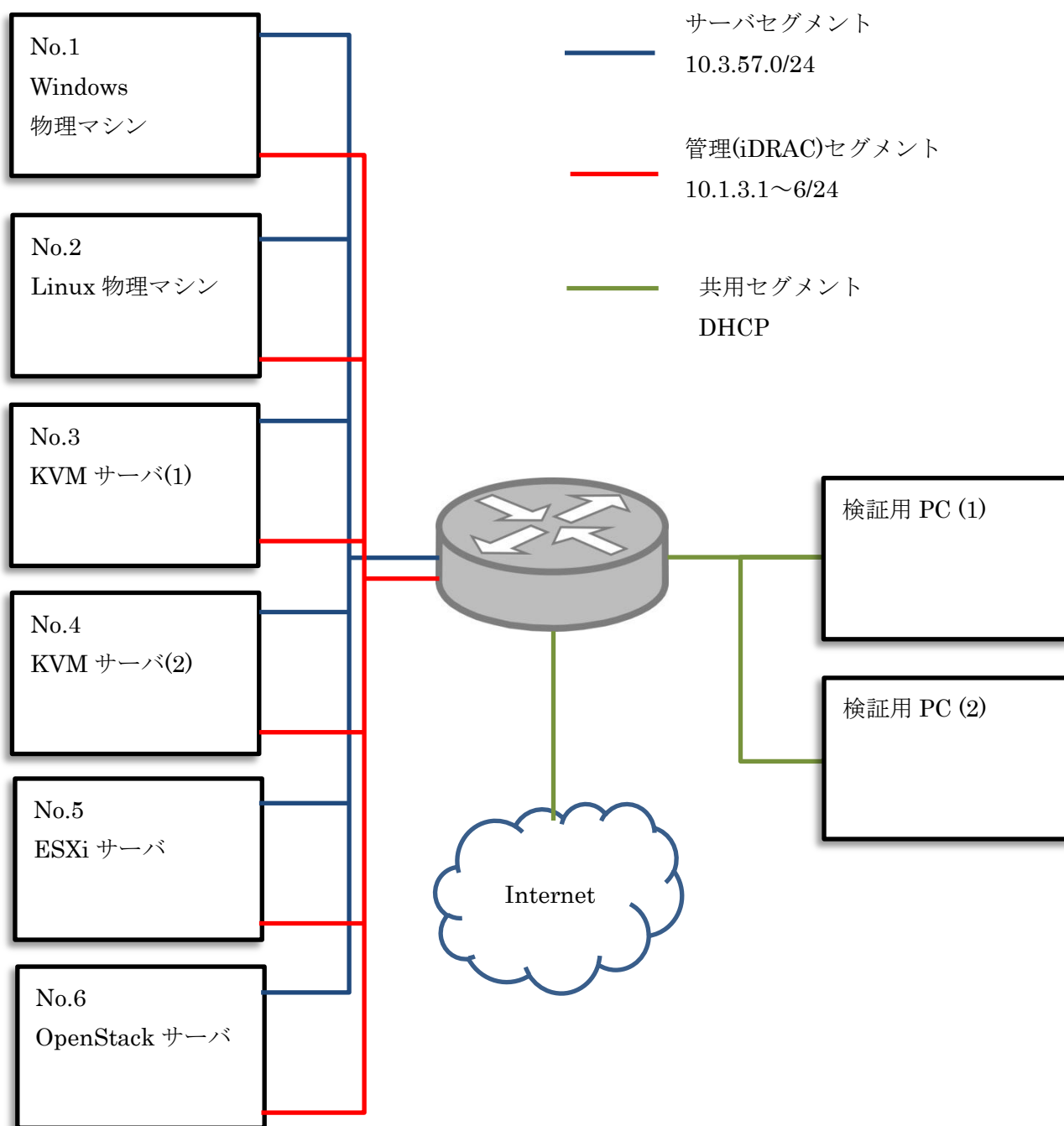
No	名称	サーバ OS	iDRAC8
1	Windows 物理マシン	10.3.57.1	10.1.3.1
2	Linux 物理マシン	10.3.57.2	10.1.3.2
3	KVM サーバ(1)	10.3.57.3	10.1.3.3
4	KVM サーバ(2)	10.3.57.4	10.1.3.4
5	ESXi サーバ	10.3.57.5	10.1.3.5
6	OpenStack サーバ	10.3.57.6	10.1.3.6

物理マシンから Redhat OpenStack & KVM 上への移行検証

2.4 ネットワーク構成

実際に検証で使用した、環境のネットワーク構成は以下となる。

- ①各サーバはサーバセグメントにフラットなネットワークで接続されている。
- ②管理セグメントは独立したネットワークに接続されており、サーバセグメントとは分離されている。
- ③共用セグメントは DHCP より IP が付与される。共用セグメントからはサーバセグメント、管理セグメントの双方に通信が可能
- ④全てのセグメントからインターネットへの接続が可能



物理マシンから Redhat OpenStack & KVM 上への移行検証

3 検証方法

3.1 物理マシン→仮想マシンへの変換(P2V)

物理マシンから直接、OpenStack イメージへの変換は行えないため、一旦、仮想マシンに変換する必要が発生する。検証は以下の2種類の方式で行った。方式およびメリット・デメリットは下表のとおりとなる。

No	方式	詳細	メリット	デメリット
1	Cold Cloning	CD 起動 OS(Live CD) を利用してサーバを起動し、サーバ内の OS が完全に停止した状態のイメージを取得する	OS を停止状態で取得できるため、ファイルやデータの不整合が発生しない	使用するハードディスクを全てイメージ化する必要があり、処理に時間がかかる
2	Hot Cloning	OS 内にコンバートツールをインストールし、サーバ内の OS イメージを取得する	OS 内でマウントされているディスクイメージの形で取得でき、未使用領域の排除も行える為、処理が早い	稼働している状態での取得となるため、ディスクの更新が行われ、データの不整合が発生する可能性がある

3.2 Cloud Cloning

物理マシン上に光学デバイス又は USB デバイスを接続し、そのデバイスにインストールされた OS を起動させることができる。このような OS は光学デバイスから起動されることが多いため、Live CD と呼称されることが多い。物理マシン上にインストールした OS とは別の OS を起動してハードディスクにアクセスすることが可能である。この場合、物理マシン上の OS は完全に停止状態であり、ハードディスク上のファイルやデータの更新は行われない。その為、データの不整合は発生せず、完全なデータを取得することが可能である。

3.2.1 Live CD の選定

LiveCD は、Linux OS の各種ディストリビューションからも提供されているが、今回は Knoppix(<http://www.risec.aist.go.jp/project/knoppix/>)を採用した。採用の理由は以下である。

- ①アレイコントローラのドライバとの互換性が高く、利用できるストレージが多い
- ②Partition Image(partimage)や Gparted(gparted)などのディスク管理ツールが標準で搭載されている
- ③GUI が標準で搭載されており、ターミナルを複数開くことにより、ヘルプからのカット&ペーストや複数タスクの実行などが行える。

物理マシンから Redhat OpenStack & KVM 上への移行検証

LiveCD が無い場合、RHEL のインストールメディアを **Rescue** モードで起動させることで Live CD の代替えとすることも可能であるが、②のツール類が存在せず、③の GUI も使用できないため利便性は少し劣る。

3.2.2 Cold Cloning を行うために必要な環境

Cold Cloning を実際に行うためには以下の環境を準備が必要となる。

①大容量 NAS

Cold Cloning では物理マシンのハードディスク上の状態をそのままの形で **RAW** 形式のイメージファイルに変換する。その為、使用されているデータ容量ではなく、最低でも移行を行う物理マシンに搭載されているハードディスク容量以上の領域が必要となる。

さらに取得したイメージファイルを各種ハイパーバイザの仮想マシンイメージに変換し、その後に **OpenStack Glance** で **OpenStack** 上に登録するイメージに変換する必要が発生する。

実際に必要となる容量は、物理マシン搭載ストレージの 3 倍程度となる。

LiveCD からマウントができ、かつ、各種ハイパーバイザから **Datastore** として利用するためには **NFS** として接続可能なストレージを準備することが望ましい。

②動作検証用ハイパーバイザ

取得を行った **RAW** 形式のイメージファイルを使用して仮想マシンとして起動を行えるか検証を行う為、仮想マシン実行環境が必要となる。**RAW** データ形式であれば、**KVM**、**VMware** ともに標準で稼働させることが可能である。

稼働を確認した後に、**OpenStack** に登録できる状態にカスタマイズする為、**cloud_init** のインストールおよび **virt-sysprep** によるネットワーク情報のクリーンアップ等の作業が必要となる。各種コマンドの実行を行うには、**Linux+KVM** 環境が現状では最適となっている。

③動作検証用サーバ

前述②のハイパーバイザを稼働させ、さらに仮想マシンイメージの稼働をカスタマイズ作業を実施するサーバが必要となる。**CPU**、**メモリ**等は最低でも仮想マシン動作分+ハイパーバイザ動作分を賄えるだけのリソース量が必要となる。

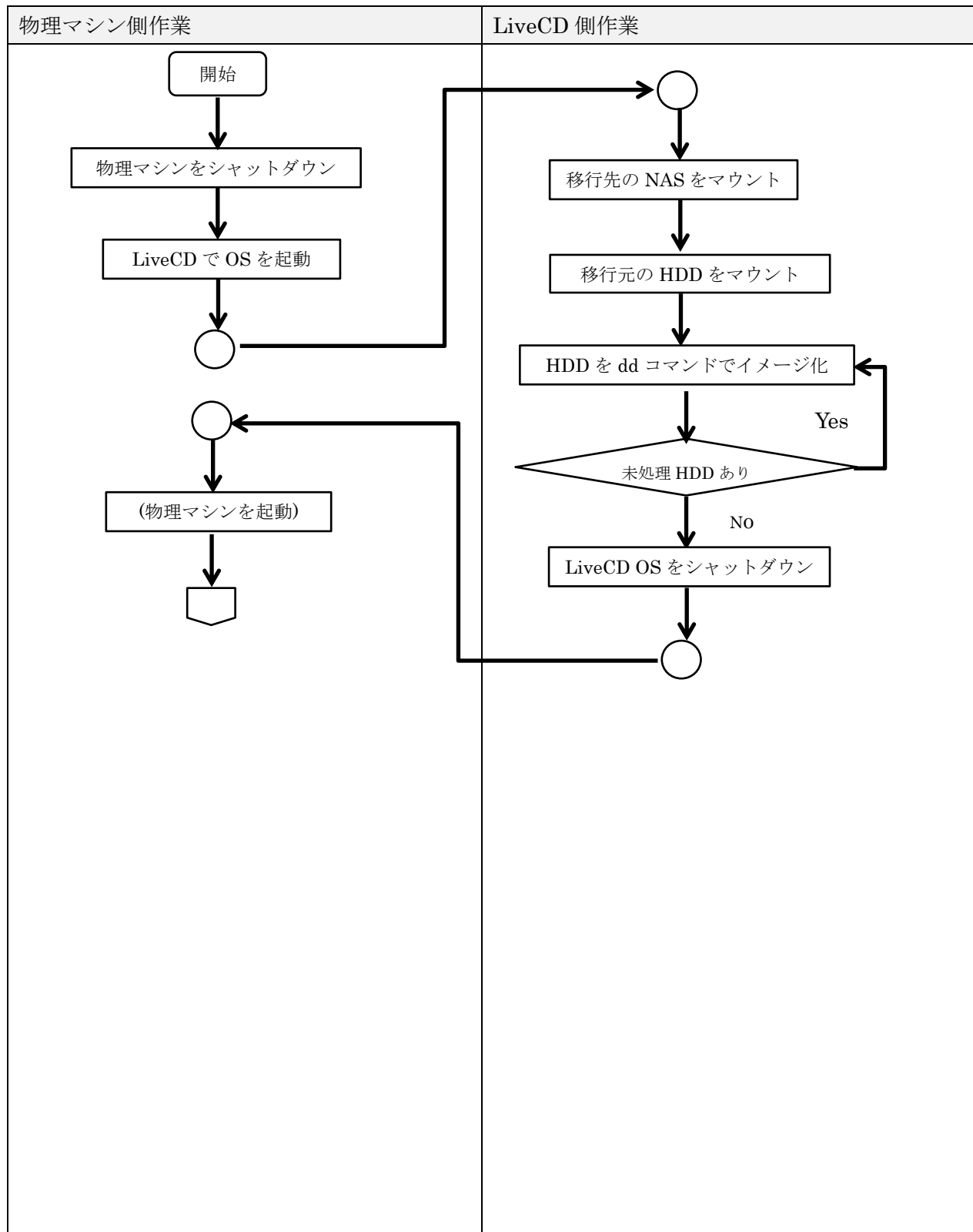
④物理マシン上の OS のインストールメディア

仮想マシンへの変換を行った場合、**MBR** の破損やハードディスクのデバイス名が変更されることなどにより、起動できなくなる場合が存在する。その場合は物理マシン上にインストールされた **OS** のインストールメディアを使用し、「修復セットアップ」や「**MBR** の上書き」や「**OS** 内部の設定ファイルの変更」等を行うことで復旧を行う必要が発生する。

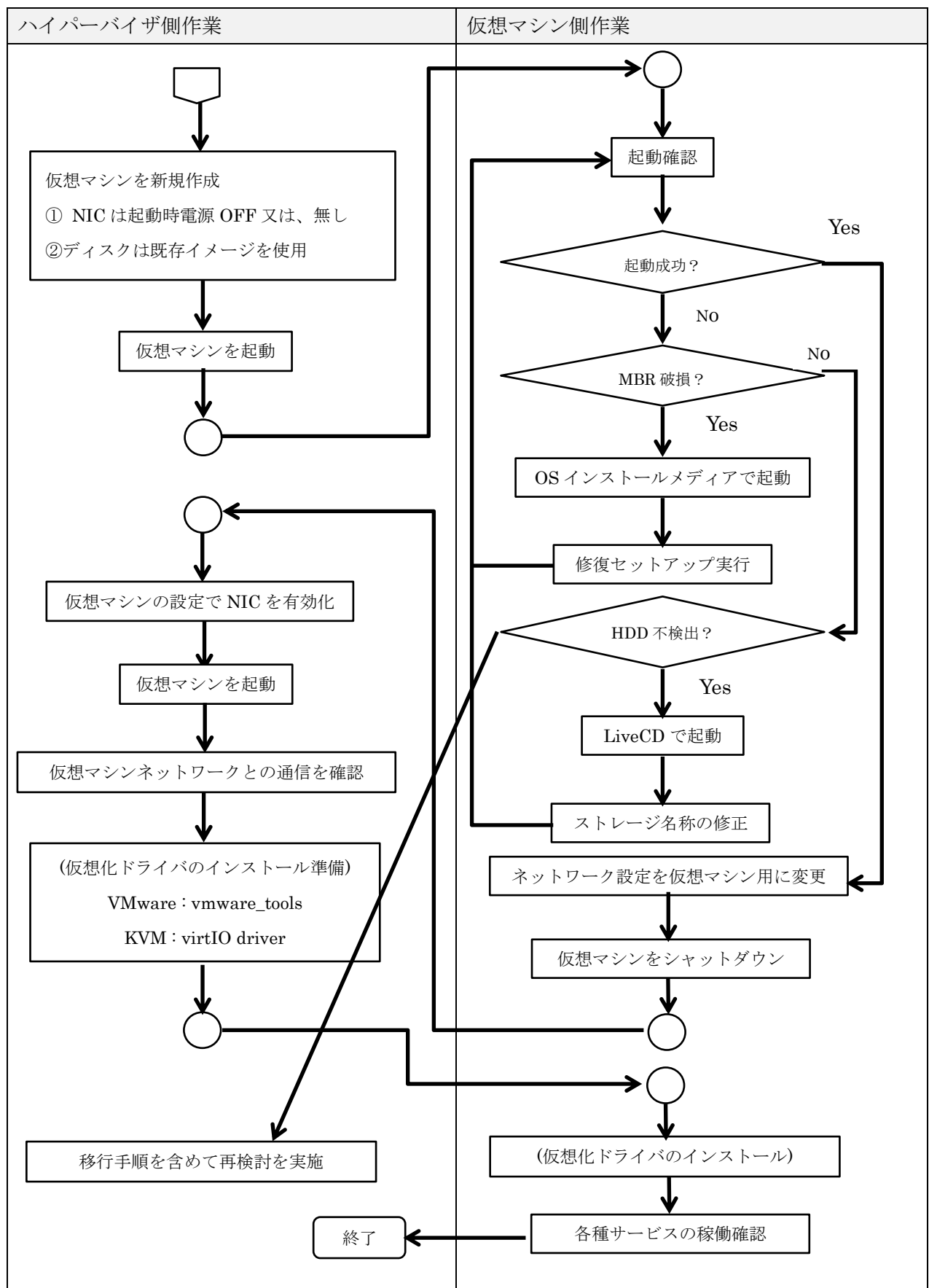
物理マシンから Redhat OpenStack & KVM 上への移行検証

3.2.3 Cold Cloning 作業の流れ

Cold Cloning 方式の移行の流れは以下になる。



物理マシンから Redhat OpenStack & KVM 上への移行検証



物理マシンから Redhat OpenStack & KVM 上への移行検証

3.2.4 Cold Cloning 作業の注意点

①データ破壊のリスクが存在

`dd` コマンドを利用することで、ハードディスクイメージをファイルに出力することもできるが、逆にハードディスクをファイルで上書きすることも可能である。バックアップの専用ツールではなく基本コマンドであるため、コマンドパラメータを間違えた場合の安全機構は全く存在しない。コマンド実行の際は十分に注意を払うことと共に、実行前に別の手段でバックアップを取得しておくことを推奨する。

