Empowered by Innovation NEC

OpenStack を
商用サービスに活かす勘所

2015年2月3日

NEC ソフトウェア技術統括本部 吉山晃

自己紹介：吉山晃

1995年	• Slackware Linux でフリー UNIX クローン（Linux）を使いはじめる 以来 Linux/OSS コミュニティで活動
1998年	• NEC入社 • 個人的に Linux/OSS コミュニティで活動
2001年	• OSSソリューションセンター（当時）に異動 以来、Linux/OSS 全般の SI・障害対応等に従事
2010年	• OSS のクラウド基盤担当になる • 同年発表された OpenStack の評価を開始
2011年	• 国内の OpenStack コミュニティに参加。イベント対応、勉強会主催など
2013年	• NEC Cloud IaaS 構築プロジェクトに従事 • OpenStack Foundation公認エバンジェリスト「OpenStack Ambassador」に選任

はじめに

本セッションの目的

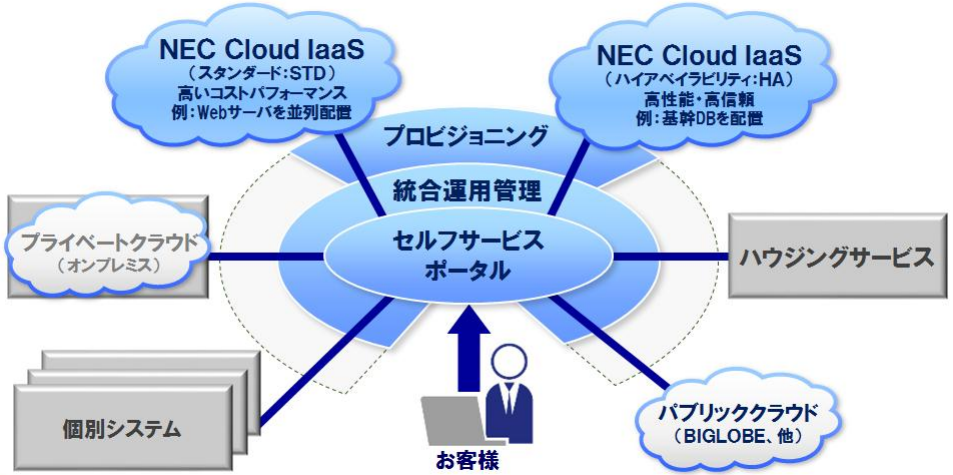
NEC Cloud IaaS(他のプロジェクト含む)で得られた
OpenStack 利用の IaaS 設計・構築・運用のポイント紹介

OpenStack を用いたサービス

- 多数あるOpenStackの機能のどれを提供する／しないのか

クラウド基盤＝大規模システム

- VMに関するSLAの考え方
  高SLA型 ／ 低SLA＋α型
- スケーラビリティは常に大きな問題
  ボトルネックになりやすいのはどこか、
  どうすれば回避できるか、
  OpenStack ではシステムをどう
  スケールしていけばよいのか、
  何を監視すべきか



NEC Cloud IaaS全体像

クラウド基盤の開発・運用

- 開発・保守サイクルをどう回していけば良いのか
- OpenStack のバージョンアップはどうするのか

商用サービスにおける
OpenStackの設計・運用

クラウド基盤の冗長化

一番数の多いVMホストサーバの冗長化は課題

制御系サーバ・ネットワークゲートウェイ・クラウドAPI

- 従来のクラスタリング手法で冗長化は可能
  ⇒1サーバが止まっても各々のサービスは継続

VMホストサーバ

- 冗長化手法が無いわけではないが高コスト
  - 高価な無停止サーバ
  - 複数ハイパーバイザによる稼働VMのミラーリング
- 共有ストレージを用いたVMのディスクデータ保全
  - 最悪VMホストサーバがクラッシュしてもVMのデータは維持できる
  - 1ストレージを多数のVMホストサーバが同時アクセス→ボトルネック

VM の SLA

VMホストサーバ障害に関するVMのSLAポリシーは主に2通り

高SLA型：基本的にサービス提供者が障害対応

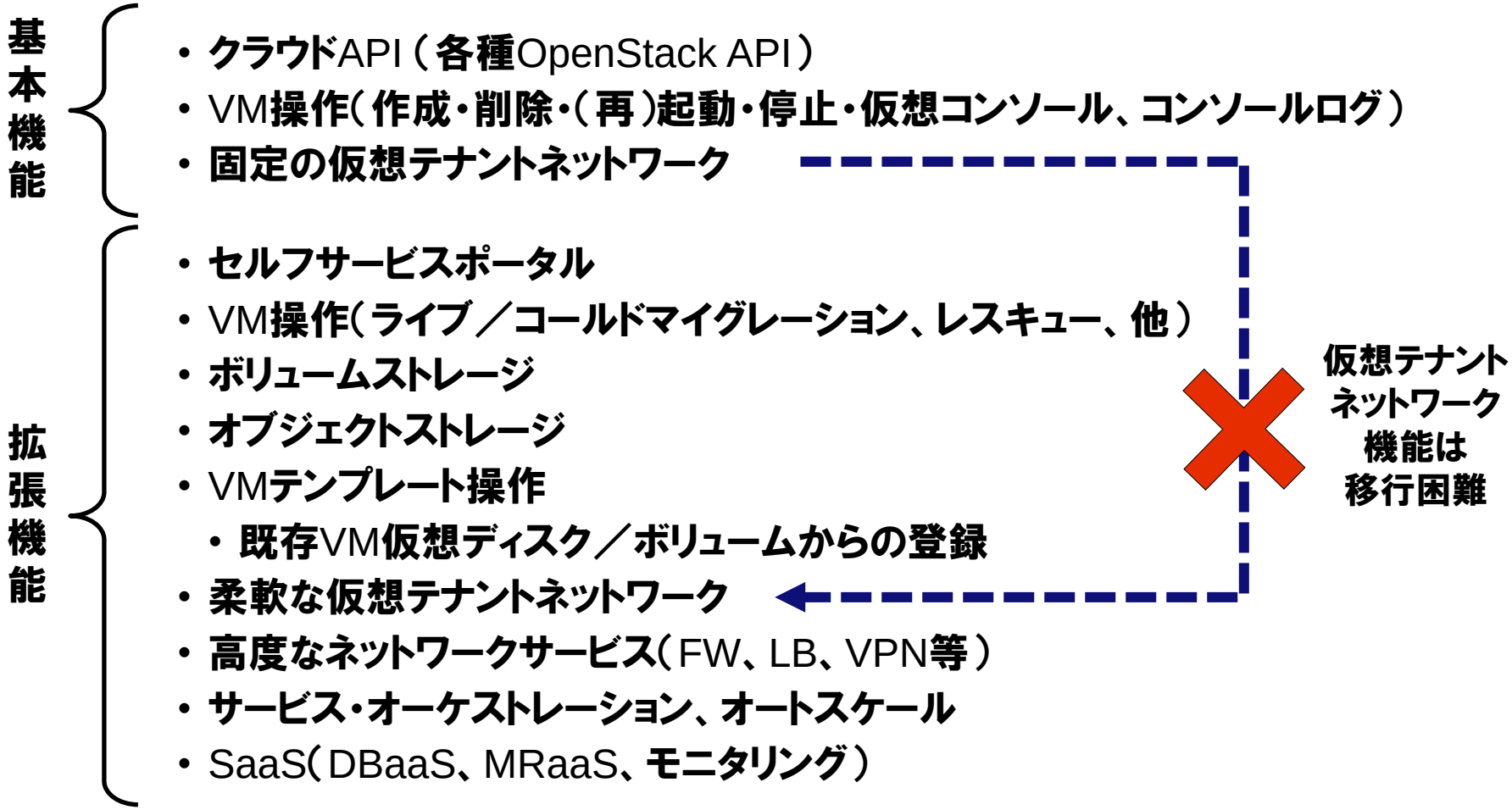
- VMホストサーバ障害時には別サーバでVM起動する
- VMやその上のアプリケーションの死活監視もサービス提供者が実施する

