

NEW WORLD. NEW NETWORK.

Brocade: Your Partner for the New IP

BROCADE[®]

ヤフーが手掛けるベアメタルOpenStackと ブロード流「自動化」

【本日のゲストは】

西山 貴彦様

ヤフー株式会社

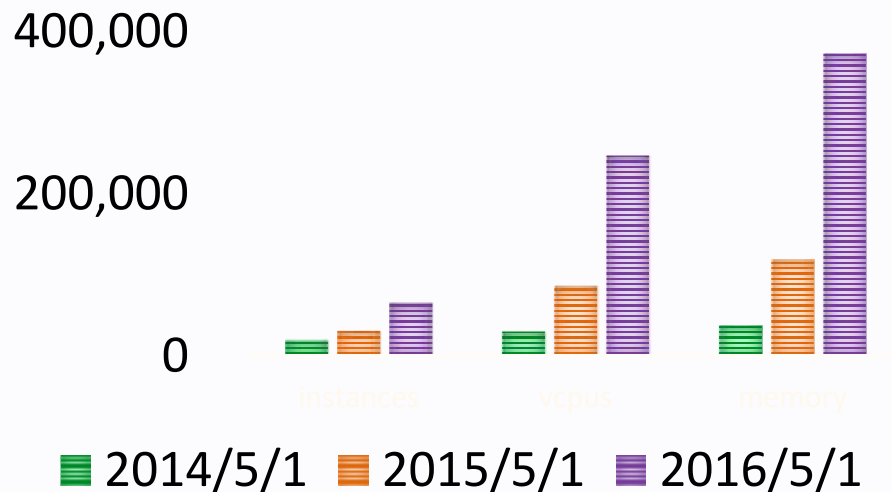
システム統括本部 サイトオペレーション本部



第一部 バアメタル OpenStack

ヤフーさんの現在のインフラ規模
について教えてください

ヤフーでのOpenStack(VM)の規模



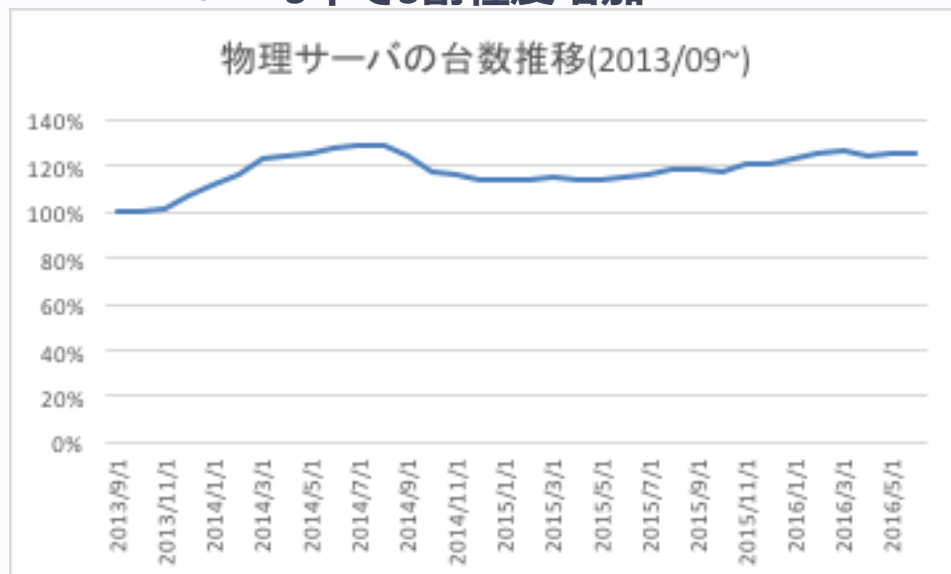
- 前年比でインスタンス数が1.5~2倍程度増えている
- ハイパーバイザの強化により1インスタンス辺りのリソースも増えている