②ハードウェア Raid

ハードウェア Raid コントローラを使用している場合、複数のハードディスクが使用されていても、OS からは Raid で構成された領域が 1 個のハードディスクとして認識されている為、実際のハードディスク構成を意識する必要は無く、OS から認識できるディスク構成でイメージ化を行う。ただし、アレイコントローラのドライバが LiveCD 上に実装されていない場合は、アレイが認識できない可能性がある。ディストリビューションによって対応に差異があるため、事前に認識可否を確認しておく必要がある。OS 標準でドライバが無くアレイが認識できない場合、USB メモリに OS をインストールしてアレイコントローラのドライバを組み込むことで、アレイに対応する LiveCD を自作することで対応を行う必要がある。

③ソフトウェア Raid

前述のハードウェア Raid とは異なり、LiveCD からはハードディスク 1 台ずつが Raid 構成無しで存在する形で認識される。ハードディスクイメージファイル取得の際は Raid で使用している全てのハードディスクのイメージを取得し、仮想マシンに構成する際も同じ順番で認識されるように構成する必要がある。

④NAS (iSCSI/FC/NFS/Samba)

実際の構成では、物理マシン上のストレージを直接ではなく、NAS 上から一部領域を提供して使用していることが非常に多い。ハードウェア接続の iSCSI、FC などはハードウェアとして直接 NAS に接続されており、OS 上からは SCSI コントローラ+ハードディスクと認識される。LiveCD で認識が行えれば、ディスクアレイに接続されているハードディスクと全く方法でイメージの取得が可能だが、デバイスドライバが必須となった場合、前述のハードウェア Raid コントローラと同様に USB 起動の対応を行う必要が発生する。

ソフトウェア iSCSI の場合、LiveCD で起動した OS 上にソフトウェア iSCSI を構成し、物理マシン上の OS と同一の IP アドレス、イニシエータ名を設定することにより接続が可能となる。接続後は物理マシン上に直接接続されているハードディスクと同様にイメージの取得が可能である。

物理マシンから Redhat OpenStack & KVM 上への移行検証

NFS/Samba の場合は、ディスクイメージとして取得するのには適さない。複数サーバから共用するネットワークストレージとして利用するのであれば、NFS/Samba ファイルサーバとして単体で移行を行うか、ファイルシステムを tar ボールで固めて別に移行する形を推奨する。

⑤ハードディスク容量

Cold Cloning では前述のとおり、ハードディスクの全領域をデータの有無にかかわらずイメージとして取得することとなる。その為、ハードディスクの容量が大きい場合、長時間サーバを停止させる必要が発生する。発生する時間のほぼすべてが物理マシン上のハードディスクからの読み込み時間と移行用ストレージへの転送＋書き込み時間である。停止時間を短縮するためには以下の対策が考えられる。

- (1) 物理マシンと移行用ストレージを高速デバイス(e-SATA 又は USB 3.0)で直接接続する。
- (2) 転送用ストレージに NAS を使用する場合は、物理マシンと NAS 間を可能な限り高速なネットワークで接続する。
- (3) 書き込み、転送速度を確保するため、移行作業は並行で行わない。

⑥仮想マシン上での HDD のデバイス名

物理マシン上では（特に HP 製の Raid コントローラを使用していた場合）認識されているハードディスクのデバイス名が論理デバイス名となるため、ベンダー特有なものになることがある。

(例) HP 製サーバのハードディスクのデバイス

ハードディスク 1 台全体 : /dev/cciss/c0d0

パーティション : /dev/cciss/c0d0p1, /dev/cciss/c0d0p2 …

移行したデータを KVM 上で起動を行う場合には、以下のデバイス名で認識が行われる

ハードディスク 1 台全体 : /dev/sda

パーティション : /dev/sda1, /dev/sda2 …

その為、OS 上に設定されたハードディスクデバイス名と実際のデバイス名が異なるため、起動時に指定されたハードディスクが認識できず、起動が行えなくなってしまう状況が発生する。

Linux の場合、/boot/grub2/device.map と /etc/fstab の変更を行うことで復旧が可能である。fstab のデバイス名に UUID が使用されている場合は UUID の部分を直接指定 (/dev/sda1 を指定) することで対応が可能となる。正常起動後は UUID が採番されているので、UUID= の形に戻しておくことを推奨する。

Windows の場合、本件事象は発生しなかった為、OS 側で対応を行っているものと考えられる。

物理マシンから Redhat OpenStack & KVM 上への移行検証

⑦仮想化ドライバのインストール

VMware および KVM の仮想化環境上でのサーバのディスクやネットワークの I/O の高速化を図るために各ベンダーより仮想化ドライバが提供されている。VMware からは `vmware_tools` が提供され、I/O の高速化以外にもハイパーバイザからの仮想マシンの制御や監視にも利用されている。KVM からは Windows OS 向けに VirtIO ドライバが提供されている。仮想マシンとして移行したハイパーバイザ上で継続利用をするのであれば、インストールが必須となる。

⑧仮想マシン上でのネットワークインターフェース

Redhat 系 Linux の RHEL 6.1/CentOS6.1/Fedora 15 以降の場合、前述の HDD のデバイス名以外にもネットワークインターフェースの設定ファイルの名称が物理デバイスの接続方法ごとに変更される仕様となっている。

(例) NEC サーバに Inter 製 NIC が PCI Express 上で接続されている場合

```
ifcfg-enp13s0,ifcfg-enp14s0 ...
```

(例) Dell 製サーバに Bloadcom 製の NIC がオンボードで搭載されている場合

```
ifcfg-em0,ifcfg-em1 ...
```

ハイパーバイザは基本的に仮想 NIC の設定ファイルは従来通り `ifcfg_eth0`、`ifcfg_eth1`…であるため、物理デバイスで命名された設定ファイルの内容が有効にならない。仮想マシンとして起動する時点で、仮想 NIC を接続する為、その設定に合わせた MAC アドレスや UUID も変更する必要があるため、設定ファイルは新規で作成する必要が発生する。

⑨OS、プロダクトのライセンス

物理マシンから仮想マシンに移行を行う場合、移行を完了し、切替が終了するまで、並行稼働を行う可能性が高いと考えられる。その場合は、有償製品であれば、仮想マシンで稼働させるライセンスを追加購入し、移行後速やかにライセンスの切替を行う必要が発生する。移行を行う場合は事前にライセンスの調達を行い、移行時点で速やかにアクティベートを行う必要がある。

仮想マシン移行時点で物理マシンを完全に停止するのであれば、ライセンスの増加は発生しない。ただし、OS のライセンスが物理マシンにバンドリング(OEM ライセンス)されているものの場合は、同一物理マシン以外で動作させることはライセンス的に違反となる。

さらに Oracle や MicroSoft の RDB 製品などのライセンスは「仮想化ソフトウェア上で製品を動作させる場合、稼働する物理マシンの全プロセッサがライセンスカウントの対象」となっている製品が存在する。

各 OS、プロダクトのライセンス条件については移行前に確認を行い、必要なライセンスを事前に準備しておくことが必要である。

物理マシンから Redhat OpenStack & KVM 上への移行検証

3.2.5 Hot Cloning を行うために必要な環境

Hot Cloning を実際に行うためには以下の環境の準備が必要となる。

①大容量 NAS

理由、用途については Cold Cloning と同様の為、割愛とする。

②動作検証用ハイパーバイザ

物理マシンから移行を行った VMware 仮想マシンとして起動を行えるか検証を行う為の実行環境となる。VMware 環境であれば、ESXi、Server、Player のいずれでも利用可能であるが、本番環境と同一のハイパーバイザにしておくのが望ましい。

③動作検証用サーバ

前述②のハイパーバイザを稼働させ、さらに仮想マシンイメージの稼働をカスタマイズ作業を実施するサーバが必要となる。CPU、メモリ等は最低でも仮想マシン動作分+ハイパーバイザ動作分を賄えるだけのリソース量が必要となる。

④VMware vCenter Converter 端末

VMwareConverter を稼働させる Windows 端末。OS は Windows Vista 以降のクライアント OS でも Windows 2003 以降のサーバ OS のいずれでも利用可能。物理マシンではなく、仮想マシンでも動作するため、前述②の VMwara サーバ上に構築することも可能

⑤移行用のネットワーク

物理マシン、VMware サーバ、vCenter Converter 端末、はそれぞれが通信可能な状態のネットワークに接続する必要が発生する。物理マシンのハードディスクイメージの転送を行うため、NAS も含めて占有を行える高速ネットワークに接続する形が望ましい。

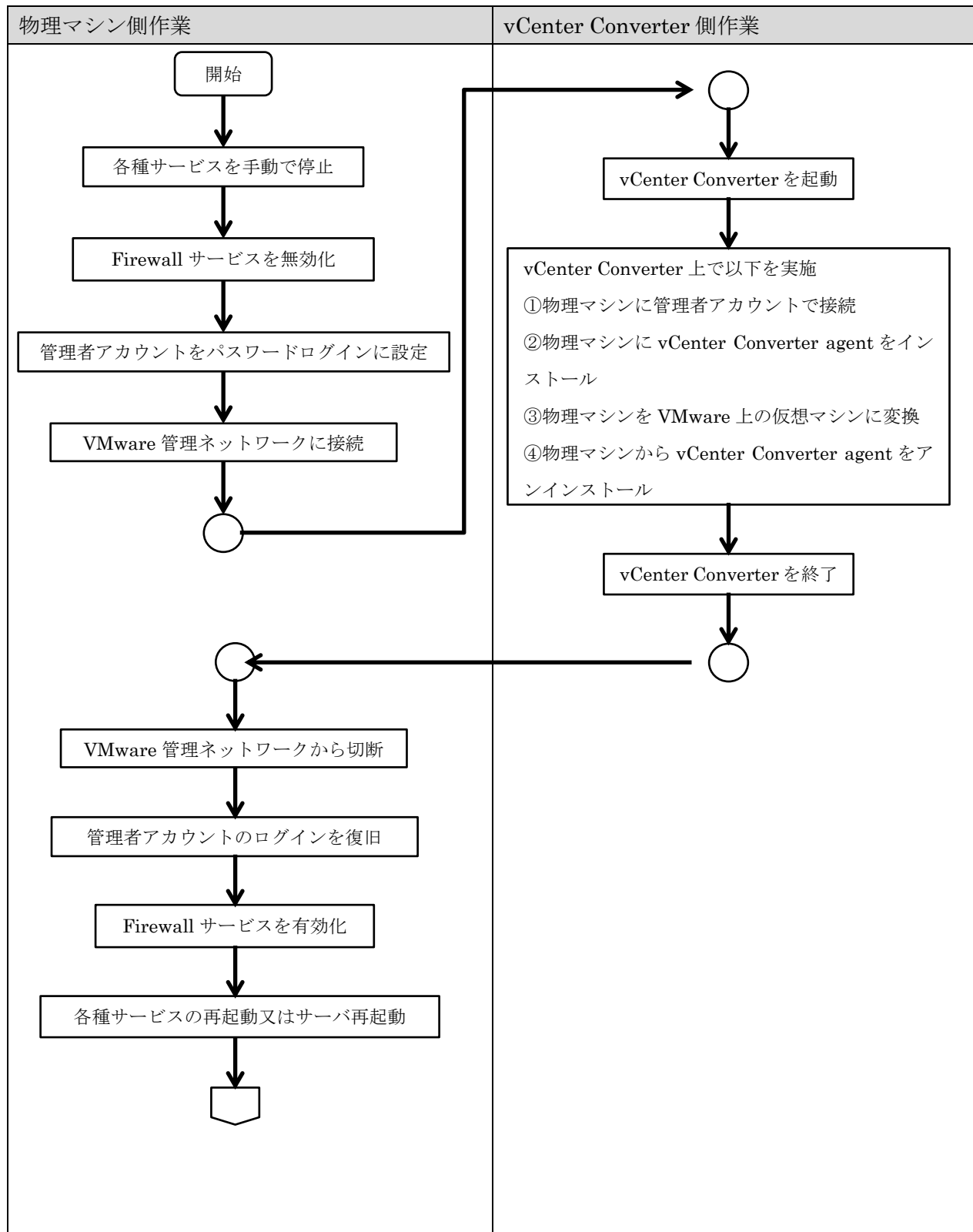
⑥物理マシン上の OS のインストールメディア

理由、用途については Cold Cloning と同様の為、割愛とする。

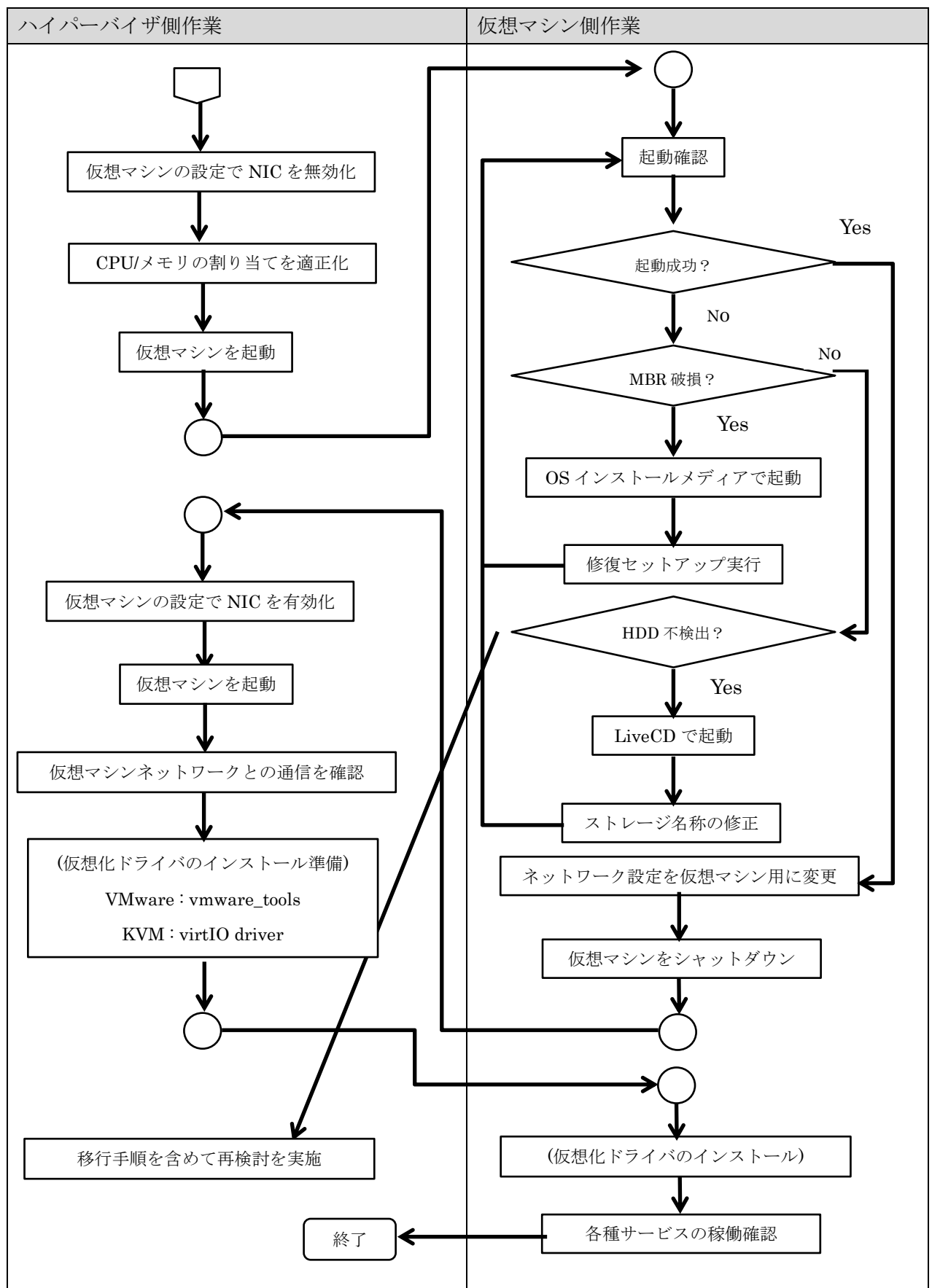
物理マシンから Redhat OpenStack & KVM 上への移行検証

3.2.6 Hot Cloning 作業の流れ

Hot Cloning 方式の移行の流れは以下になる。



物理マシンから Redhat OpenStack & KVM 上への移行検証



物理マシンから Redhat OpenStack & KVM 上への移行検証

3.2.7 Hot Cloning 作業の注意点

①セキュリティ上のリスクが存在

Windows の場合、vCenter Converter agent をインストールするためには、通常ブロックされているサーバの管理者アカウントをリモートからパスワード認証で接続できるように変更する必要が発生する。さらにリモートからのインストールを実行するために、UAC も Firewall も無効にする必要が発生する。Linux の場合は vCenter Converter agent のインストールは行われないが、管理者アカウントでの外部からの SSH 接続の許可と Firewall/Selinux の無効化は同様である。