低SLA＋α型：基本的にユーザが障害対応

- 「クラウド基盤障害時、最悪VMやそのデータが消失する可能性がある」という事をサービスの大前提とする
  - ユーザにはVMが消失しても困らない冗長システム設計をしてもらう
- 高SLA化オプションを有償で提供する
  - 必要ならユーザがVMのバックアップを取る
  - 必要ならユーザが高SLAのストレージ（ボリューム）やDBaaSを利用する
  - 必要ならユーザがVMやその上のアプリケーションの死活監視を行う

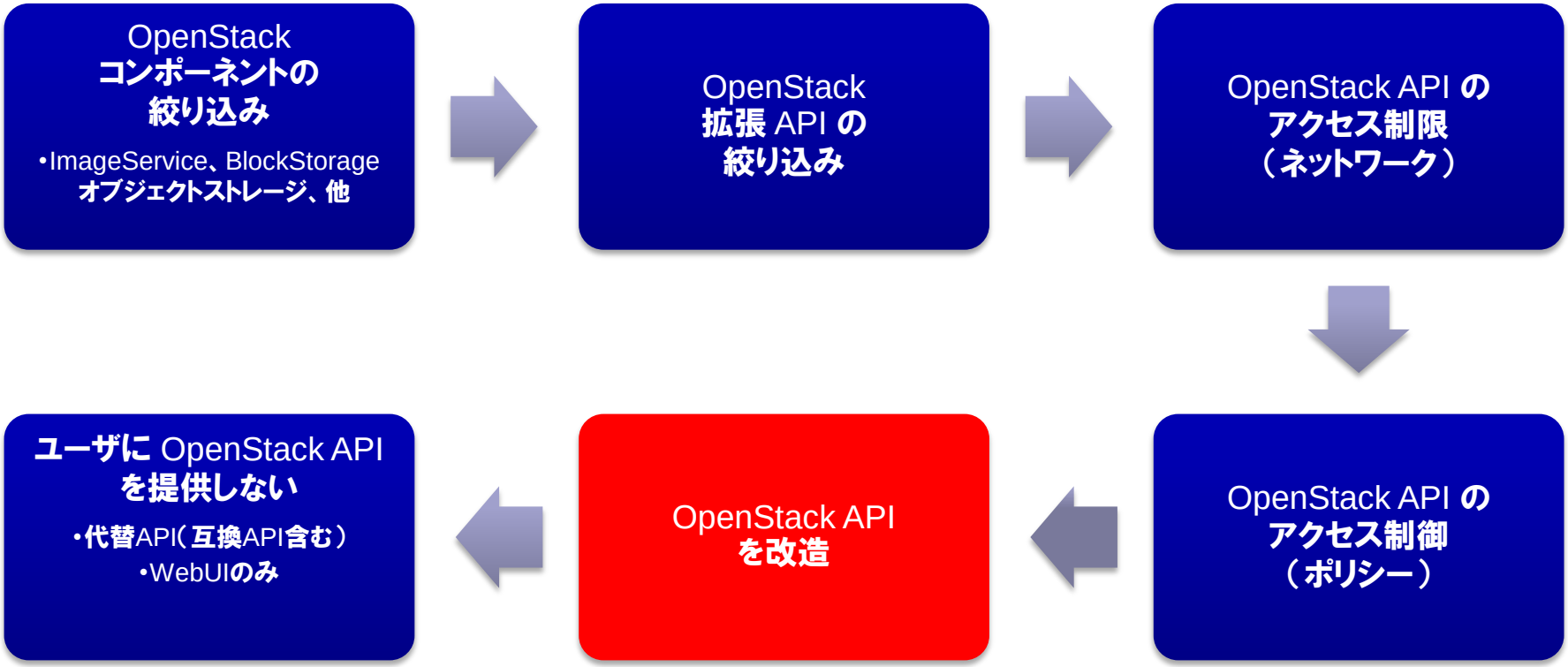
クラウドサービスの機能

OpenStack は機能豊富。サービスで使用する機能範囲の見極めを。
「小さく産んで大きく育てる」は重要。但し拡張困難な部分あり



機能の絞り込み

設定で出来る範囲の当該REST APIのアクセス禁止が基本。
古い OpenStack API はバリデーションが不十分。今後に期待



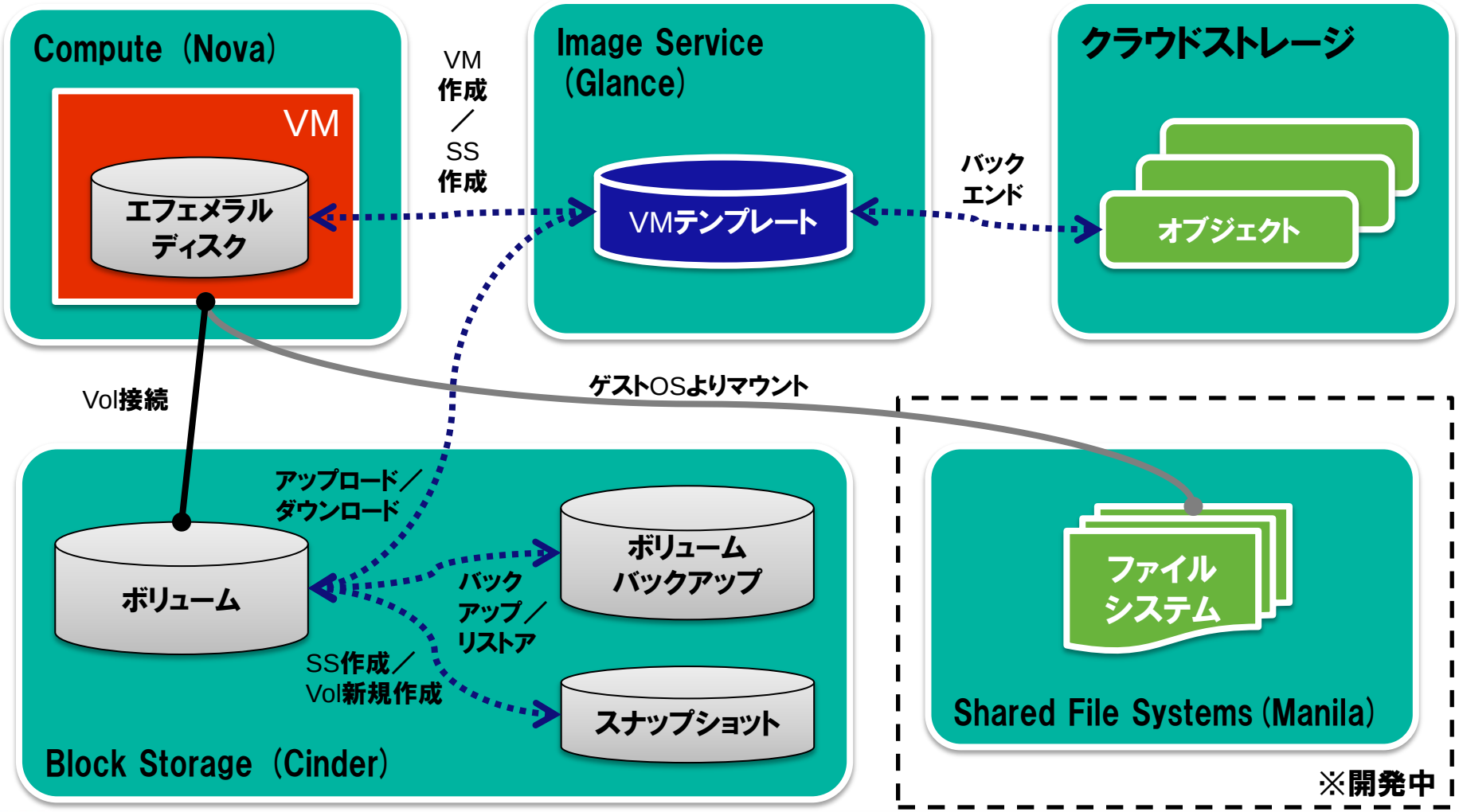
ハイパーバイザ

Computeではハイパーバイザ毎にサポート機能の差異あり

	ドライバの種類	備考
①	KVM、QEMU	OpenStack の標準。対応機能・実績共に最多。
②	XenServer/XCP	①と同等の対応機能の多さ。
③	VMware Hyper-V PowerKVM	①②ほどではないが比較的対応機能が多い。 要件次第では有利な場合あり。 ※VMware は商用ライセンス・商用SDNが必要