伸び率 (YoY)	2015年	2016年
インスタンス	156%	211%
vCPU	286%	284%
メモリ	316%	312%

仮想化を推進されていますが
物理サーバってこれからも
増えるんですか？

ヤフーでの物理サーバの状況

- 仮想化も進んでいるが物理サーバ台数も拡大している
 - ◆ OpenStackの社内展開開始(2013/09)後も増えている
 - 3年で3割程度増加



物理サーバが増える理由

- VMでは性能が足りないサービスが存在する
 - ◆ CPUリソースが重要なサービス
 - Hadoopなど
 - ◆ ディスクIOが激しいサービス
 - Elastic Searchなど
 - ◆ レイテンシが重要なサービス
 - 広告配信など
 - ◆ 大容量ディスクが必要なサービス
 - オブジェクトストレージのデータノードなど

ベアメタルプロビジョニング
で登場する IroniC って？

About Ironic

- OpenStackにおけるベアメタルプロビジョン用コンポーネント
 - ◆ <https://wiki.openstack.org/wiki/Ironic>
 - ◆ Icehouseでスタートし、Kiloでintegratedプロジェクトに



Ironicで出来ること(1/2)

■ ベアメタルノードの情報管理

- ◆ ハードウェアスペック、macアドレス、ipmi接続情報
 - ipmi接続情報をnode-create時に入力
 - Ipmi = Intelligent Platform Management Interface
 - ネットワーク越しにサーバを直接操作する規格
- ◆ ハードウェアスペックとmacアドレスを筐体から取得
 - Ironic-Inspectorを利用
 - 専用ramdiskでpxebootして取得
- ◆ (Mitakaから)raid構成の管理

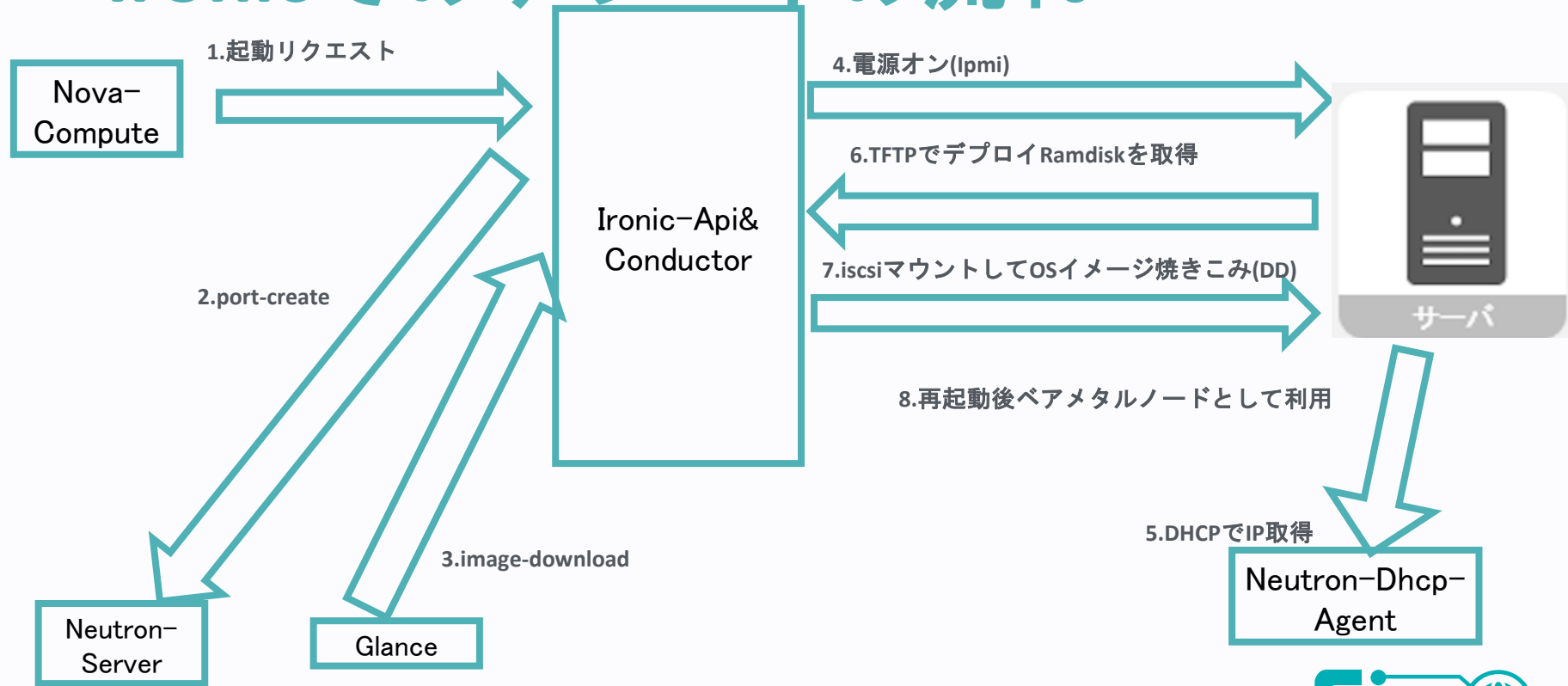
Ironicで出来ること(2/2)