外部に公開されているサーバでは行ってはいけない対応であるため、サーバイメージ取得時は公開セグメントをオフラインとし、vCenter Converter を実行するセグメント以外を遮断することを推奨する。

②システムに変更が発生

vCenter Converter agent がインストールやセキュリティ設定の変更が発生するため、移行元となる物理マシンは本番稼働状態とは異なる状態となる。当然、移行先の仮想マシンにも設定は継承されるため、移行元、移行先ともに設定の復元および稼働確認が必須となる。

OS およびデータ破損の可能性も考慮し、事前にバックアップを取得することを推奨する。

③データ不整合のリスクが存在

実行中の OS から HDD イメージを取得するため、アプリケーションやサービスが稼働している状態では更新と取得タイミングのラグによりファイルやデータに不整合が発生する可能性がある。

vCenter Converter 実行前に Web、RDB などのサービスは必ず停止し、Windows update や yum-updatesd 等の自動更新サービスも停止する必要がある。

ウィルス対策ソフトもサービスを停止する必要があるが、製品によっては、一定時間で自動的にサービスを起動させてしまう製品も存在する。作業中は再起動がかからないように設定を行うか、不可能な場合はアンインストールも検討する必要がある。

④ハードウェア Raid

OS 上から認識されているストレージ情報のままで変換が可能である。OS からドライバも組み込まれているため、Cold Cloning のように特殊な起動ディスクの作成も不要である。

⑤ソフトウェア Raid

vCenter Converter ではソフトウェア Raid で構成されたシステムからは、正常にイメージを取得することが不可能である。Cold Cloning を選択するか、まったく別の方式を検討する等の対応が必要となる。

【ご参考】 VMware vCenter Converter Standalone 5.5.1 Release Notes

https://www.vmware.com/support/converter/doc/conv_sa_551_rel_notes.html

物理マシンから Redhat OpenStack & KVM 上への移行検証

⑥NAS (iSCSI/FC/NFS/Samba)

iSCSI/FC は OS 上から認識されているストレージ情報のままで変換が可能である。OS からドライバも組み込まれているため、Cold Cloning のように特殊な起動ディスクの作成も不要である。

NFS/Samba については Cold Cloning と同様である。

⑦ハードディスク容量

大容量ハードディスクを持つ物理マシンを移行する場合の懸念点は Cold Cloning と同様であるが、vCenter Converter はデータの存在する領域のみを移行する性能対策が行われており、Cold Cloning と比較するとはるかに高速で処理が可能である。生成されるディスクイメージは VMware の vmdk 形式となっており、Thin provisioning 形式を利用すれば、仮想化環境のストレージ利用量の削減も可能である。(性能は若干劣化するので注意が必要)

⑧仮想マシン上での HDD のデバイス名

vCenter Converter で移行を行う場合、OS 内のデバイス名の設定の書き換えも合わせて行われる。Cold Cloning と同様の問題は発生しない。ただし、vCenter Converter を使用して Linux 物理マシンの変換を行うと、"/boot"と "/" の領域が別のイメージファイルに分割されてしまう。

【変換前：Linux 物理マシン上のディスク構成】

/dev/sda		
/dev/sda1	/dev/sda2	
500MB	95.5GB	
Linux	LVM	
ext4	swap	ext4
/boot	tmpfs (/dev/mapper/VolGroup-lv_swap)	/ (/dev/mapper/VolGroup-lv_root)

【変換後：vCenter Converter 変換後の仮想マシンのディスク構成】

/dev/sda	/dev/sdb	
/dev/sda1	/dev/sdb1	
500MB	5GB	94.5GB
Linux	LVM	
ext4	swap	ext4
/boot	tmpfs	/ (/dev/mapper/VolGroup-lv_root)

物理マシンから Redhat OpenStack & KVM 上への移行検証

本事象は VMware 上で仮想マシンを動作させる点において全く問題はないが、変換した仮想マシンイメージを OpenStack の仮想マシンイメージに利用する場合には大きな問題となる。Windows に関しては本事象は発生しない。

⑨仮想化ドライバのインストール

Cold Cloning と同様の対応が必要。

⑩仮想マシン上でのネットワークインターフェース

Cold Cloning と同様の問題が発生するため、対応が必要。

⑪OS、プロダクトのライセンス

Cold Cloning と同様の問題が発生するため、対応が必要。

物理マシンから Redhat OpenStack & KVM 上への移行検証

4 OpenStack の仮想マシンイメージ (Cloud image)

OpenStack 上で仮想マシンを作成するためには、Cloud image と呼ばれる OpenStack で管理できる形にカスタマイズした仮想マシンイメージを OpenStack 上に登録する必要がある。Cloud image の入手方法は以下の2つからの選択となる。

- ①各ディストリビュータが公開している Cloud image をダウンロードして登録する。
- ②インストールメディアを利用して Cloud image を自作する。

本検証は物理マシンからの変換を行った仮想マシンをベースにカスタマイズを行い、OpenStack で Cloud image として利用可能な状態を実現することを目的としている。まずは、通常の物理マシン・仮想マシンと OpenStack で利用する Cloud image 形式の差異を理解する必要がある。

4.1 Cloud image の特徴

Cloud image には通常の OS のインストール状態とは異なる以下の特徴がある。

- ①cloud-init がインストールされており、初期構築時にサーバの各種設定を外部パラメータで行う機能が実装されている。
- ②初期構築を行う為の管理者権限を持つユーザアカウントが登録されている。
- ③ネットワークデバイスが接続されておらず、設定も消去されている。
- ④ハードディスクは1台のみで構成されている。
- ⑤Swap 領域は持たない。
- ⑥OS ごとにハードディスクのパーティショニングやファイルシステムの形式に制限が存在する。

Cloud image の登録を行う Glance の機能では複数のハードディスクイメージは登録できない。また、OpenStack 仮想マシンイメージの標準以外の形式で領域を作成している場合、cloud-init をフレーバーの設定サイズまでディスクを自動的に拡張する機能が正常に動作しない問題も発生する。

正常に動作している場合は、Linux であれば "/" の領域、Windows であれば"C:\¥" の領域がフレーバーで指定したサイズまで初期構築時に拡張されて提供される。ただし、自動拡張が行われなくても、仮想マシンイメージのファイル自体は拡張されているため、OS にログインしてコマンドで拡張することは可能だが、OpenStack の利便性の一部が損なわれる状態になる。

OS ごとの特徴と標準インストールした場合の仮想マシンとのを以下にまとめる。

物理マシンから Redhat OpenStack & KVM 上への移行検証

4.2 Linux の Cloud image

Linux を標準インストールした物理マシンのハードディスクの構成と同容量で構成された OpenStack で利用可能な仮想マシンイメージは以下となる。

(例) 100GB のハードディスク 2 台を使用して Linux をインストールした場合

4.2.1 物理マシン上のディスク構成

物理マシンに現行の Linux を標準 (自動設定) でインストールすると以下のような構成になる。

① /boot が先頭の独立領域で作成される。

② 次に swap 領域が作成される。

③ 1 台目のハードディスクの残り空き領域 + 2 台目の空き領域を LVM で確保し連続領域を作成。

※ 大容量ディスクの場合、"/" 構成したあとの余りが全て "/home" として別領域に作成される。

/dev/sda		/dev/sdb	
/dev/sda1	dev/sda2		/dev/sdb1
500MB	5GB	94.5GB	100GB
Linux	LVM		
ext4	swap	ext4	
/boot	tmpfs	/ (/dev/mapper/VolGroup-lv_root)	

4.2.2 Cloud image のディスク構成

Cloud image の標準のディスク構成は以下となる。

① ハードディスクは 1 台で領域は 1 つのみ。

② 領域は Linux で ext3 or ext4 でフォーマットされている。(LVM 使用不可)

③ swap 領域は無し

/dev/sda
/dev/sda1
200GB
Linux
ext4
/

物理マシンから Redhat OpenStack & KVM 上への移行検証

4.3 Windows の Cloud image の特徴

Windows server(2008 以降)を標準インストールした物理マシンのハードディスクの構成と同容量で構成された OpenStack で利用可能な仮想マシンイメージは以下となる。

(例) 100GB のハードディスク 2 台を使用して Windows server をインストールした場合

4.3.1 物理マシン上のディスク構成

物理マシンに現行の Windows server を標準でインストールすると以下のような構成になる。

- ①/boot が先頭の独立領域で作成される。
- ② 1 台目のハードディスクの残り空き領域が 1 ボリュームとなり c:¥に割り当てられる。
※任意に変更可能
- ③swap 領域は c:¥内にファイルで作成される。
- ④ 2 台目インストール時は特に使用されない。

※インストール時に領域確保を行えば、いずれかのドライブにマウントされる

/dev/sda		/dev/sdb
/dev/sda1	/dev/sda2	-
100MB	99.9GB	100GB
NTFS	NTFS	-
/boot	c:¥	-

4.3.2 Cloud image のディスク構成

Cloud image の標準のディスク構成は以下となる。

- ①ハードディスクは 1 台で領域は/boot と c:¥の 2 つ。
- ②領域は NTFS でフォーマットされている。
- ③swap 領域が別途提供されるため、OS 中の swap ファイルは存在しない。

/dev/sda	
/dev/sda1	/dev/sda2
100MB	99.9GB
NTFS	NTFS
/boot	c:¥

物理マシンから Redhat OpenStack & KVM 上への移行検証

4.4 物理マシンイメージから Cloud image への変換手順 (案)

前述のとおり、実際の物理マシンのディスク構成と Cloud image には大きな乖離があり、P2V のように「ツールの実行のみでのコンバート」は残念ながら行えない。物理マシンイメージから Cloud image に変換を行う為には、以下の作業を行う必要が発生する。

4.4.1 Linux の場合

Linux の場合は、使用しているハードディスクの台数やファイルシステムの構成にかかわらず、ファイルシステムのパスは "/" からとなる。そのため、Cloud image 用の仮想マシンを新たに構築してファイルのみを tar で取得して展開することで Cloud image に移行することが可能である。

実際に行う作業手順は以下となる。

①Cloud image 用仮想マシンを移行元サーバと同一 OS で構築

移行対象のサーバを構築したものと同一インストールメディアを使用し、仮想マシンを新規構築する。その際の構築パラメータは以下となる。

- (1)ディスクイメージ形式 : qcow2
- (2)アレイコントローラ : VirtIO
- (3)ハードディスク台数 : 1 台
- (4)ハードディスクサイズ : 移行元で使用しているディスクサイズ + 若干の予備
- (5)NIC
- (6)インストールモード : minimal
- (7)パーティション構成
 - パーティション : 通常パーティション × 1
 - マウントポイント : "/"
 - フォーマット : ext4 or ext3

②Cloud image 用の仮想マシンをインストールメディアの Rescue モードで起動

① で作成した仮想マシンにインストールメディアの CD を再度マウントした状態で起動し、インストーラを起動させる。起動したらメニュー画面の "Rescue installed system" を選択する。

物理マシンから Redhat OpenStack & KVM 上への移行検証

③ /boot と /etc/fstab 以外を削除

- (1) Rescue モードで起動すると、仮想マシンイメージ内のディスクが /mnt/sysimage 以下に読み取り専用でマウントされている。
- (2) `mount -o rw,remount /mnt/sysimage` で書き込み可能状態に再マウントする。
- (3) /boot と /etc/fstab を残し、それ以外のファイルを全て削除する。
- (4) シャットダウンし、Rescue モードを終了する。

④ 移行元サーバをシャットダウン

移行を行う移行元のサーバをシャットダウンする。

⑤ 移行元サーバをインストールメディアの Rescue モードで起動

- (1) インストールメディアを CD ドライブにマウントして移行元サーバを起動させ、インストーラを起動させる。起動したらメニュー画面の "Rescue installed system" を選択する
- (2) 起動したら、/mnt/sysimage に "/" 以下のボリュームがマウントされていることを確認する。

⑥ データバックアップ用のストレージをマウント

- (1) データバックアップ用のストレージを Rescue モードで起動した OS にマウントする。
- (2) 移行元が物理サーバの場合、USB 外付け HDD 等のリムーバブルデバイスを接続する。
- (3) 移行元が仮想サーバの場合、仮想マシンイメージファイルを新たに作成しマウントする。
※①～②した仮想マシンイメージのディスクでもよい。

⑦ 移行元マシンの "/" 以下を⑥で接続したストレージにバックアップ

- (1) `cd /mnt/sysimage` でバックアップ取得ディレクトリに移動する
- (2) 以下のコマンドで移行対象ファイルをバックアップする。
`# tar zpcf [バックアップデバイス]/backup.tar.gz .* --exclude ./boot --exclude ./mnt/fstab`
- (3) ①～②の仮想マシンイメージの "/" を直接マウントした場合は、バックアップの取得と解凍を
で一気にできる。

```
# tar cpf - .* -exclude /boot --exclude /mnt/fstab | (cd [移行先の "/"]; tar xpvf -)
```

本処理を実行した場合、処理⑬までスキップできる。

- (4) 作成した tar ファイルは "iso イメージファイル" か "仮想ハードディスクイメージファイル" に格納する。

Linux 上で tar ファイルをフォルダに格納し以下のコマンドで iso イメージファイルが作成可能。

```
# mkisofs -r -J -o [iso イメージファイル名] [tar ファイル格納フォルダ]
```

物理マシンから Redhat OpenStack & KVM 上への移行検証

⑧ ① で作成した Cloud image 用の仮想マシンにインストールメディアと tar フォルダを格納した仮想イメージファイルをマウントして Rescue モードで起動する。

iso イメージなら仮想 CD デバイスを、ハードディスクイメージなら仮想ハードディスクデバイスを追加する。

⑨ Rescue モードで起動した OS に Cloud image 用の仮想マシンの "/" とバックアップイメージのデバイスがマウントされていることを確認

(1)ハードディスクは基本的には/mnt 以下にそれぞれのファイルシステムでマウントされている。

(2)マウントされていない場合は /mnt 以下にフォルダを作成しマウントを実行

```
# mkdir /mnt/sysimage /mnt/backup
```

```
# mount /dev/vda1 /mnt/sysimage
```

```
# mount /dev/vdb1 /mnt/backup
```