④	Xen/Libvirt	SUSE や Oracle が提供する商用 OpenStack で採用
⑤	LXC/libvirt Docker	コンテナベースであり、ゲストOSが限定される。 また対応機能も比較的少なく、充分にテストされていない
⑥	Baremetal	古い物理マシン用ドライバ。廃止予定
⑦	Ironic	新しい物理マシンサービス用ドライバ （OpenStack Bare Metal）

ストレージ

OpenStack ではストレージの種類が多い



ストレージバックエンド

Compute	Image Service	Block Storage（ボリューム、スナップショット）	Block Storage（ボリューム バックアップ）	クラウドストレージ
VMホストサーバ内蔵ストレージ	サーバ内蔵ストレージ	（サービス提供しない）	（サービス提供しない）	（サービス提供しない）
NFS等の共有ファイルシステム	HTTP (Web)	商用ストレージ装置	OpenStack Object Storage	OpenStack Object Storage
Ceph (RBD)	OSS分散ストレージ（Ceph (RBD) 等）	OSS分散ストレージ（Ceph (RBD) 等）	Ceph (RGW)	Ceph (RGW)
	OpenStack Block Storage	その他（Linux iSCSI、NFS等）	IBM TSM	RiakCS
	OpenStack Object Storage			
	GridFS			
	Amazon S3			