- ベアメタルノードの操作

- ◆ ipmiを利用
- ◆ 電源操作、電源状態管理
- ◆ コンソール接続
 - 但しNova-Compute側のドライバに未実装のためNova-Consoleとは連携できない

- ベアメタルノードへのイメージのデプロイとクリーニング

Ironiicでのデプロイの流れ



従来と比べたメリット

- 使っていないサーバの入れ替えがしやすい
 - ◆ 従来: 廃棄するまで同じ利用者がサーバを用途を変えて使っていた
 - ◆ Ironic: 使うときのみ払いだす = その他のサーバはすぐ変えられる
 - 管理者側でライフサイクルが可能になる
 - ◆ ラックが歯抜けにならなくて済む
- (利用者からみて)利用開始までの時間が早い
 - ◆ 従来: 物理サーバの発注から利用開始まで1ヶ月
 - ◆ Ironic: デプロイするだけなので10分

(参考)ラックの歯抜け問題

- 最初はラックを埋めるように設置される
- サーバのライフサイクルがサービス依存のため捨てるタイミングがずれたり、違う拠点に移設されたりする
- 残されたサーバが歯抜けになってしまう
 - ◆ 他所から移設される場合もあるが限界がある
- ラックが埋まらなくなり、スペースが無駄になってしまう

現在のヤフーさんでの
Ironic 取組み状況は？

ヤフーにおけるIronicの取り組み

- Kiloリリースから検証を開始
- Libertyリリースのクラスタを複数構築
 - ◆ 開発環境1クラスタ
 - ◆ プロダクション環境4クラスタ
 - PaaSのバックエンドやElastic Searchなどで利用中
 - 1クラスタ数百台規模