(3)CD の場合は以下のコマンドでマウント

```
# mount -o loop /dev/sr1 /mnt/backup
```

⑩ ⑦ で取得したバックアップを Cloud image 用の仮想マシンの "/" に展開

以下のコマンドで移行元マシンから取得したファイルを展開

```
# tar xzpvf /mnt/backup/backup.tar.gz /mnt/sysimage
```

⑪ Rescue モードで起動した OS をシャットダウン

⑫インストールメディアとバックアップストレージを Cloud image 用の仮想マシンから取り外す。

⑬Cloud image 用の仮想マシンを起動する。

⑭システムが正常に稼働していること、データが破損していないことを確認

以上に作業で Cloud image にカスタマイズ可能な仮想マシンへの移行が行える。

本方式で、移行が難しかったソフトウェア Raid や特殊なアレイコントローラを使用しているサーバも移行を行うことが理論上は可能と考えられる。

物理マシンから Redhat OpenStack & KVM 上への移行検証

4.4.2 Windows の場合

Windows の場合、ハードディスクが複数台存在する場合や D:¥以降の複数のパーティションが存在する場合、カスタマイズの作業が必要となる。実際に行われる変換は以下のイメージとなる。

【変換前】

ハードディスク×3台で構成。それぞれにドライブレター C:¥ / D:¥ / E:¥

用途はそれぞれ、C:¥ = OS / D:¥ = DB / E:¥ = log に使用する。

D:¥はデータが増加した場合、ディスクサイズの拡張を行いたい。

/dev/sda	
/dev/sda1	/dev/sda2
/boot	C:¥
boot-mgr	OS

/dev/sdb
/dev/sdb1
D:¥
DB

/dev/sdc
/dev/sdc1
E:¥
log

【変換後】

ハードディスク×1台にパーティションを作成し、データのコピーを実施

/dev/sda			
/dev/sda1	/dev/sda2	/dev/sda3	/dev/sda4
/boot	C:¥	E:¥	D:¥
boot-mgr	OS	log	DB

ハードディスクを1台に集約することで Cloud image としての登録が可能な状態となる。OpenStack のインスタンス起動時のディスク領域の拡張が終端領域に行われる。その為、拡張が行われる必要のある D:¥のパーティションを最後のパーティションに設定している。

物理マシンから Redhat OpenStack & KVM 上への移行検証

上記の例の場合、移行手順は以下となる。

- ①物理マシンの場合、一旦 P2V を行い、仮想マシンとして稼働させる。
- ②仮想マシン設定の変更で boot 領域と C:¥のパーティションを持つ仮想ディスクを拡張する。
- ③バックアップ用の一時領域の仮想ディスクを追加する。容量は D:¥+E:¥のバックアップを取得できる分だけ必要となる。
- ⑤仮想マシンを起動する。
- ⑥追加したバックアップ用のディスクをフォーマットしマウントする。
- ⑦Web server や RDB のサービスを停止する。
- ⑧Windows 2012 server 標準の Windows server バックアップで D:¥、E:¥をそれぞれ、バックアップ用のディスクにバックアップする。
- ⑨D:¥、E:¥のドライブレターを D:/E:以外に変更する。
- ⑩拡張した 1 台目のディスクの空き領域に E:¥、D:¥のパーティションを作成する。
- ⑪バックアップから E:¥、D:¥のデータを復元する。
- ⑪仮想マシンをシャットダウンする。
- ⑫仮想マシンの設定で 2 台目移行のハードディスクを取り外す。
- ⑬仮想マシンを起動し、稼働を確認する。

この作業では、ハードディスクの台数を 1 台に集約することで OpenStack への登録は可能となるが、フレーバーに対応したディスクの自動拡張は行なえない。Linux 同様にファイルレベルで C:¥にコピーを行うことも不可能ではないが、設定ファイルやレジストリのドライブレターやパスの記述を完全に修正することは「ほぼ不可能」だと考えられる。

4.5 物理マシンから OpenStack への移行の限界

Linux/Windows とともに P2V を行った仮想マシンイメージをそのまま、完全な Cloud image に変換することは制約が大きく、かなり困難な作業となる。ただし、cloud-init のディスク領域の自動拡張機能が動作しない以外は、正常に動作する。

この部分は手動で対応するなど、ある一定の妥協を許容できるのであれば、実現は可能である。

以降の章で実際に物理マシンから移行した仮想マシンイメージを使用して、OpenStack Glance での登録および OpenStack Nova でのデプロイを行う検証を行った結果とその手順について記載する。

物理マシンから Redhat OpenStack & KVM 上への移行検証

5 仮想マシンイメージから Cloud image への変換

OpenStack への Cloud image の登録は OpenStack Glance のコマンドを利用して行い、登録を行ったイメージは OpenStack Nova のコマンドを利用し、インスタンスとして起動することができる。物理マシンから移行を行う場合は、必ず一旦、仮想マシンに変換し、必要なカスタマイズを行った後に Cloud image に登録する必要が発生する。本書では実際に検証を行った結果を元に取りまとめた手順を記載する。

5.1 Glance で登録可能な Cloud image のファイル形式

Cloud image のファイル形式には以下の制約が存在する。

- ①ファイルは1個であること。(内部の領域は複数ボリュームに区切られていても登録は可能)
- ②仮想マシンとして起動可能であること。
- ③イメージファイルのデータ形式は以下であること。

形式	特徴
RAW 形式	物理ハードディスクをそのままの状態(生=RAW)で取得した仮想マシンイメージファイルの形式。未使用領域もそのままの形でイメージ化される為、サイズは大きい領域がすでに確保されているため、アクセス速度は速い。
qcow2 形式	QEMU を利用したハイパーバイザで利用される仮想マシンイメージファイル。実際にデータの存在する領域のみを保存したファイルとなるため、サイズが小さく、バックアップやコピーなどが容易だが、ディスク使用領域が増えるたびにファイルの拡張又は、差分ファイルの作成・拡張が行われるため、アクセス速度が若干遅い。
vmdk 形式(Thick)	VMware の仮想マシンイメージファイルの形式。RAW 形式と同様に未使用領域が確保されている為、同様の特徴を持つ。設定ファイルの[filename.vmdk]と実際のデータファイルの[filename-flat.vmdk]の2個のファイルで構成される。ファイルの形式にはバージョンが存在し、ファイル形式のバージョンによって使用できる VMware の製品やバージョンが異なる。
vmdk 形式(Thin)	VMware の仮想マシンイメージファイルの形式。qcow2 形式と同様に実際のデータが存在する領域のみが確保されている。ただし qcow2 とは異なり、追加データの差分ファイルでの保有は行えない。それ以外の特徴は vmdk 形式(Thick)と同様。
VHD 形式	VirtualPC 用に開発された仮想マシンイメージファイル。容量固定、容量可変(縮小も可)、差分更新にも対応。Hyper-V、Xen、VirtualBOX などハイパーバイザに幅広く対応。

物理マシンから Redhat OpenStack & KVM 上への移行検証

5.2 OpenStack で利用可能な Compute node

OpenStack では compute node として KVM、Xen、VMware ESXi、Microsoft Hyper-V が利用可能である。Glance で登録した形式がどの形式であっても、Nova が各 Compute node に対応した形式にデータ形式を変換する為、インスタンスの構築が可能である。

5.3 QEMU ディスクイメージ管理ツール (qemu-img)

QEMU を利用するハイパーバイザの仮想マシンイメージの作成、編集、変換、設定の確認などの管理を行うツールとして qemu-img が提供されている。各種仮想マシンイメージファイルの確認や変換を行うために必要となるため、KVM も含めて準備をしておくことが望ましい。

Cloud image ファイルの Glance での登録や Nova によるデプロイを行う場合、Cloud image ファイルのサイズが大きいと処理に非常に時間がかかってしまい、待ち時間による効率の悪化のみならず、タイムアウトなどにより処理が中断されてしまう事象が発生する。OpenStack の Glance/Nova とともにエラー処理が完全でなく、「処理中」の状態のままとなってしまうことが発生する。

それを防ぐためには、Cloud image は処理時間削減の為、シンプロビジョニングが行える形式を採用する形が望ましいと考えられる。Compute node が KVM であれば、qcow2 形式を採用したい。

qemu-img コマンドでは以下のフォーマットへの双方向の変換が可能となっている。

RAW / qcow2 / bochs / cloop / cow / dmg / nbd / parallels / qcow / vdi / vmdk / vpc / vvfat

各フォーマットに対応バージョンの制限が存在する。例えば vmdk 形式は仮想マシンバージョン 7 or 8 のみに対応で 9 以降の現行バージョンに対応していない為、事前にコンバートが必要となる。

各種仮想マシンイメージから本コマンドを利用し、qcow2 イメージへの変換を行うことにより、前述の問題が大きく解消される。

5.4 その他準備が必要な環境

実際に仮想マシンイメージを Cloud image に変換する為にはハイパーバイザ上で仮想マシンへの追加プロダクトのインストールおよび設定の変更が必要となる。その為、以下の環境の準備が必要となる。

①実行検証用サーバ仮想化環境 (KVM 環境 or VMware 環境(Compute node に合わせる))

②インターネット接続環境

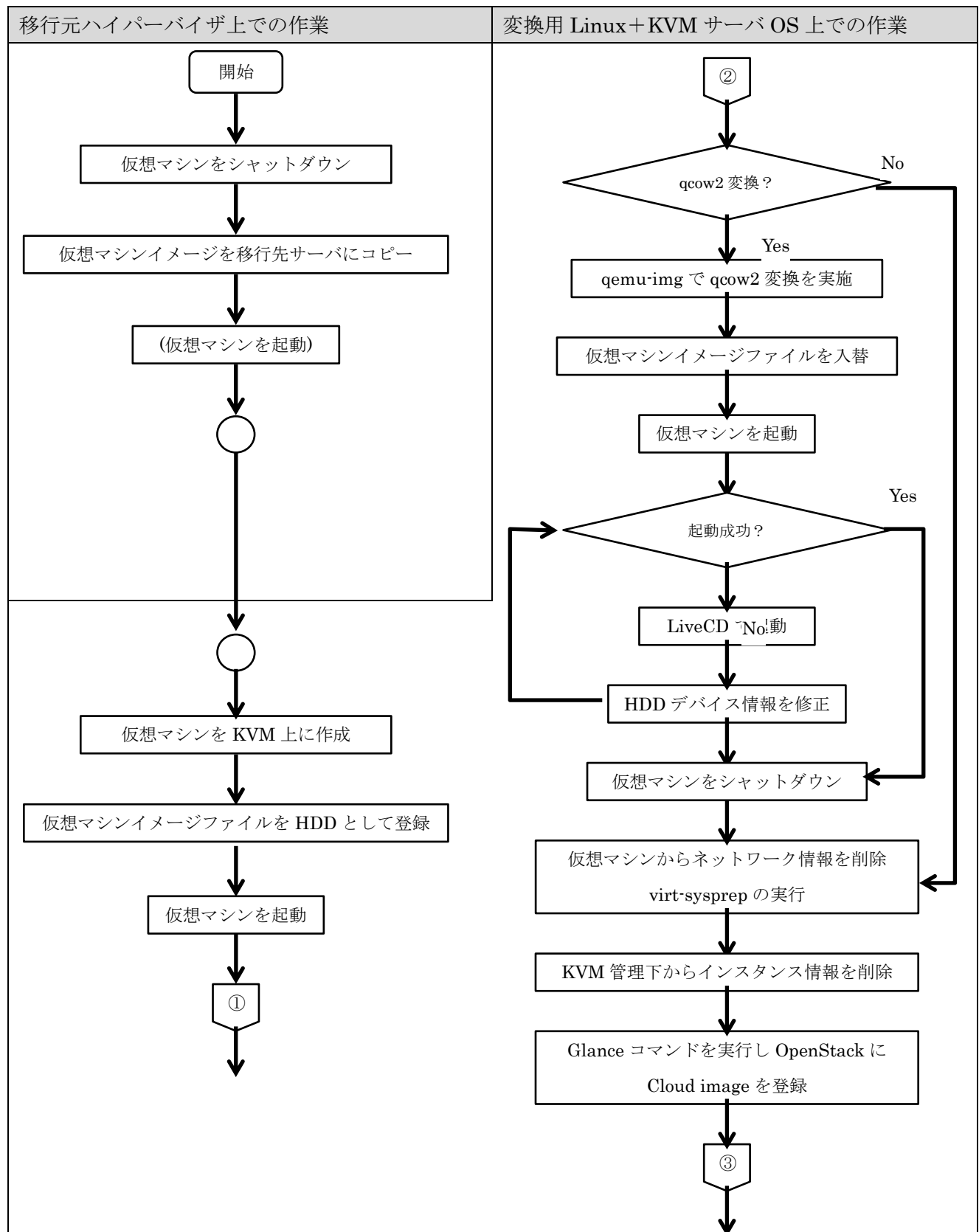
③Linux+KVM サーバ仮想化環境 (検証時は RHEL 6.6 x86_64 を使用)

※lbguestfs-tools パッケージも追加インストール (virt-sysprep コマンドを使用)

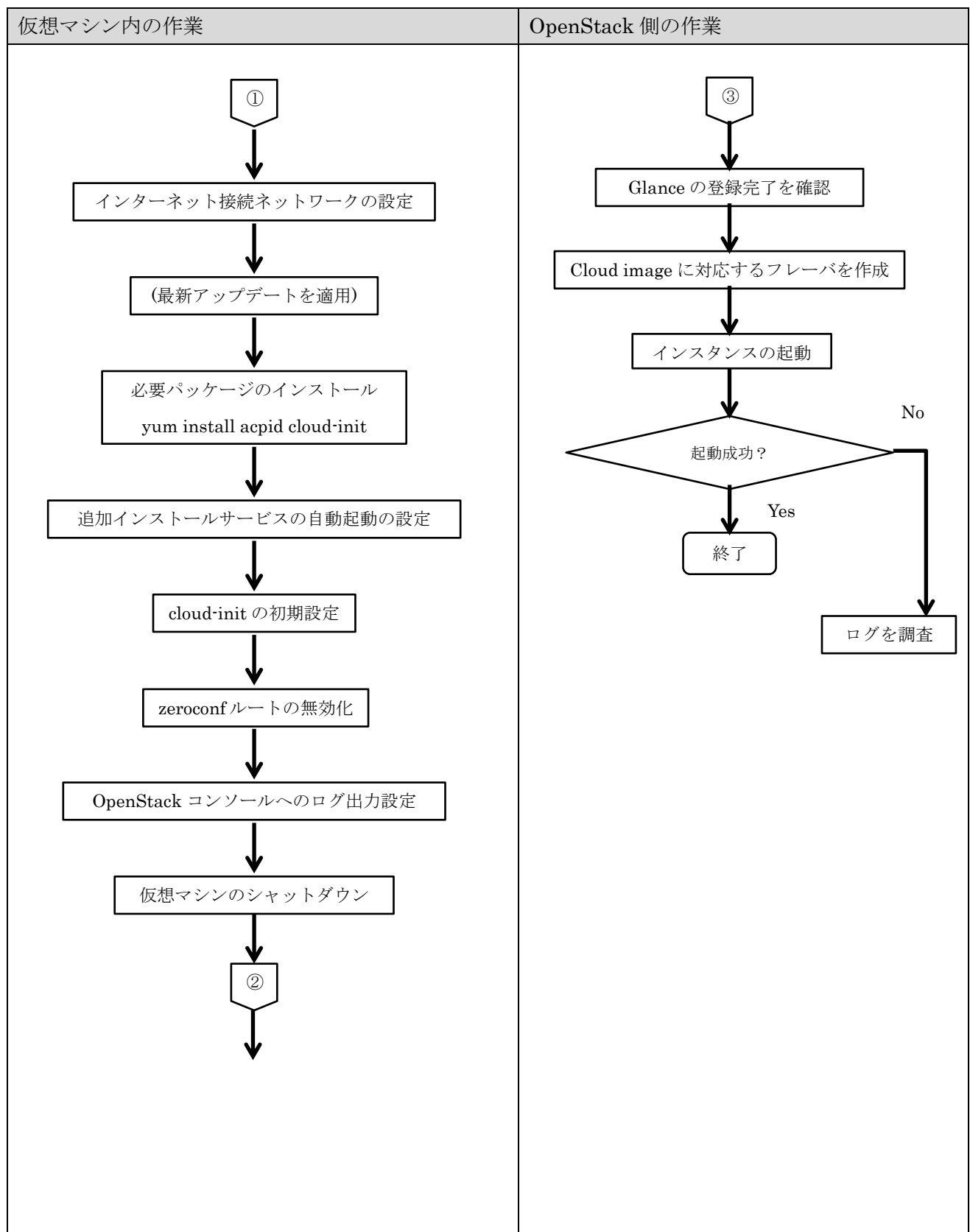
前述の qemu-img コマンドや Cloud image カスタマイズの際に Linux 仮想マシンのネットワーク情報の初期化を行う virt-sysprep 等のツールを使用する場合は、③の環境が必須となるが、イメージファイルの型変換や、virt-sysprep を使用しない場合は必ずしも必要ではない。