ストレージ・サービスの組み合わせ例

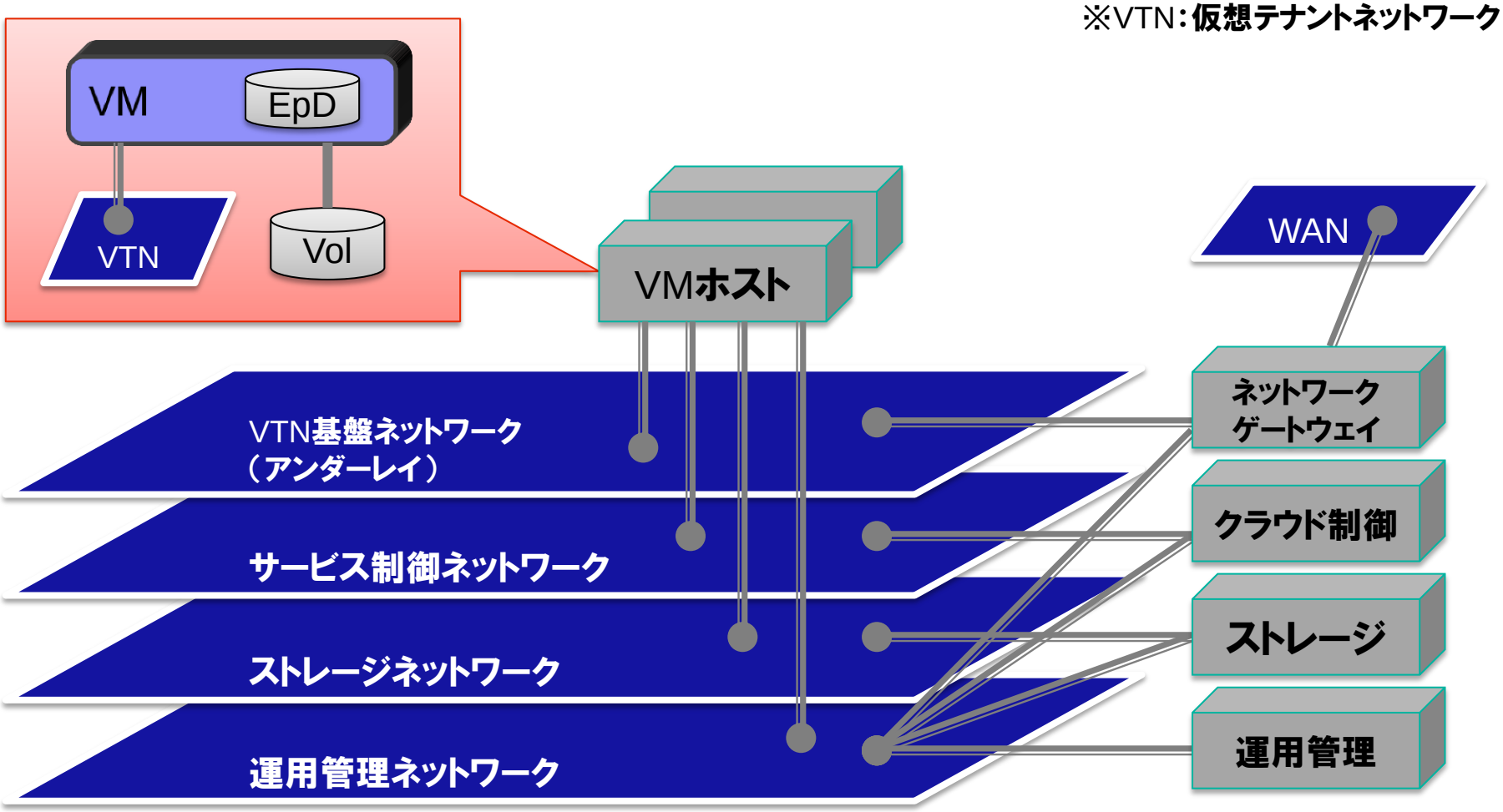
SLAと密接に関わる部分。他社サービスを参考に検討を。
ストレージのスケーラビリティとVM上のディスクI/O帯域制限が重要

	Compute	Block Storage（ボリューム、スナップショット）	Block Storage（バックアップ）	Image Service	備考
①	内蔵ストレージ	有	Object Storage	有	大手パブリッククラウドで見られる構成。AWSの構成に近い
②	商用NFSサーバ	商用ストレージ装置	-	有	ディスクがボトルネックにならない為の設計・構築上の配慮が重要
③	Ceph (RBD)	Ceph (RBD)	Ceph (RGW)	Ceph (RBD)	最近OpenStackコミュニティで注目されている構成。
④	商用NFSサーバ	-	-	有	ライブマイグレーション可能。VMホスト障害でもデータ消失せず。ディスクがボトルネックにならない為の設計・構築上の配慮が重要
⑤	Ceph (RBD)	-	-	有	ライブマイグレーション可能（Juno以降）。VMホスト障害でもデータ消失せず。Cephクラスタ必要
⑥	内蔵ストレージ	-	-	有	内蔵RAIDを使っても運用は大変

ネットワーク

クラウドサービス基盤のネットワークは種類が多い
環境毎の要件に合わせてLANの分割・共用化を

※VTN：仮想テナントネットワーク



仮想テナントネットワークのバックエンド

SDN構成推奨。ネットワーク性能・管理面でメリット大

	種類	備考
①	Networking + **商用**SDN**ベース**	**ネットワークの性能面・管理面でのメリットあり。但しコストはかかる**
②	Networking + OSS SDN**ベース** ※OpenContrail、OpenDayLight、MidoNet **等**	**ネットワークの性能面・管理面でのメリットあり。ゲートウェイでネットワーク製品が利用可能なもの有り。**
③	Networking **ベース（**SDN**無し）** ※OVS, LinuxBridge **等**	Networking **の管理サーバがボトルネックになりやすく、スケーラビリティに課題**
④	nova-network **ベース（**VLAN**等）**	**フラットな**IP**アドレス管理。テナント毎に1つの**VTN**、シンプルな**FW**機能のみ。マルチホストなど便利な点もあるが、**OpenStack Networking (Neutron) **への移行方法が未だ提供されていない**

クラウド基盤開発・保守

商用サービス基盤では開発環境・評価環境・本番環境が必要

開発環境

- 新機能・バグ修正パッチ・テストケース開発
- ソフトウェア設定変更・バージョンアップ検証
- 自動プロビジョニングツール（Chef等）の構成設定（レシピ）開発