物理マシンから Redhat OpenStack & KVM 上への移行検証

5.4.1 Linux KVM イメージの場合

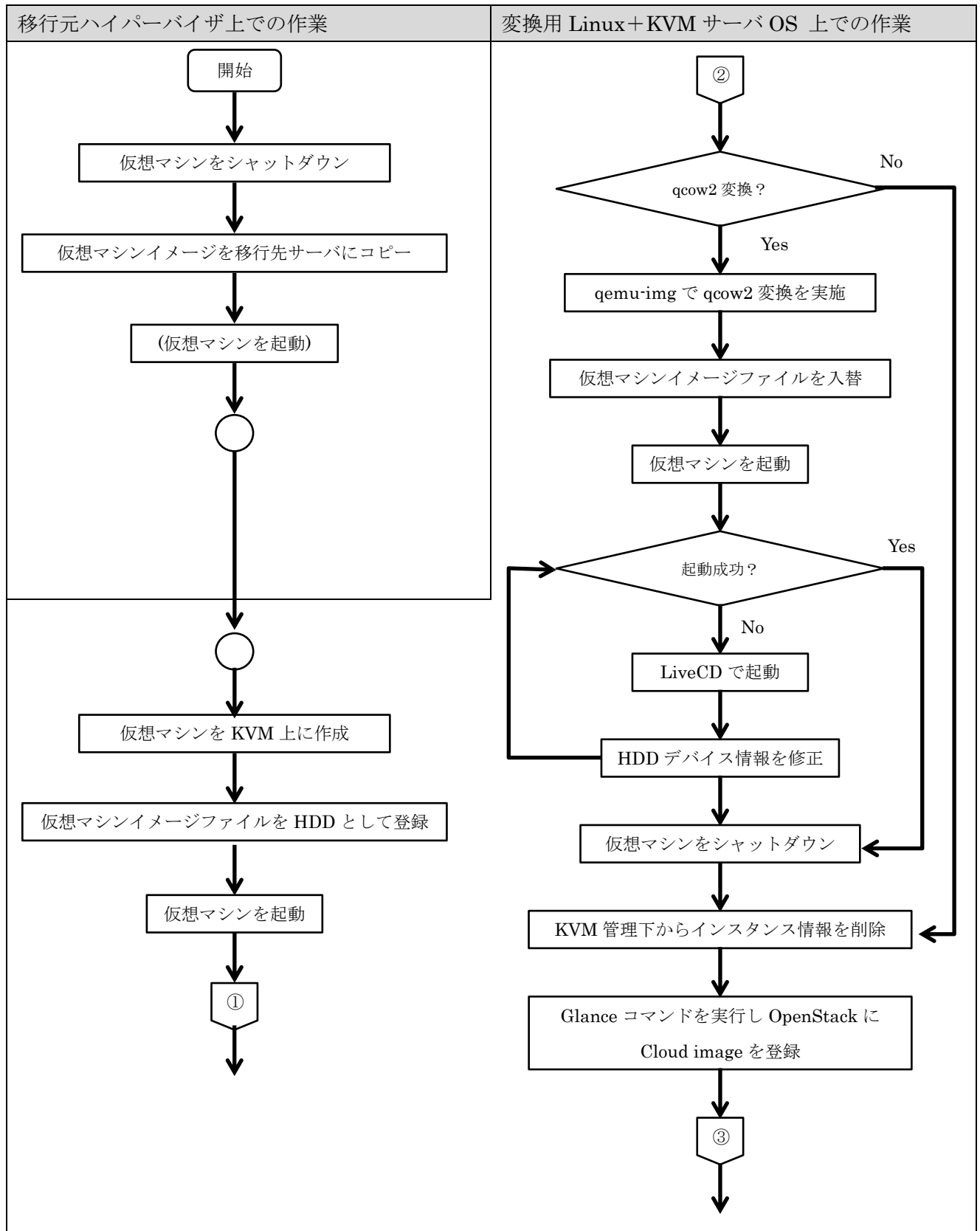


物理マシンから Redhat OpenStack & KVM 上への移行検証

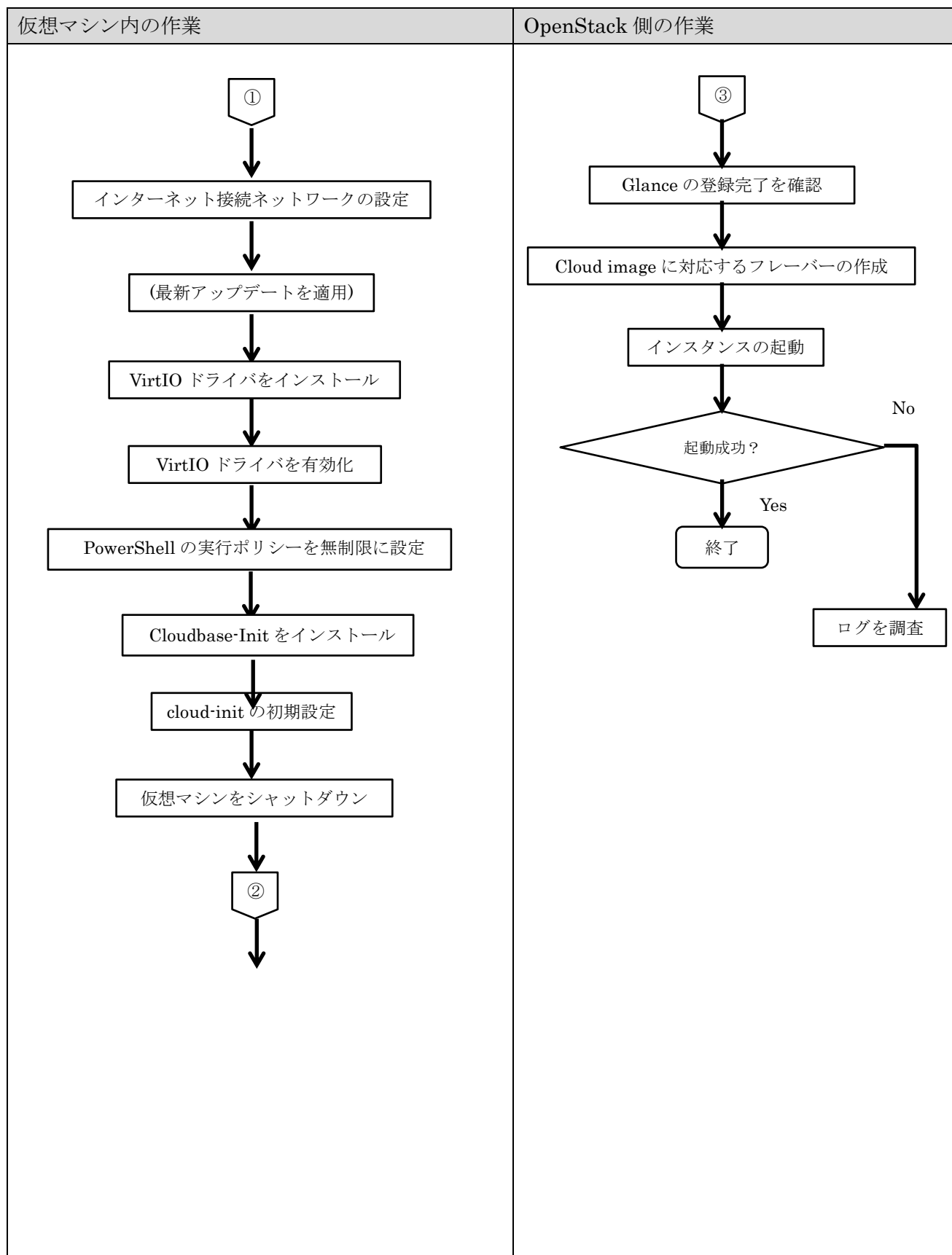


物理マシンから Redhat OpenStack & KVM 上への移行検証

5.4.2 Windows KVM イメージの場合

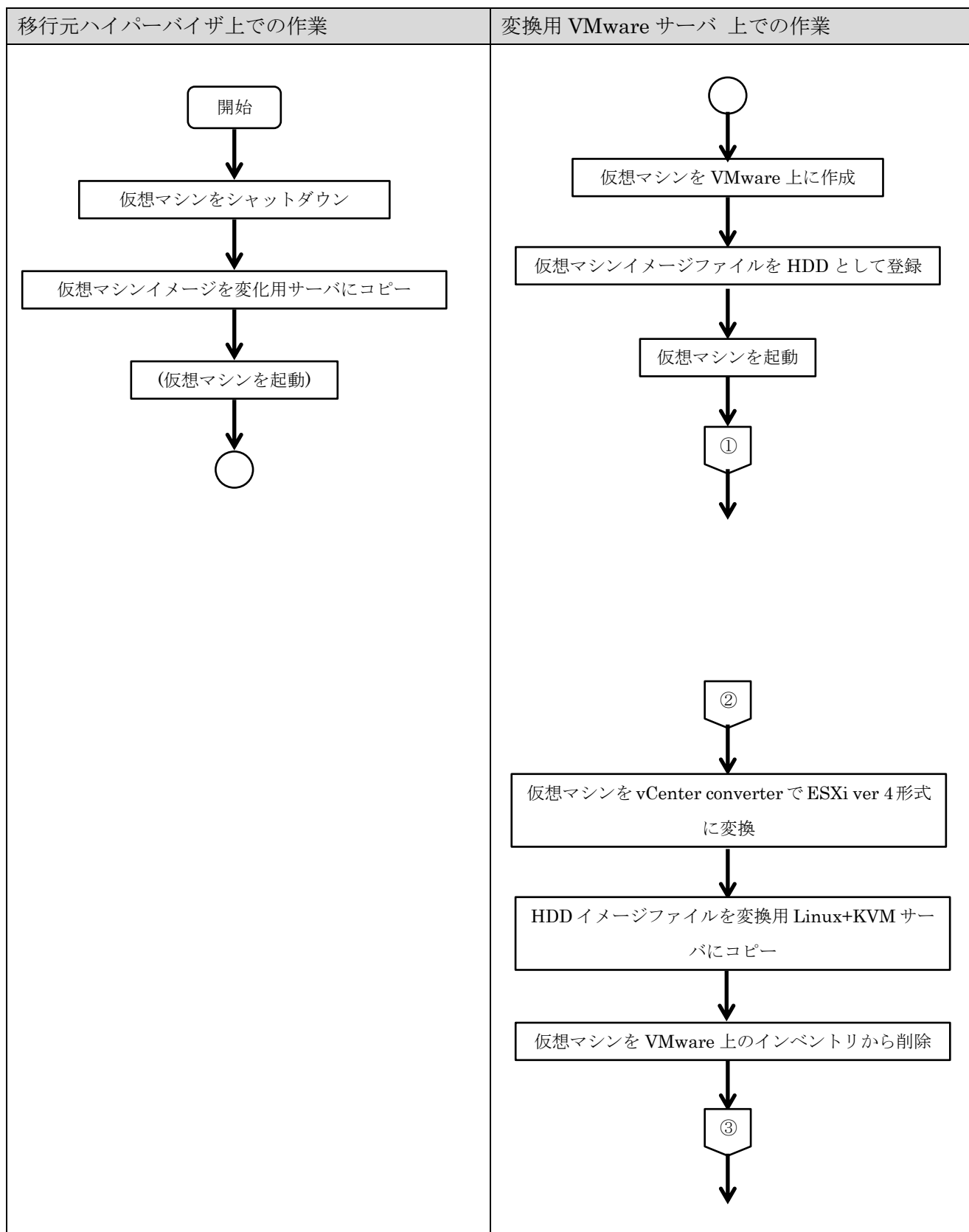


物理マシンから Redhat OpenStack & KVM 上への移行検証



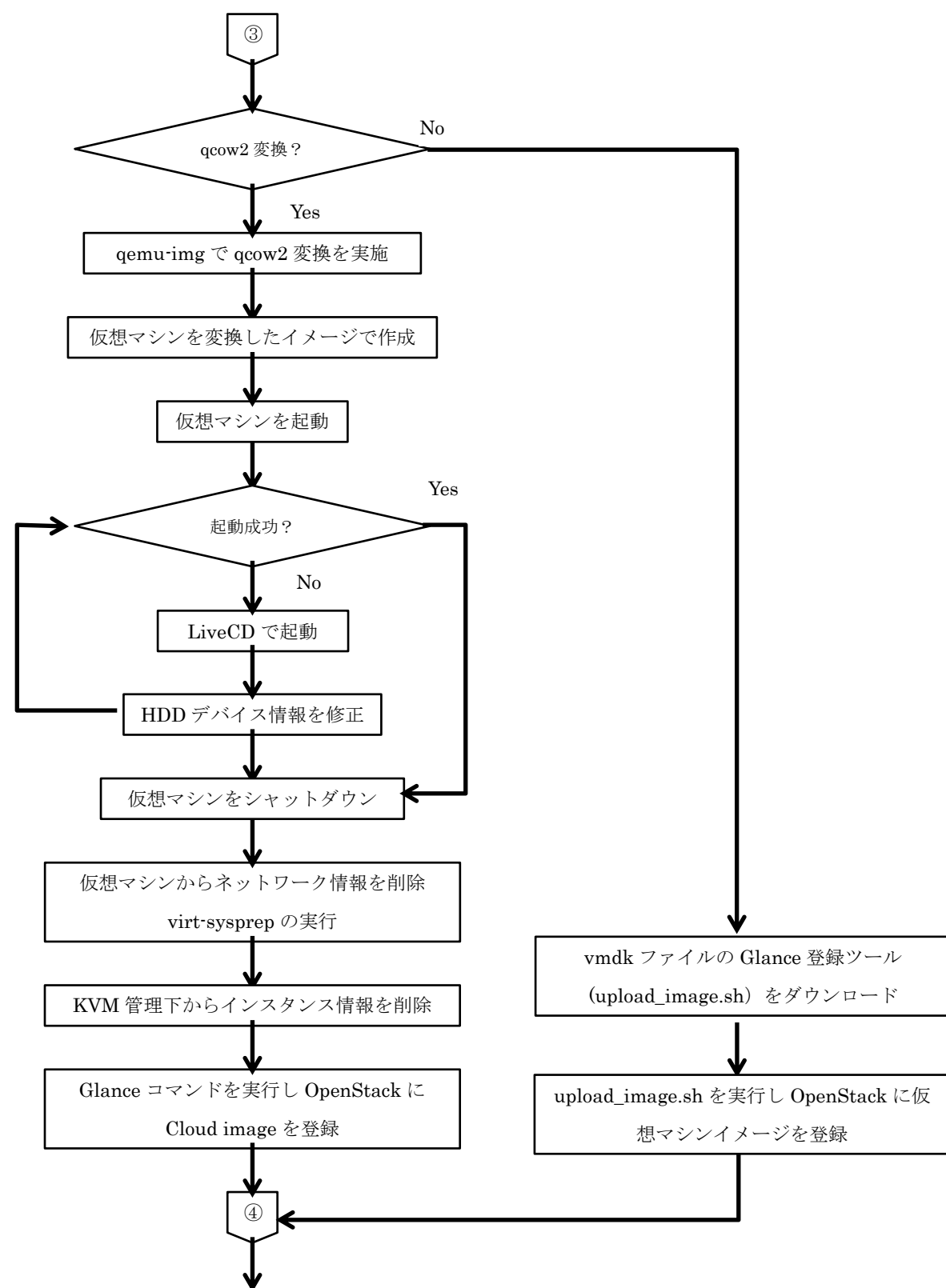
物理マシンから Redhat OpenStack & KVM 上への移行検証

5.4.3 Linux VMware イメージの場合

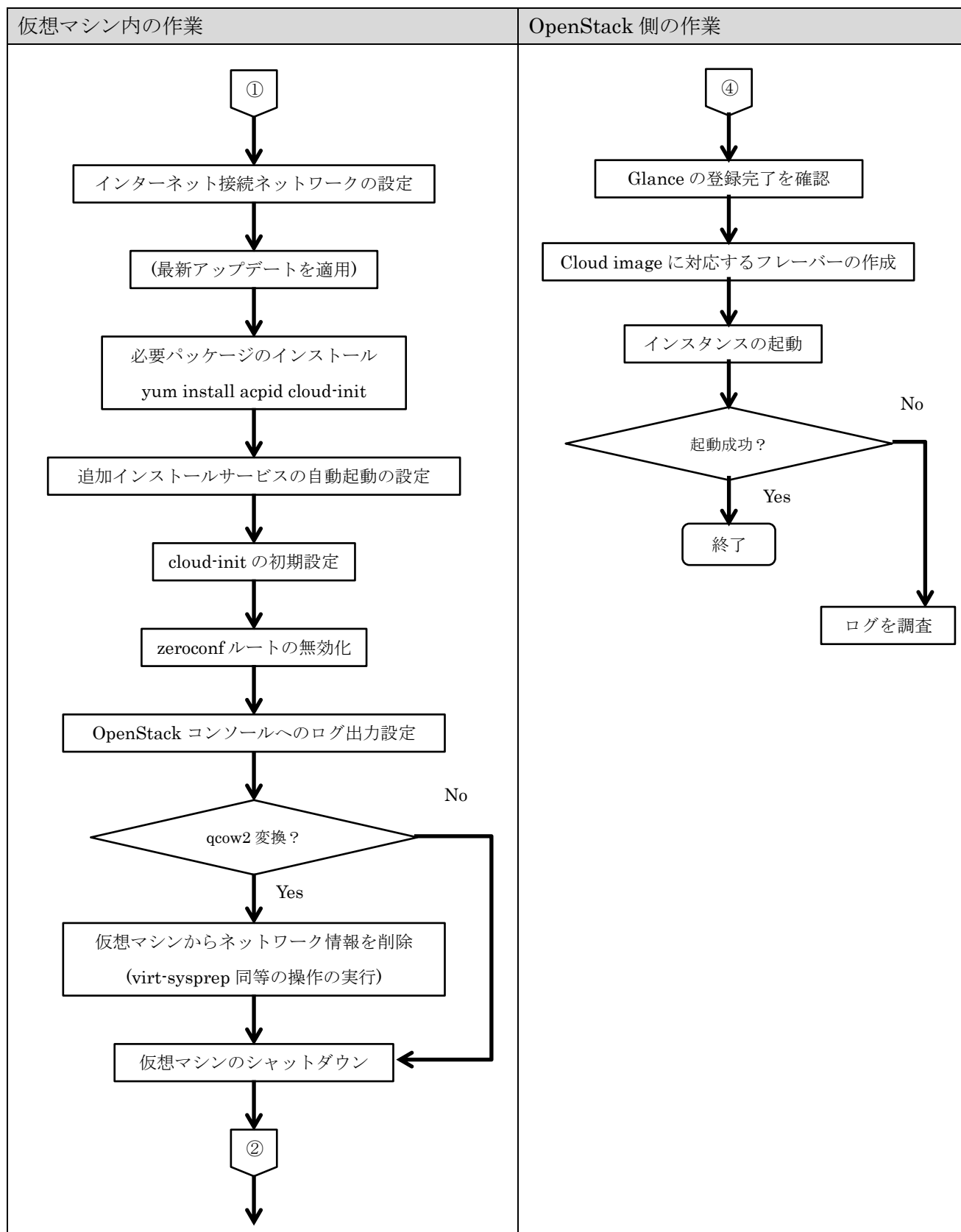


物理マシンから Redhat OpenStack & KVM 上への移行検証

変換用 Linux+KVM サーバ OS 上での作業

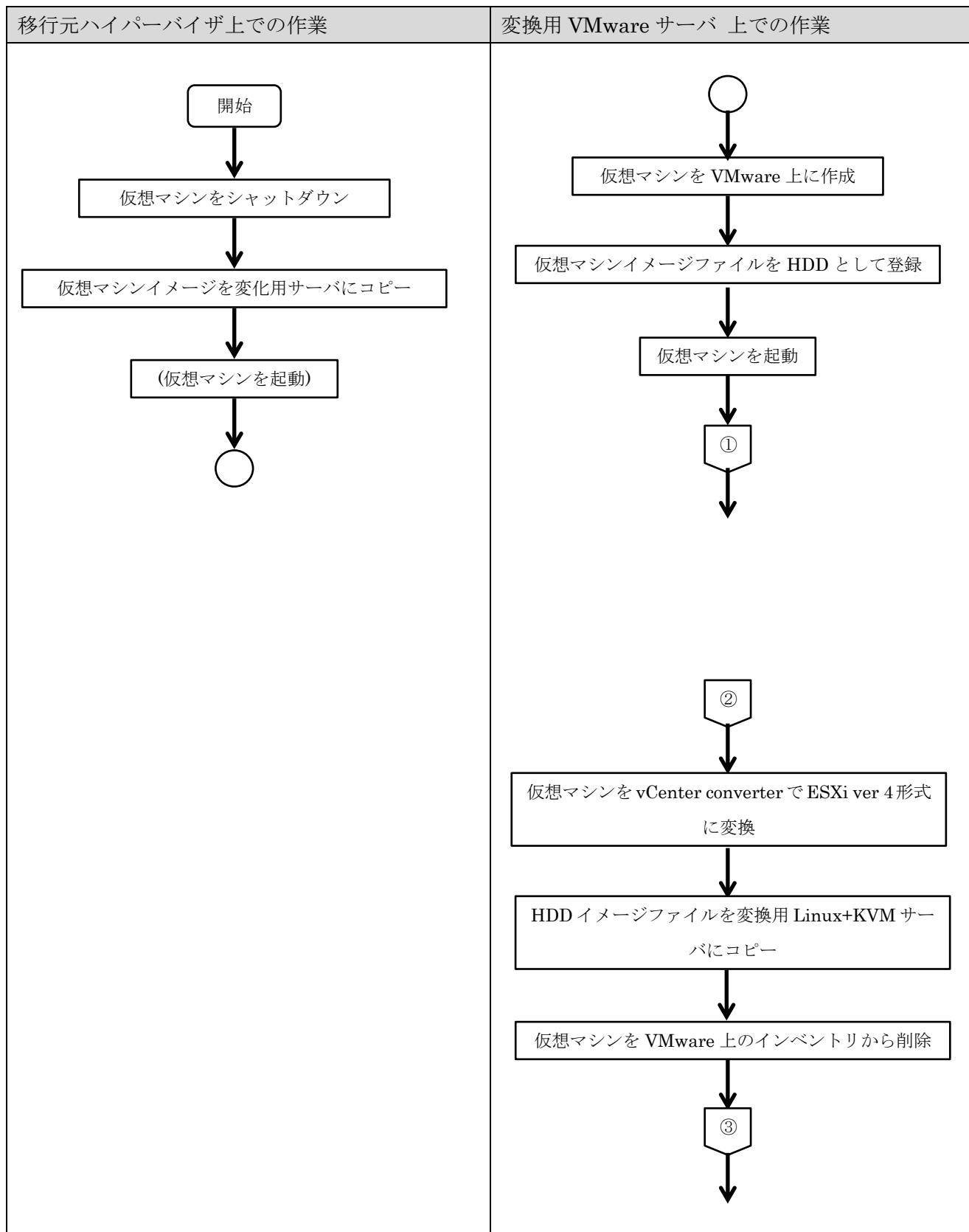


物理マシンから Redhat OpenStack & KVM 上への移行検証



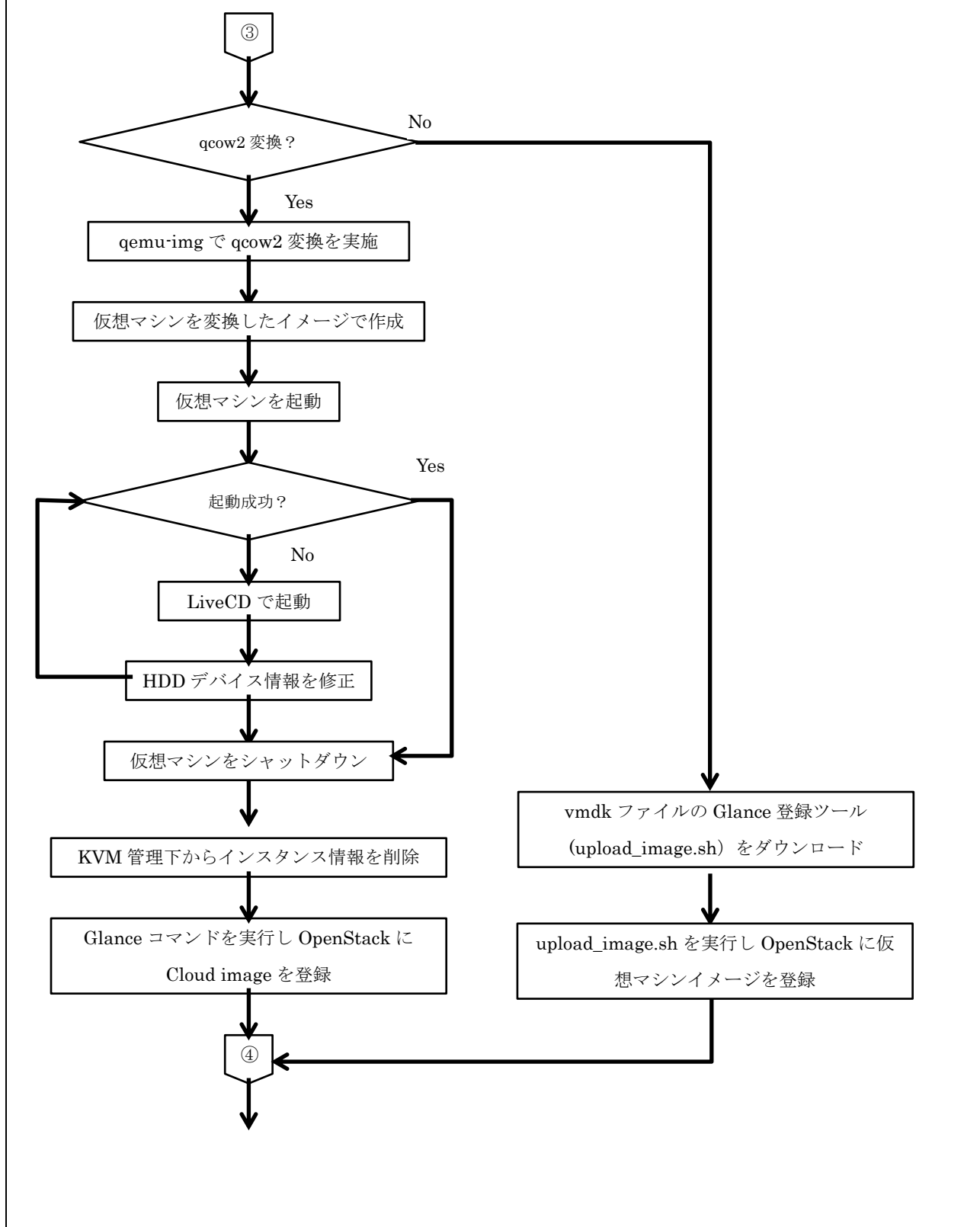
物理マシンから Redhat OpenStack & KVM 上への移行検証

5.4.4 Windows VMware イメージの場合

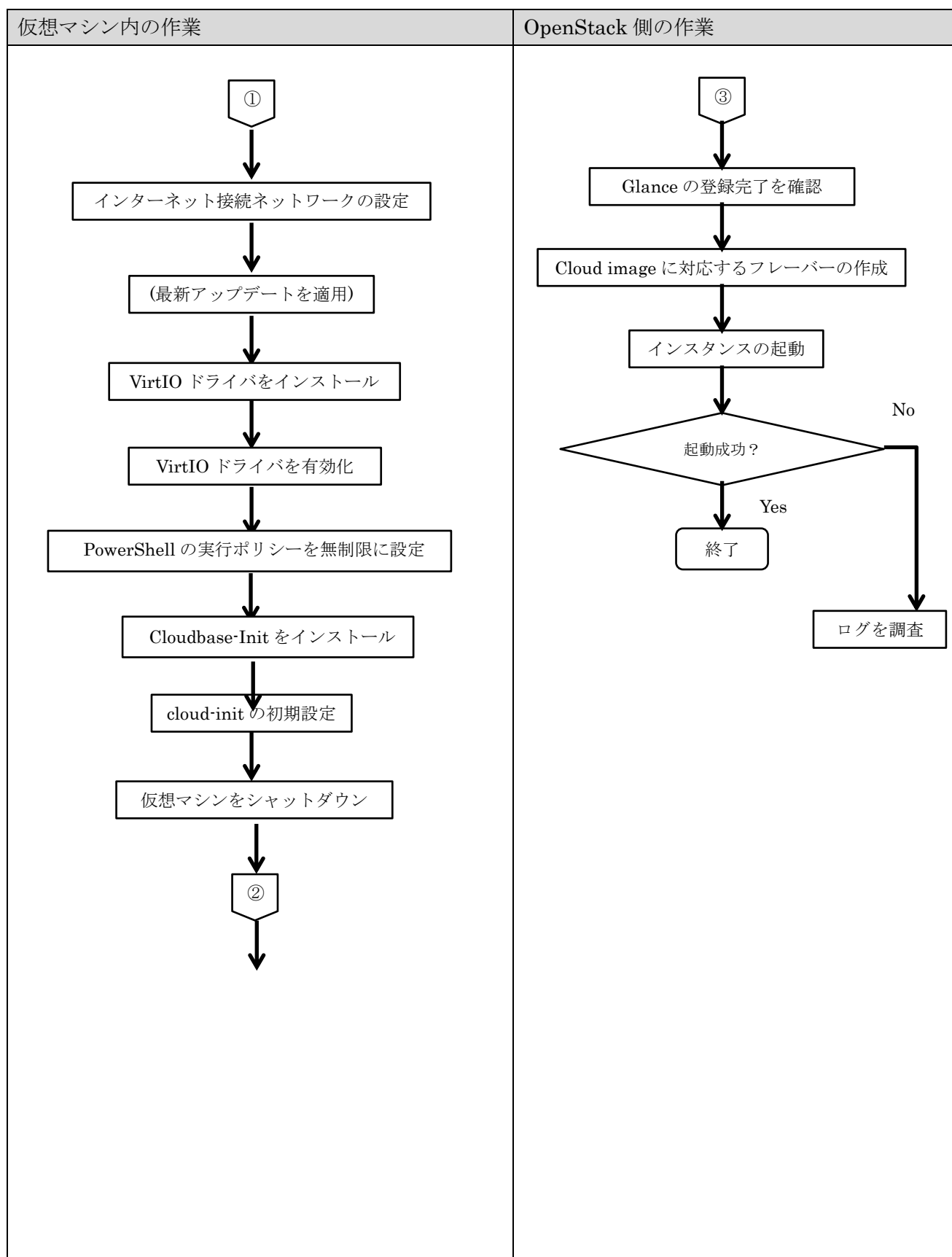


物理マシンから Redhat OpenStack & KVM 上への移行検証

変換用 Linux+KVM サーバ OS 上での作業



物理マシンから Redhat OpenStack & KVM 上への移行検証



物理マシンから Redhat OpenStack & KVM 上への移行検証

5.5 仮想マシンからの Cloud image 作成作業の注意点

5.5.1 ディスク容量による発生する問題と対策

RAW 形式や vmdk(Thick)形式のシックプロビジョニングのイメージファイルを Cloud image に利用することが可能だが、大容量のイメージファイルを利用した場合に以下の問題が発生する。