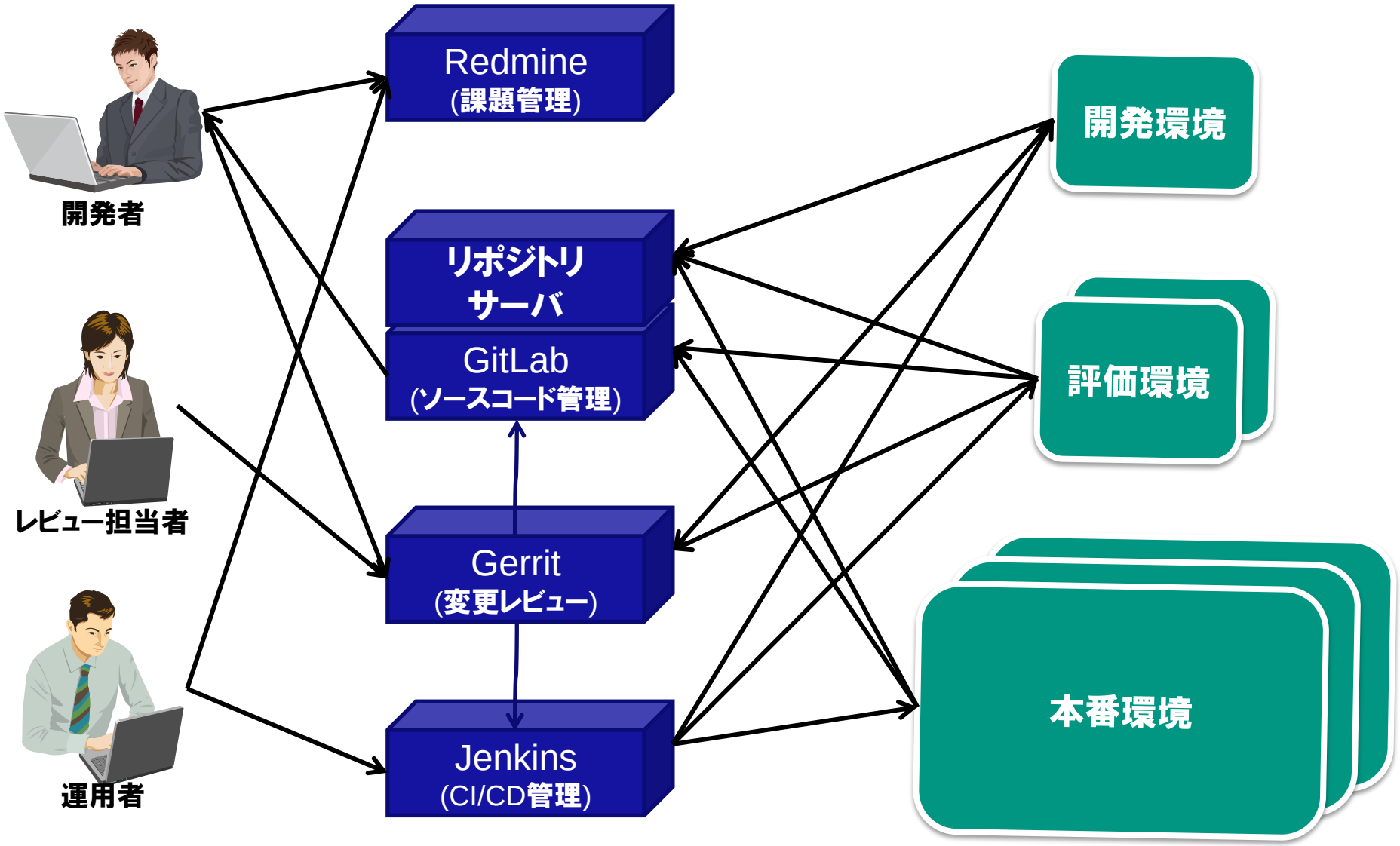
評価環境　※本番環境と同じハードウェア構成

- 開発環境した諸々の変更内容適用後の機能・性能テスト
- 本番環境への変更適用・切り戻し作業リハーサル

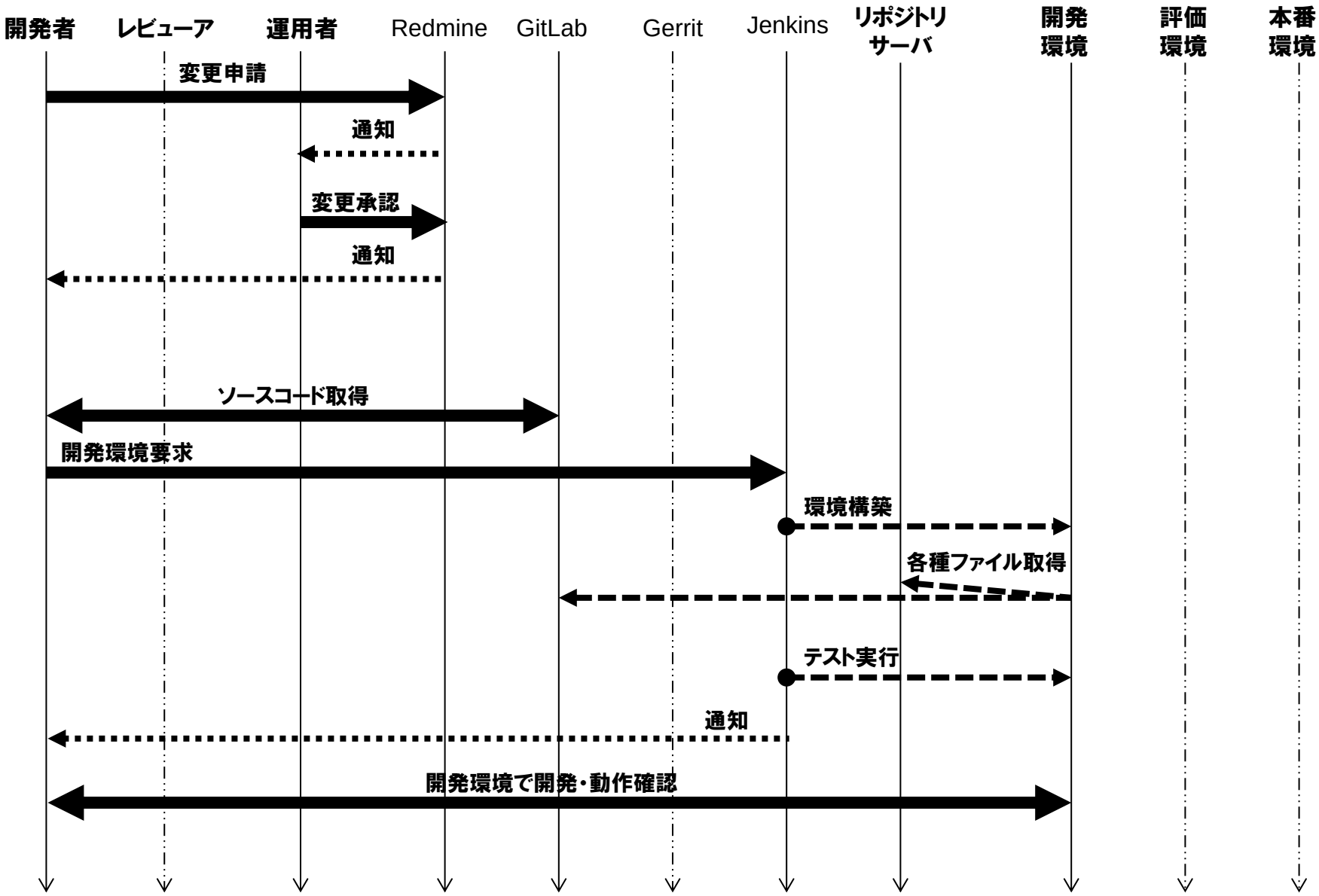
本番環境

- 商用サービス提供
- 定期的な自動テスト実行（正常性確認）

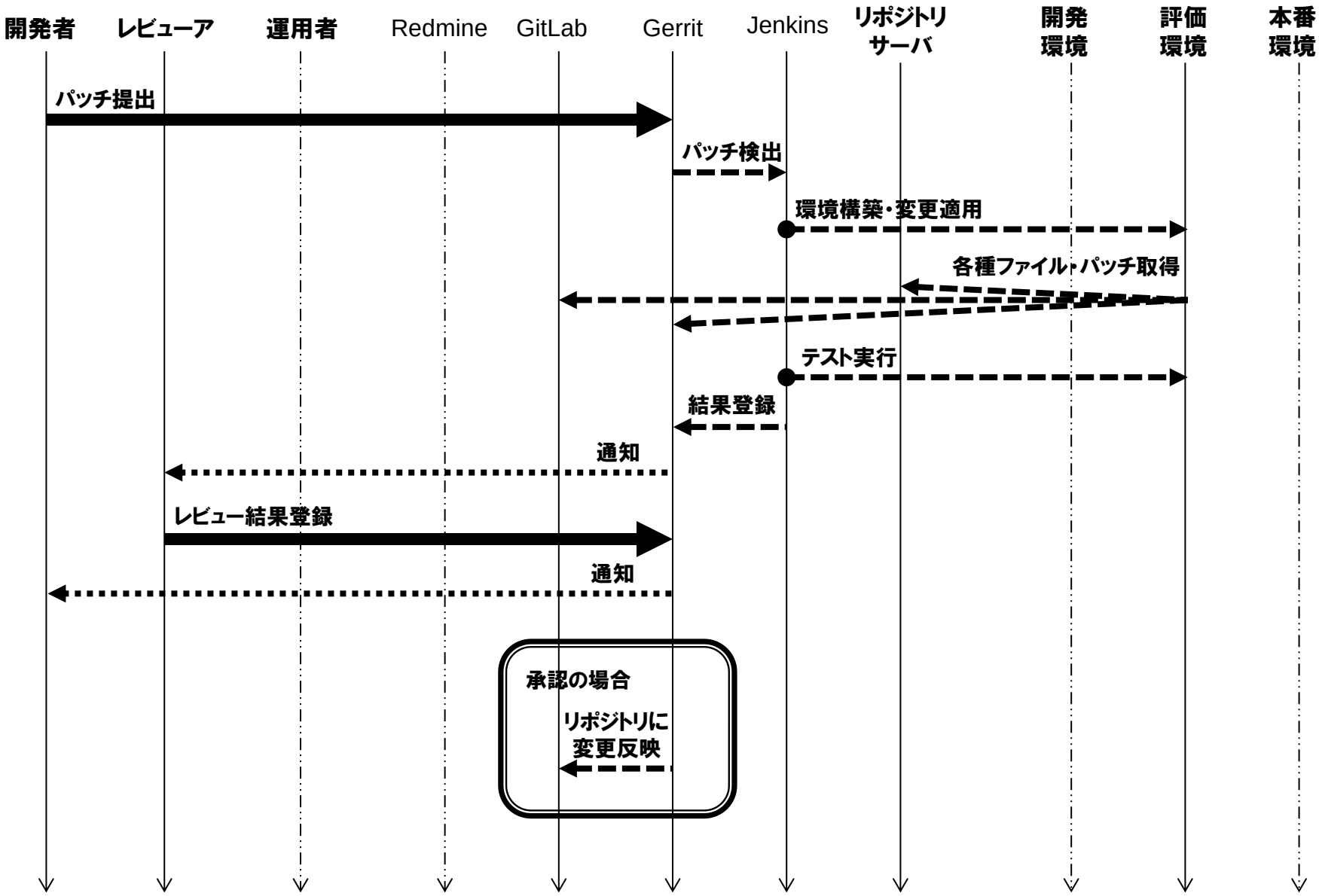
ソフトウェア管理のシステム構成例



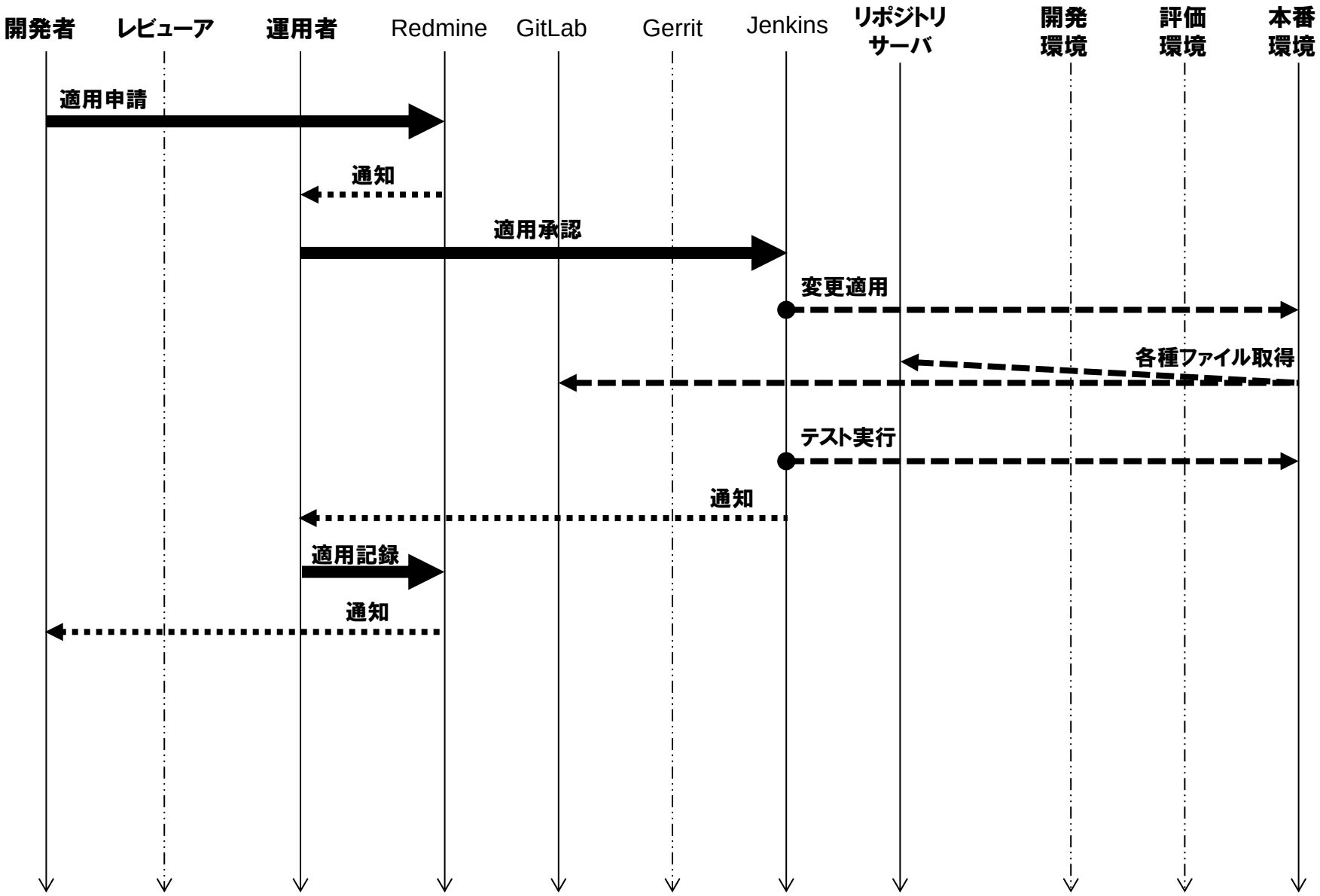
ソフトウェア変更ワークフローの一例（開発）



ソフトウェア変更ワークフローの一例（開発～評価）



ソフトウェア変更ワークフローの一例（本番適用）



自動プロビジョニングツール

各種OpenStack インストーラで内部利用されているケース有り