(1)仮想マシンイメージファイルのダウンロードやコピーに時間がかかる。

(2)OpenStack Glance 登録時に 30 分を超えると、処理が中断されてしまう。

OpenStack Keystone のセッションオブジェクトの有効期限がデフォルトでは 30 分となっている為、Glance の処理が 30 分を超えると処理中にセッションが無効化され中断状態となった。実際に 500GB のイメージファイルの登録を行った際には 4 時間程度の時間がかかっていた。対策として、Keystone のセッション有効期限のタイムアウト値を大幅に伸ばす必要がある。

(3)登録した Cloud image を OpenStack Nova でインスタンスとして起動するのにも時間がかかる。

実際に 500GB のインスタンスの起動を行った際には Glance と同様に 4 時間程度の時間がかかった。大容量の Cloud image は運用の利便性が著しく棄損される為、qcow2 形式や vmdk(thin)形式などのシンプロビジョニングなイメージファイルを採用することを検討したい。

5.5.2 Cloud image の形式の選択

Cloud image の形式の選択は、OpenStack の Compute node に合わせる形が望ましい。OpenStack は Cloud image からのインスタンスの起動時に仮想マシンの形式を Compute node のハイパーバイザに合わせて変換を行ってくれる。

vmdk 形式の Cloud image を KVM にデプロイした場合、イメージの形式が vmdk から qcow2 に変換されると同時にハードディスクコントローラも SATA から VirtIO に切り替わる。内部で認識されるストレージデバイス名も /dev/sdX から /dev/vdX に変更する必要が発生する。この仕様により実際は 1 つの vmdk の Cloud image ファイルを VMware と KVM の両方の Compute node で共用することは起動ができなくなる原因ともなるため、運用が非常に難しくなる。

仮想マシンから Cloud image へのカスタマイズの環境も含めて同一のハイパーバイザで構築することを推奨する。OpenStack + KVM で環境を構築する場合、特別な事情が無い限り cloud image を qcow2 形式で作成し VirtIO ドライバが動作することまで確認しておきたい。

5.5.3 VMware の場合の移行元仮想マシンイメージの形式

VMware の仮想マシンイメージはバージョンが複数存在し、vCenter Converter で出力できる形式もバージョン 4/8/9/10 の何れか選択が可能で、その仮想マシンイメージの形式が各バージョンで異なっている。検証時の OpenStack(Juno)のバージョンではバージョン 8 以外では動作しなかった。これは OpenStack でインスタンス起動時にイメージを生成する qemu-img コマンドが上記バージョン以外には対応していない為と考えられる。その為、移行元となる仮想マシンイメージの取得時は vCenter Converter の仮想マシンバージョン 8 形式で取得することが必須となる。

物理マシンから Redhat OpenStack & KVM 上への移行検証

5.5.4 シックプロビジョニング・シンプロビジョニング形式イメージのメリット・デメリット

前述のとおり、大容量のシックプロビジョニングのイメージファイルを利用することのデメリットは非常に大きい。シンプロビジョニングを利用することで解消される問題も存在するが、別の新たな問題も発生する。それぞれの形式のメリット・デメリットは以下となる。

これらの特性を理解した上で、どの形式を選択するか判断が必要である。

形式	シックプロビジョニング	シンプロビジョニング
特徴	①ハードディスクサイズ=ファイルサイズ ②空き領域がファイル中に存在する ③更新はデータファイルの更新箇所の上書きで記録する ④書き込み時にファイルのリサイズが不要 ⑤ディスクが連続領域で作成される。	①ハードディスクサイズ=使用ディスクサイズ ③空き領域は無し ③更新はファイルサイズの追加 又は 別の差分ファイルの追加で記録する ④書き込み時にファイルのリサイズが行われる ⑤ディスクが不連続領域になる可能性がある
メリット	①ディスクアクセスが若干早い ②更新が多くなっても性能劣化が起きにくい	①データが存在する分のみディスクが使用される ②インスタンスの起動やコピーやバックアップの時間が削減できる
デメリット	①データが存在しない分までディスクが使用されてしまう ②インスタンスの起動やコピーやバックアップに時間がかかる	①ディスクアクセスが若干遅い ②更新が多くなると書き込み領域の断片化により性能劣化が大きくなる

物理マシンから Redhat OpenStack & KVM 上への移行検証

5.5.5 仮想マシンイメージのカスタマイズ

Cloud image を作成する為には仮想マシンイメージをカスタマイズする必要が発生する。手順としては OpenStack 公式の「OpenStack 仮想マシンイメージガイド」の手順に準拠することとなる。

<http://docs.openstack.org/ja/image-guide/content/index.html>

①複数台の仮想ハードディスクを使用している場合

前述のとおり、OpenStack Glance ではハードディスクイメージファイルは 1 個しか登録できない。そのため、複数台の仮想ハードディスクを利用している場合は、変換前に 1 個に統合する必要が発生する。

複数台をソフトウェア Raid 構成で利用している場合は、単純なコピーでは処理が行えないため、V2C ではなく、Cloud image 用に再構築を行うことを推奨する。

②ネットワークの設定

移行元のハイパーバイザから仮想マシンイメージファイルをコピーすると移行元のサーバに設定された IP アドレスや MAC アドレスがそのままの状態に移行されてしまう。移行元のハイパーバイザとカスタマイズ用のハイパーバイザが同一ネットワークに接続されている可能性が高く、移行元、移行先のサーバを同時に起動してしまうと、IP アドレス、MAC アドレスの重複が発生し、ネットワーク障害が発生する可能性がある。

カスタマイズ用のハイパーバイザに移行した仮想マシンを最初に起動させる時点では、NIC を接続しないか、NIC を無効にして IP アドレス、MAC アドレスの情報を変更してから NIC の接続を行う対応が必要となる。

③SWAP 領域

SWAP 領域は OpenStack のインスタンス起動時にフレーバーの設定によって、SWAP 専用の仮想ディスクが提供される。移行元の SWAP 領域は削除しておく必要がある。ただし、Windows サーバに関しては可変長ファイルの SWAP 領域であるため、既存設定に問題がなければ OS 内のファイルを利用し、フレーバーの側の SWAP サイズを 0 にしておく方が、起動時から SWAP が利用できることにより安定すると考えられる。

物理マシンから Redhat OpenStack & KVM 上への移行検証

④大容量データの取扱い

RDB 内のデータや大きな容量のファイルシステムをもつボリュームなどを Cloud image 内に持つことは Cloud image のサイズの増加を招き、運用利便性の低下を招く。OpenStack で構築されるサーバはインスタンスを削除すると、サーバ自体が消失してしまう「揮発性」を持っているため、永続的なデータは Cloud image 中ではなく、永続的ストレージか外部ストレージに格納する必要がある。実際の対応方法は以下が考えられる。

(1)データバックアップを外部ストレージに保有

- Cloud image 作成時はデータを削除してからイメージ化する。
- 構築後に外部のバックアップストレージからデータの復旧を行う。

(2)OpenStack Cinder で永続ボリュームを作成

- 起動時は Web や RDB などのサービスを停止しておく。
- 初回起動時に Cinder ボリュームを作成し、マウントする。
- 初回起動時に DB やファイルシステムを Cinder ボリューム上に作成し、バックアップからのデータ移行を行う。
- 2 回目以降は Cinder ボリュームを自動マウントし、サービスも自動起動する。

※初回起動用と 2 回目以降の 2 種類の Cloud image を作成する対応が必要

(3)iSCSI/NFS 等の外部ストレージをマウント

- 外部の iSCSI や NFS などのネットワークストレージにあらかじめ DB やファイルシステムを構築しておく。
- 起動時は Web や RDB などのサービスを停止しておく。
- 起動後に iSCSI や NFS などのネットワークストレージに接続する。
- ネットワークストレージに接続後、サービスを起動する。

⑤P2V 作業時の設定の復元

Hot Cloning の P2V で作成した仮想マシンはセキュリティ設定の解除やサービスの停止などのカスタマイズが行われている。その設定を復元する必要が発生する。

■Windows 仮想マシン固有の設定

- (1) 管理権限ユーザのパスワードログインの制限
- (2) UAC の制限の設定
- (3) FireWall の復元
- (4) 停止サービスの復元

物理マシンから Redhat OpenStack & KVM 上への移行検証

■Linux 仮想マシン固有の設定

- (1) 管理権限ユーザのパスワードログインの制限
- (2) Firewall の復元
- (3) 停止サービスの復元

⑥PTF の適用

Cloud image から構築されたインスタンスは揮発性な為、起動中の OS 内に対して行った設定の変更や PTF の適用や update などとは保存されず、インスタンスの再構築を行った際に消失してしまう。最初に Cloud image 化を行う前に各種 update の適用は可能な限り行っておきたい。実際に後述する cloud-init をインストールする際に依存するパッケージのバージョンが古く、エラーになることもあるので、その点を踏まえても実施しておきたい。

⑦仮想化ドライバのインストール

前述のとおり、OpenStack 上の Cloud image として利用する為には、Compute node のハイパーバイザに対応した仮想化ドライバをインストールする必要がある。仮想化ドライバをインストールし、有効化した状態で、仮想マシンとして正常に動作することを確認の上で後続処理を行う必要がある。

KVM 上で仮想マシンのカスタマイズを行う場合、RAW 形式、vmdk 形式ともに qemu-img コマンドを利用し、qcow2 形式に変換する必要がある。

VMware はデフォルトでハードディスクイメージは SCSI コントローラで接続されているが、KVM には SCSI コントローラは存在しないため、IDE コントローラを使用して仮想マシンイメージを作成する。

起動したら VirtIO ドライバをインストールした後に、IDE コントローラを VirtIO コントローラに変更して起動を確認しておく必要がある。その際にハードディスクデバイス名が変わってしまう可能性がある。SCSI/IDE の場合 /dev/sdX [X:a-z]だが、VirtIO は /dev/vdX [X:a-z]となる。その為、VirtIO に切り替えた時点で OS 内デバイス設定と異なるため、起動ができなくなる場合がある。

本事象が発生した場合、P2V で利用した LiveCD やインストールメディアの Rescue mode で起動し、OS 内の設定ファイルを起動するまで修正する。

OpenStack でのインスタンス起動時にデプロイされた仮想マシンが動作しない状況が検証でも多発しており VirtIO を有効化したあとの起動確認は入念に行っておきたい。

物理マシンから Redhat OpenStack & KVM 上への移行検証

⑧【重要】ここで作成した仮想マシンイメージを保存

ここまでカスタマイズを実施した仮想マシンを一旦保存する。これ以降の作業を実施すると、ネットワークが削除されてしまう等の設定変更が行われ、仮想マシンイメージとして利用するのが難しくなるため、この状態を保存しておく。

今後の運用ではPTFやupdateの適用、プロダクトのアップデートは本イメージを「テンプレート」としてコピーを行い、テストを実施する。問題が無いことが確認された時点で、更新したイメージを新たな「テンプレート」として Cloud image を作成するベースとして利用する。

⑨Cloud image 化の最終カスタマイズ(cloud-init のインストール)

■Windows 仮想マシン固有の設定

- (1) インスタンスの起動時に、Cloudbase-Init がスクリプトを実行できるようにするために、PowerShell の実行ポリシーを無制限に設定する。
- (2) デバイスマネージャーを非接続デバイス情報の表示を ON にして起動し、非接続となったデバイス情報を削除する。(物理マシン時代やカスタマイズ中の不要デバイスの情報を削除)
- (3) Cloudbase-Init をダウンロードして、インストールする。
- (4) Cloudbase-Init 設定オプションでロギングのユーザ、ネットワーク、ポートを設定する。
- (5) インストール完了後、Complete the Cloudbase-Init Setup Wizard ウィンドウで、Run Sysprep と Shutdown のチェックボックスを選択し、Finish をクリックする。
- (6) 仮想マシンをシャットダウン((5)終了後、自動的に行われる)する。

■Linux 仮想マシン固有の設定

- (1) ACPI サービス(acpid)をインストールと自動起動の設定
- (2) cloud-init のインストールと自動起動の設定
- (3) cloud-init の設定ファイルの編集
- (4) /boot/grub/menu.lst に nova console-log で出力されるように設定を追加
本設定は kernel 設定の行に直接追加する為、現状利用して kernel メニューの行をコピーして新たなメニューを作成する形にすることで、設定ミスがあった場合に旧設定で起動ができる回避策が取れる。この設定は kernel の update をすると更新されるため、設定が消える問題が懸念されるので、kernel の update を実施する場合は注意が必要
- (5) ネットワーク設定を削除 (カスタマイズ環境が VMware の場合)
 - ・ /etc/sysconfig/network-scripts/ifcfg-eth0
MAC アドレスの削除
固定 IP の場合、DHCP に変更
GATEWAY の削除
 - ・ /etc/udev/rules.d/70-persistent-net.rules
先頭のコメント行を除きネットワークカード認識の行を削除

物理マシンから Redhat OpenStack & KVM 上への移行検証

(6) 仮想マシンをシャットダウン

(7) ネットワーク設定を削除 (virt-sysprep 実行) (カスタマイズ環境が Linux+KVM の場合)

⑩仮想マシンをハイパーバイザの管理から外す

KVM は、仮想マシンを「libvirt のドメインから削除」し、VMware は「仮想マシンをインベントリから削除」を実行する。

⑪作成した Cloud image を OpenStack 管理サーバに転送する。

OpenStack にイメージをアップロードできる OpenStack 管理サーバに Cloud image としてカスタマイズした仮想マシンイメージを転送する。

⑫Cloud image を OpenStack に登録する。

qcow2、RAW 形式のイメージは通常の glance image-update コマンドを使用して OpenStack に登録を行う。

http://docs.openstack.org/ja/user-guide/content/cli_manage_images.html

vmdk 形式のイメージは vmdk 形式用の upload ツールを使用して登録する。

https://wiki.openstack.org/wiki/NovaVMware/DeveloperGuide#Upload_your_VMDK_to_glance

⑬OpenStack 管理画面でインスタンスの起動を行う

OpenStack 上で登録した Cloud image に対応するフレーバーを作成する。そのフレーバーを使用し、OpenStack Nova の機能を使用し、インスタンスの起動を行う。サーバが起動したら、移行前の状態と機能に差異が無いかの検証を行う。

5.6 仮想マシンからの Cloud image 作成の未検証部分

今回の検証作業では KVM 又は VMware の仮想マシンイメージから OpenStack + KVM の Cloud image への変換を検証したが、今回、検証を行えなかった点は大まかに以下となる。

①vdi、vhd、hdd などの KVM、VMware 以外の仮想マシンイメージからの移行

②VMware、Xen、Hyper-V などの KVM 以外の Compute node への移行

③古いバージョンの OS で構築された仮想マシンイメージの移行

今回の検証でも vmdk 形式のイメージファイルの対応バージョンの制約が存在したが、他の形式でも同等の問題が発生する可能性がある。

Compute node として使用するハイパーバイザごとに組み込む仮想化ドライバや OS 上の設定が異なる可能性もある。さらに VirtIO に対応していない OS を移行する場合はディスクコントローラとして VirtIO が使用できないため、Glance 登録時のパラメータ指定等で対応する必要も発生すると考えられる。

本番で実際の移行を行う為には、事前にリハーサルを含めた実機での検証を行う必要がある。

物理マシンから Redhat OpenStack & KVM 上への移行検証

6 OpenStack 上でのインスタンス更新の運用

OpenStack で運用を行う場合、Cloud image で構築されたインスタンスは揮発性であることを認識しておく必要がある。システムを稼働させ続けるためには OS やプロダクトの PTF 適用やバージョンアップは不可避であるが、揮発性であるため、再度インスタンスを作成しなおすと、更新は消えてしまう。

インスタンス起動後にそれまで適用した PTF やバージョンアップを適用することも可能だが、インスタンス構築時に PTF、バージョンアップの適用の作業を都度行う必要が発生する。

その為、以下の問題が発生する。

- ①インスタンス構築後に PTF、バージョンアップを適用する為、復旧に時間がかかる。
- ②適用した PTF やアップデート用のファイルや手順を保存しておく必要がある。
- ③更新が完了したサーバの稼働確認を行う必要がある。

運用を続けていくと当然だが、OS やパッケージやアプリケーションの更新が必要となる。都度 Chef や Puppet などのツールと共に自動化をすることも可能だが、インスタンスを作成する度に更新作業を順番通り行うのは時間もかかる上に、作業負荷も大きい。OS の PTF などのアップデートは Windows update や yum など更新を行うことも可能だが、その時適用したものと同一バージョンのアップデートが適用される保証はない。Ruby のようにバージョン依存の厳しいパッケージが存在した場合、システムが動作不能になってしまう可能性もある。万全を期すのであれば、Windows update や yum などの自動適用ではなく、アップデートファイルをダウンロードして保存後に適用する必要が発生する。さらにアップデートを適用した後にシステムが正常に動作するかを確認する必要もあり、障害復旧などでインスタンスを再作成する運用としては、利便性、信頼性ともに良い状況ではない。

本件を解消する運用の手順は以下の 2 つの案が考えられる。

6.1 OpenStack 上の Snapshot を利用する

OpenStack の Snapshot の機能を利用し現状イメージのコピーを作成し、更新を適用して新たな Cloud image として登録を行う方法である。OpenStack を利用する上では一般的だと考えられる。

通常、本番環境と開発・検証環境は分離していることが多いと考えられる。

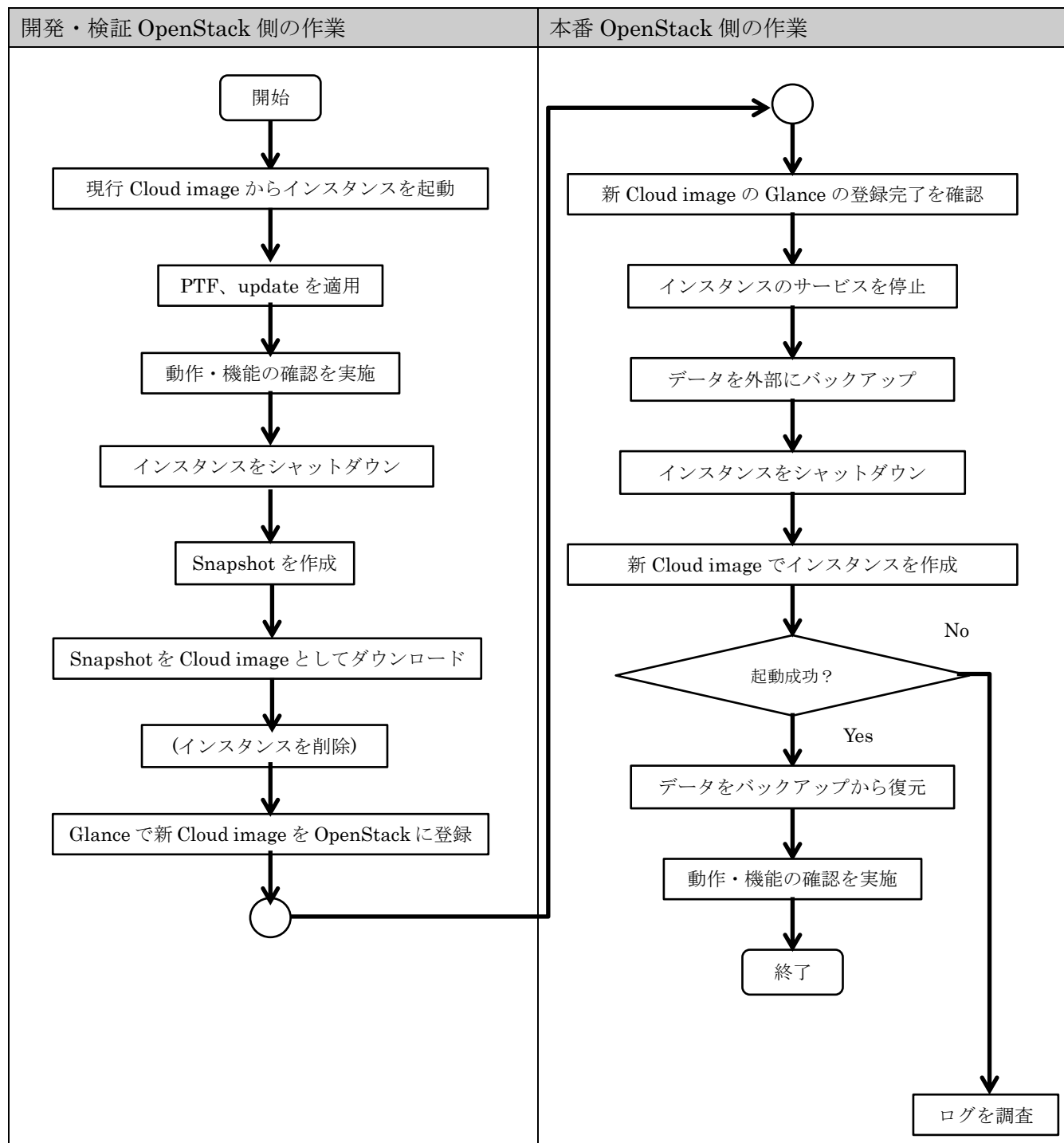
6.2 仮想マシンイメージのカスタマイズ環境を利用する

仮想マシンイメージのカスタマイズに利用したハイパーバイザの環境を利用する方法も考えられる。

Cloud image の作成に利用した仮想マシンイメージに対して更新を適用し、再度 Cloud image としてカスタマイズを行って OpenStack に登録を行う。仮想マシンイメージのカスタマイズに利用した環境を開発・検証環境として利用することで本番環境との分離を実現する。

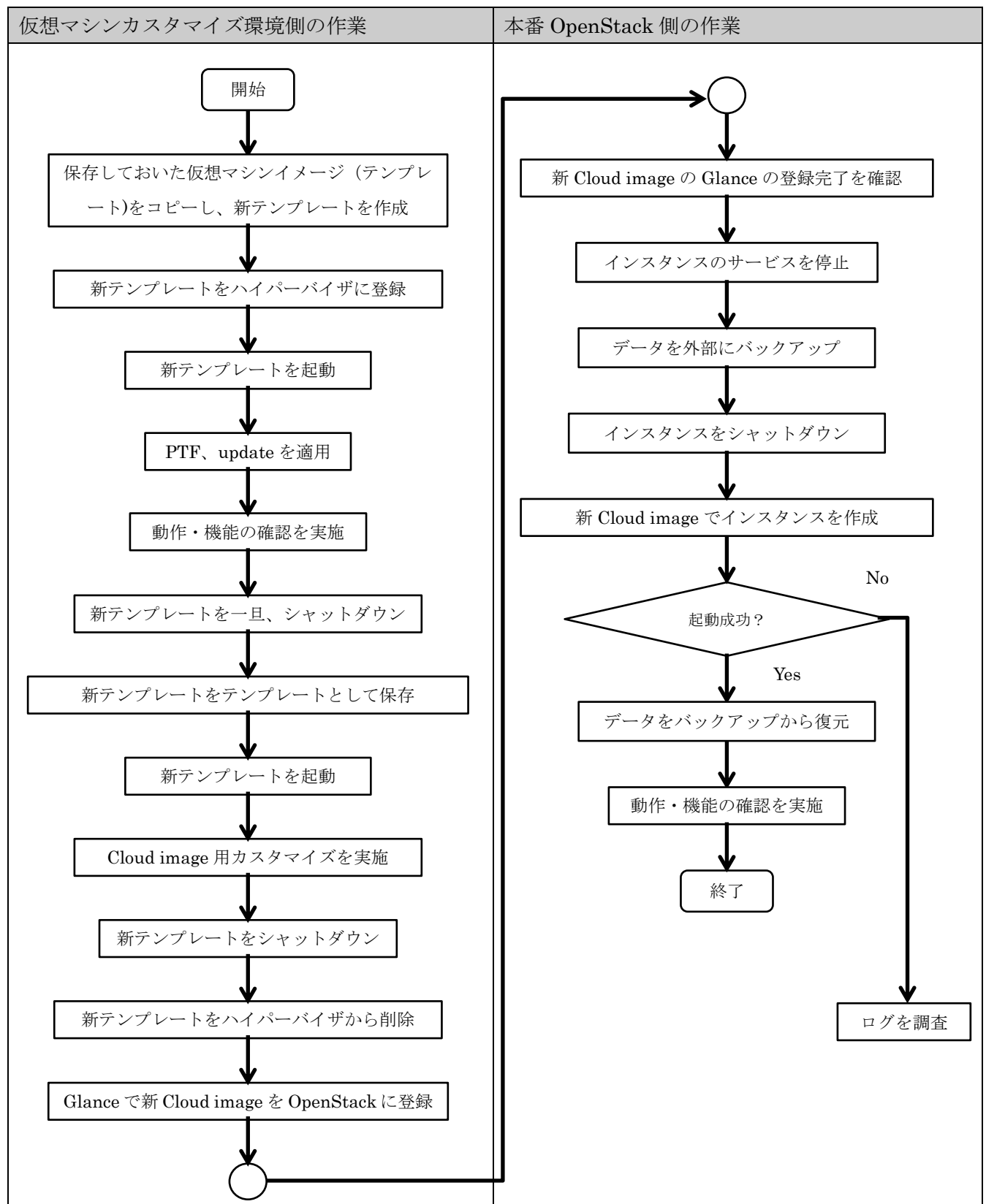
物理マシンから Redhat OpenStack & KVM 上への移行検証

6.3 OpenStack 上の SnapShot を利用する場合の運用フロー



物理マシンから Redhat OpenStack & KVM 上への移行検証

6.4 仮想マシンイメージのカスタマイズ環境を利用する倍の運用フロー



物理マシンから Redhat OpenStack & KVM 上への移行検証

本運用は OS、プロダクトのアップデートだけではなく、移行・検証環境の増幅や障害時のリカバリやディザスタリカバリ等に応用が可能である。インスタンスの起動からバックアップデータの復元と動作・機能の確認の自動化は OpenStack 環境上での運用を行う上で重要となる技術と考えられる。

7 総括

以上の検証の結果を受けて、物理マシンから（仮想マシンを経由して）OpenStack Cloud image に移行する為に必要となるインフラ基盤・技術と考慮点についてまとめる。

7.1 Cold Cloning と Hot Cloning の選択

今回は物理マシンから仮想マシンイメージを取得する為に Cold Cloning と Hot Cloning の両方の検証を行った。処理の安全性ではデータ破損の発生しない Cold Cloning を採用すべきだが、処理時間の面では Hot Cloning が圧倒的に早い。通常の物理サーバは十分に余裕ディスク容量を持っていることが多く、その空き容量までイメージ化してしまう Cold Cloning は処理時間という観点では、かなり不利である。停止時間を多く取れない場合は、Hot Cloning も十分に候補になると考えられる。

7.2 仮想マシンイメージを取得するデバイスの選択

物理マシンから仮想マシンイメージを取得する際に以下の何らかを選択するケースが多いと考える。

- ①ネットワークストレージ(NFS 等)を接続
- ②USB の外付けハードディスクを接続
- ③e-SATA の外付けハードディスクを接続

②は追加する機材が少なく、データの移送も容易だが、処理時間の面では若干の考慮が必要である。

各接続方式転送速度については以下の表を参考にして欲しい。

接続方式	USB 2.0 HDD	NFS ストレージ (1000Base-T 接続)	USB3.0 HDD	e-SATA 接続
理論転送速度	60MB/sec	128MB/sec	640MB/sec	384MB/sec
連続書込み (実測)	約 25MB/sec	約 100MB/sec	約 100MB/sec	約 180MB/sec
100GB 転送時間	4096 秒	1024 秒	1024 秒	569 秒

サーバに USB 3.0 又は e-SATA のインターフェースが接続されているのであれば、問題ないのだが、P2V を行うような旧型のサーバの場合、USB 2.0 が一般的だと考えられる。その場合、仮想マシンイメージの転送に非常に時間がかかり、あまり現実的ではないことがわかる。

USB 2.0 インターフェースしかない場合は、ネットワークストレージを利用することを検討したい。

物理マシンから Redhat OpenStack & KVM 上への移行検証

本検証では、KVM サーバ上に大容量ストレージを構築し、NFS サービスを起動させることで、P2V に使用するストレージとした。この方式であれば、外部ストレージから KVM のハイパーバイザのデータストアに転送する必要がなく、さらに作業時間を短縮することが可能である。

移行を行う場合は、複数の物理マシンを同時ではなく、シーケンシャルに 1 台ずつ行う方が、ストレージ・ネットワークを占有できるため、トータルの速度は速くなる。ネットワークも可能であれば、独立したネットワークを構築し占有できるようにしたい。

7.3 移行検証・カスタマイズ用サーバの必要性

P2V や V2C に使用するサーバは以下の条件を満たすものを推奨する。

- ①物理マシンから移行した仮想マシンを動作させるのに十分な CPU 性能
- ②複数台の仮想マシンを同時に動作させられる大容量メモリ
- ③高速かつ大容量なローカルストレージ
- ④検証用のハイパーバイザを搭載している。

移行作業には大容量ファイルの形式変換や転送を頻繁に行う。変換の途中過程のイメージファイルでの動作確認も頻繁に行うため、ハイパーバイザも搭載されていた方が転送が不要な為、利便性が非常に高い。

OpenStack 上で Cloud image を運用・管理する為にも作業+検証環境として利用できるため、永続的に使える環境として準備して置きたい。

7.4 物理サーバを Cloud image に移行するのは有効なのか？

P2V は物理サーバを大きな変更を行うことなく、クラウド環境に移行できるメリットがある。ただ、移行可否については、物理サーバのストレージの構成によるところが大きい。ソフトウェア Raid や特殊なドライバが必要なデバイスを使用していなければ、ほぼ移行は可能であると考えられる。

P2V をが適しているケースは以下だと考えられる。

- ①移行対象となる物理サーバが多数存在する為、テストが行いきれない。
- ②OS が古く、最新ハードウェアにインストールしても動作しない。
- ③サーバに搭載されているストレージがあまり大きくない。

もしクラウド環境で OS インストールから構築ができるのであれば、P2V を無理に選択する必要はない。

さらに OpenStack への移行だが、Cloud image 標準のディスク構成が、現行の OS のインストールの標準構成と乖離が大きく、物理マシンイメージからの完全な Cloud image への移行は仮想マシン状態でのカスタマイズに相当手間がかかることとなる。ただ、問題になるのはディスク領域の自動拡張が動作しない点に絞られており、その部分をあきらめることができれば、移行は可能と言える。

物理マシンから Redhat OpenStack & KVM 上への移行検証

7.5 事前検証は必ず実施

今回の検証を行うに当たり、事前に各ベンダーや OpenStack 公式資料のみではなく、ネット上で公開されている情報をある程度調べて、作業を開始した。ところが、想定外のところで問題が発生し、その調査の為に時間が取られることが、多々発生した。

今回の移行手順でも記載したが、移行を行う手段や経路は複数選択可能で、障害が発生した場合も別の手段を選択することが可能である。

作業計画や手順書を作成した後、必ず「実機」を使用しての事前検証を行うことを強く推奨する。

7.6 移行後の実運用の検討が重要

物理サーバとクラウド環境では運用作業が大きく変わる箇所が発生する。ハードウェアの死活監視は無くなるが、共用環境になることにより、リソース使用量の監視が必要となる。クラウドの機能を活用する為にオートスケールやクラスタリングなども追加が可能である。バックアップやリカバリの運用も変更する必要が発生するだろう。

特に OpenStack での管理に移行を行う場合、単純にクラウド上に移行するより遥かに多い変更点が発生する。

移行計画と合わせて、運用設計についても十分に検討を行い手順が確立してから移行を行うべきである。

8 謝辞

最後に、本検証にあたり、検証環境・機材をご提供頂きましたデル株式会社様、OS・プロダクトの評価用ライセンスをご提供頂きましたヴィエムウェア株式会社様、日本マイクロソフト株式会社様、レッドハット株式会社様、さらに技術課題の解決にご支援を頂きましたデル株式会社 増月孝信様、村上洋介様、レッドハット株式会社 助川学様、小島克俊様に感謝の意を申し上げます。