名称	Agent	OSインストール	設定言語	Pull／Push	備考
Puppet	有	不可	独自	Pull	古参のツール。PackStack、FUEL、eDeploy 等は内部で Puppet を利用している。
Chef	有	不可	独自	Pull (Push)	Puppet より後発のツール。元々はプル型だが、最新版や商用版ではプッシュも可能。Crowbar、Compass 等は内部で Chef を利用している
Ansible	無 (sshd)	不可	YAML	Push	Chef より後発のツール。学習が容易。プッシュ型の為、順序だったシステム構築が可能。スケーラビリティは比較的低い
Salt	有	不可	YAML	Push	Ansible よりスケーラビリティがある。OpenStack コミュニティではややマイナー。
Juju/ MAAS	有	可	各種スクリプト	Push	MAAS と組み合わせる事で、OSインストールからシステム構築まで可能。設定の冪等性を開発者がスクリプト内で保証する必要あり。

自動テストツール

Tempest・Grenade は開発・評価環境用、Rally は評価・本番環境用

Tempest

- OpenStack **の結合テストツール**
- OpenStack **正式プロジェクトの1つ**
- OpenStack **開発ではパッチ毎に** Tempest **によるテストが自動実行される**
- OpenStack **の各ソフトウェアの正常系テストと異常系テストの両方を多数持つ**
  →本番環境の正常系試験ツールとしてはあまり適していない

Grenade

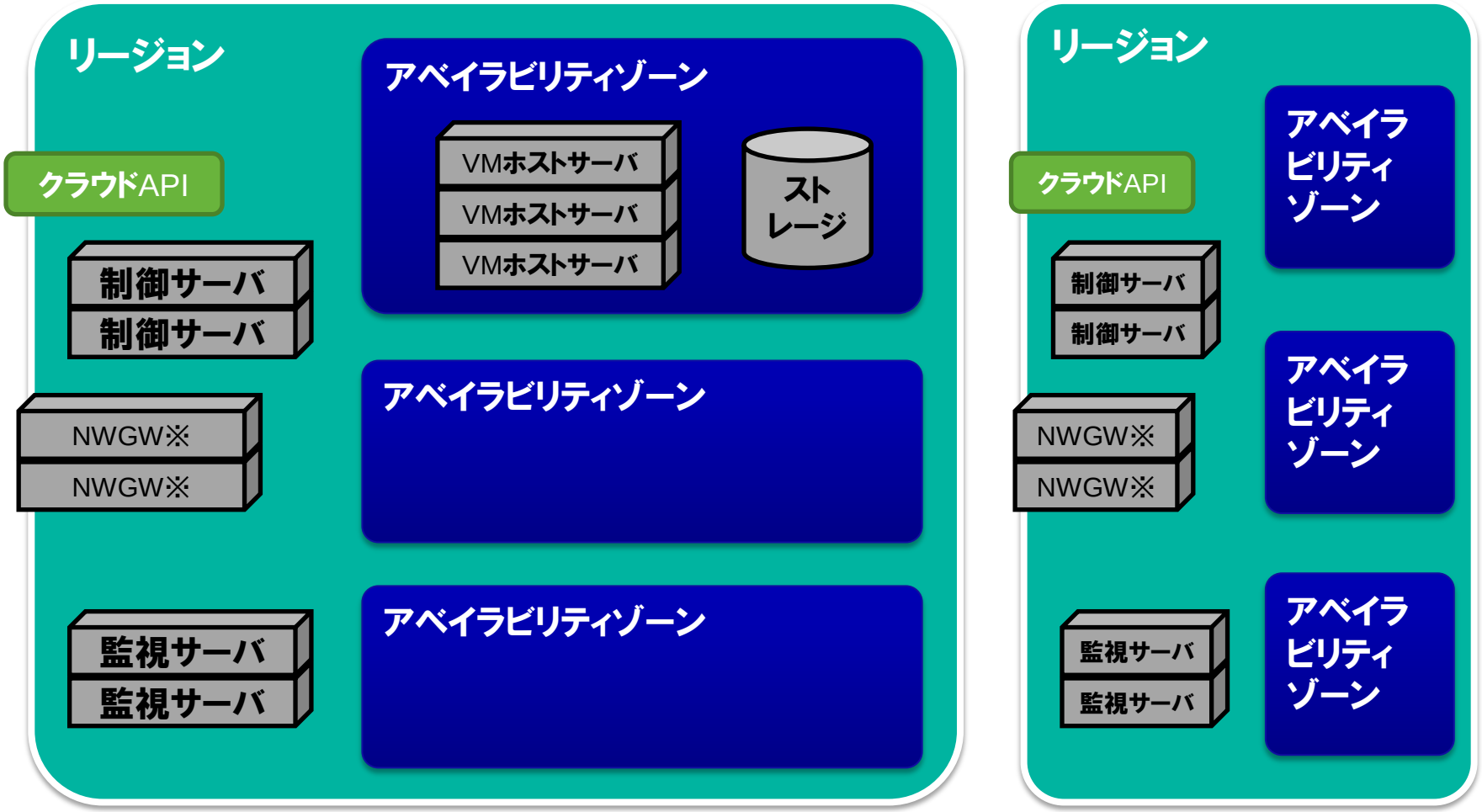
- OpenStack **のアップグレードテスト・ツール**
- OpenStack **開発ではパッチ毎に** Grenade **によるテストが自動実行される**

Rally

- OpenStack **のベンチマークツール**
- **まだ** OpenStack **正式プロジェクトではないが、最近の正式プロジェクトで** Rally **用のテストケースが用意されたものが出てきた**
- OpenStack **主要ソフトウェアの正常系テスト・ツールとして注目されている**

クラウド基盤拡張

OpenStack におけるクラウド基盤拡張のパターンは3種類



※NWGW：ネットワークゲートウェイ

OpenStack バージョンアップ

「OpenStackバージョンアップ＝新リージョンデプロイ」が無難

既存の OpenStack 環境をコミット毎にバージョンアップ

- OpenStack 各プロジェクトのパッチコミットマージ毎に反映（DevOps的）
  ⇒大きな変更が減る一方、頻繁にリリースを繰り返す事になり、負担大

既存の OpenStack 環境をリリース毎にバージョンアップ

- OpenStack は半年毎にメジャーバージョンアップがある
  ⇒半年毎にリリースを行う事になるので、それなりに負担大
- OpenStack は1バージョン前からのマイグレーションのみ対応
  ⇒OpenStack プロジェクトで検証済みのバージョンアップが可能

OpenStack 新バージョンの新リージョンをデプロイする

- 既存の OpenStack 環境はバージョンアップしない
- 半年毎の OpenStack リリースを追いかけなくても良い
- リージョン毎のサービス寿命を定義しておく事が重要

クラウド基盤監視

監視対象が多いので、監視しすぎない事が重要
必須監視項目は「サービスが正常動作しているか」

制御サーバ

- PING・プロセス死活・各種ログ
- 各種性能情報
  - クラウドAPI応答時間
  - DBクエリ時間
  - メッセージキュー長

ネットワークゲートウェイサーバ

- PING・プロセス死活・各種ログ
- 外部ネットワーク疎通確認

ストレージ

- PING・障害通知・各種ログ
- ディスク使用量

VMホストサーバ

- PING・プロセス死活・各種ログ
- メモリ・ディスク使用量

まとめ

商用サービスにおける OpenStack の勘所

VM の SLA について明確なポリシーを定める

- 高SLA型　／　低SLA＋α型

上記ポリシーに則したクラウド基盤を設計する

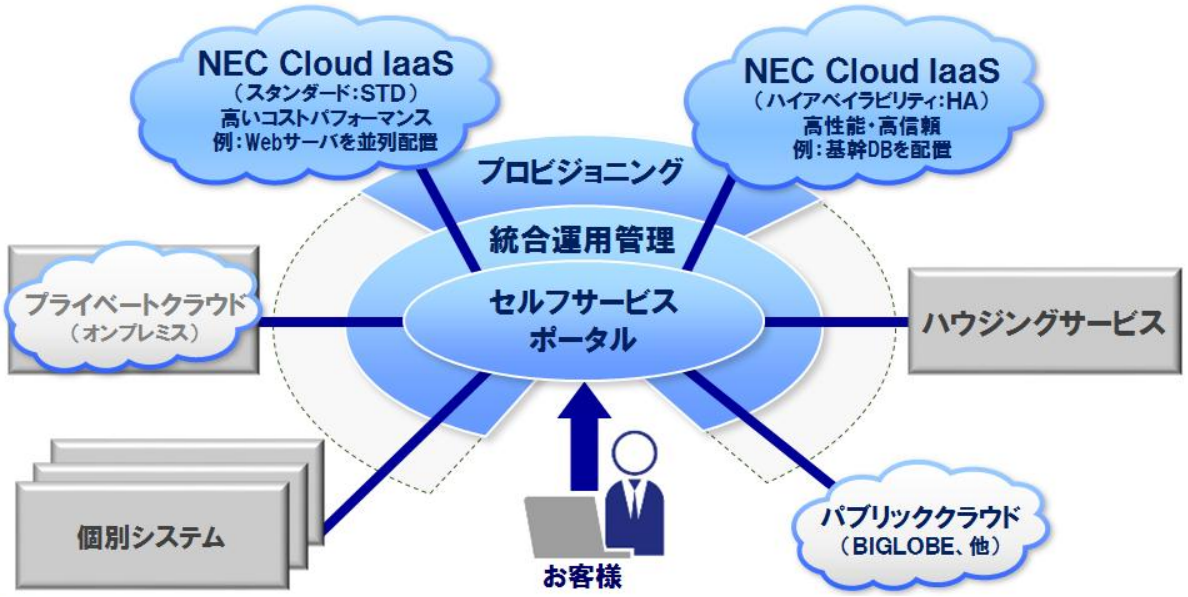
- OpenStack はストレージやネットワークのバックエンドの選択肢が多い
- 組み合わせは他社サービスを参考に
- クラウドサービス＝セルフサービス

クラウド基盤の構築・運用・保守は可能な限り自動化する

- クラウド基盤＝大規模システム。手動構築・保守は破綻する
- クラウド基盤構築・運用・保守は連続している
- クラウド基盤開発・保守で既存のテスト・ツールを活用する

さいごに

OpenStackを活用したNEC Cloud IaaSを今後も発展させていきます



NEC Cloud IaaS

Orchestrating a brighter world

世界の想いを、未来へつなげる。

未来に向かい、人が生きる、豊かに生きるために欠かせないもの。
それは「安全」「安心」「効率」「公平」という価値が実現された社会です。

NECは、ネットワーク技術とコンピューティング技術をあわせ持つ
類のないインテグレーターとしてリーダーシップを発揮し、
卓越した技術とさまざまな知見やアイデアを融合することで、
世界の国々や地域の人々と協奏しながら、
明るく希望に満ちた暮らしと社会を実現し、未来につなげていきます。

Empowered by Innovation