ネットワークまわりの要件と課題は？

ネットワーク周りの課題

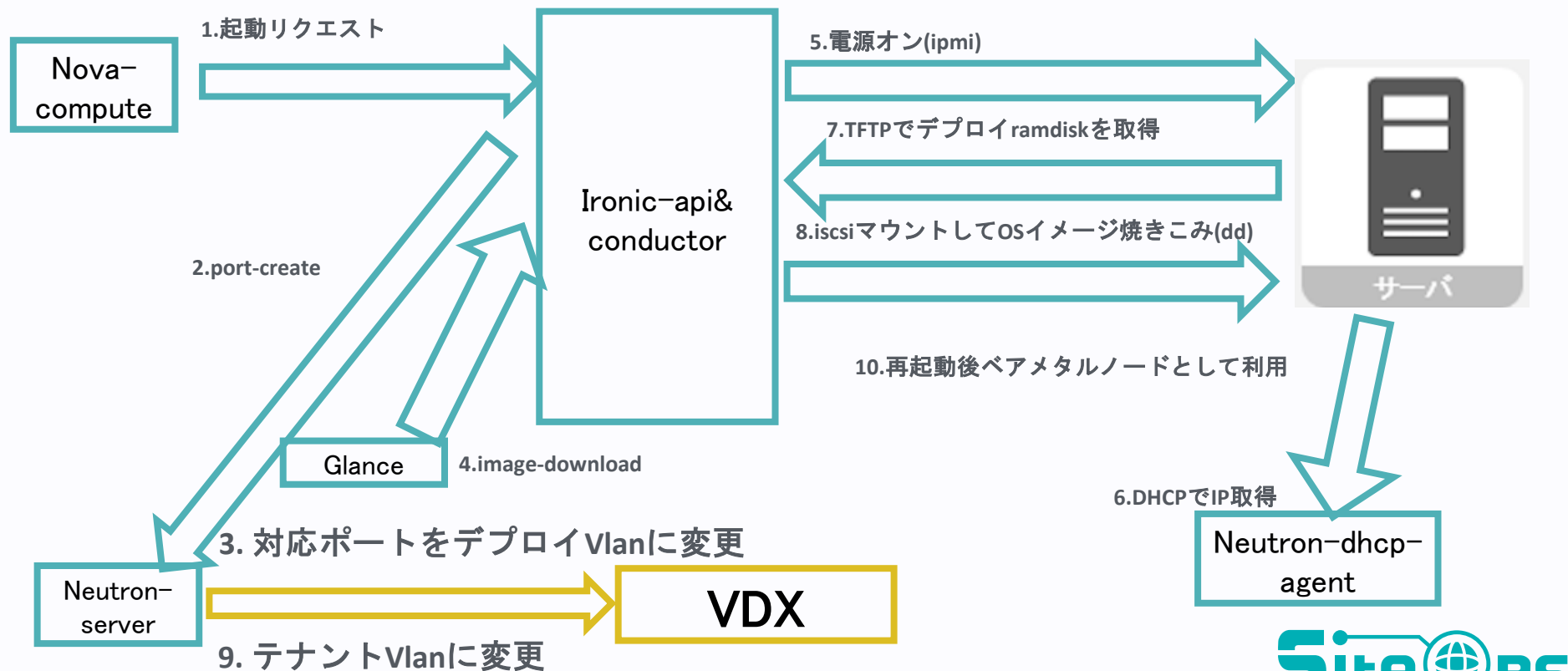
- 事実上flatの単一ネットワークでしか使えない
 - ◆ サーバがつながっているポートのVlanが変えられない
 - ◆ ノードに対するタグ付けもできない
 - ◆ 複数NWを用意してもデプロイ出来ない可能性がある
 - スケジューラからはノードに対応するVlanが分からない
 - 違うVlanを設定したノードでデプロイを始めてしまう可能性
 - 通信できないのでタイムアウト -> エラーになる
- ヤフーでは1クラスタ1Vlanとすることで回避している
 - ◆ クラスタが乱立する

今、どんなものを
検証されていますか？

IroniC Brocade VDX Plugin

- 単一のflatネットワークからの解放
- エッジスイッチのVlanをIroniCと連動して変更
 - ◆ デプロイ完了まではデプロイ用Vlan
 - ◆ デプロイが終わったらテナント用Vlanに切り替え
 - 再起動するとテナント用Vlanで疎通できる
- ネットワークごとにクラスタをわけなくて良くなるはず
- Mitakaから利用可能

デプロイの流れ(with VDX)



ここで、大事なものは？

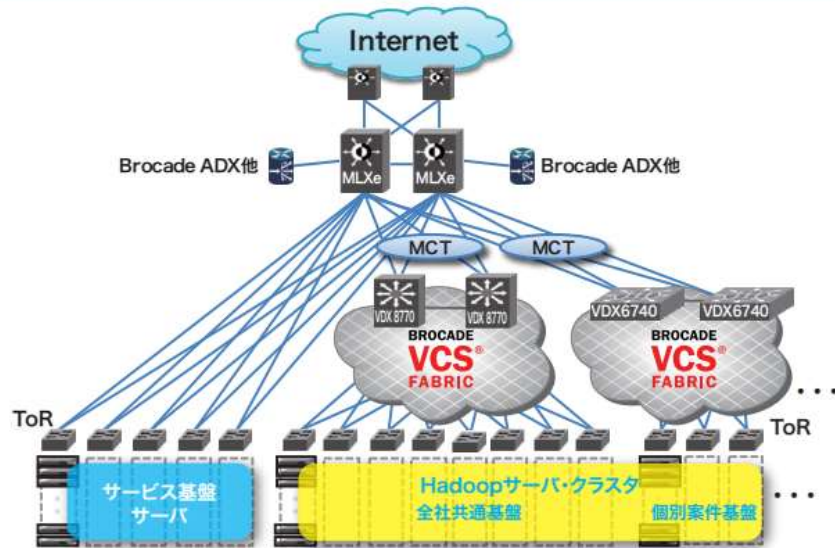
つ・な・が・る・こ・と

ヤフーさんのデータセンターを支える ブロードのファブリックスイッチ VDX

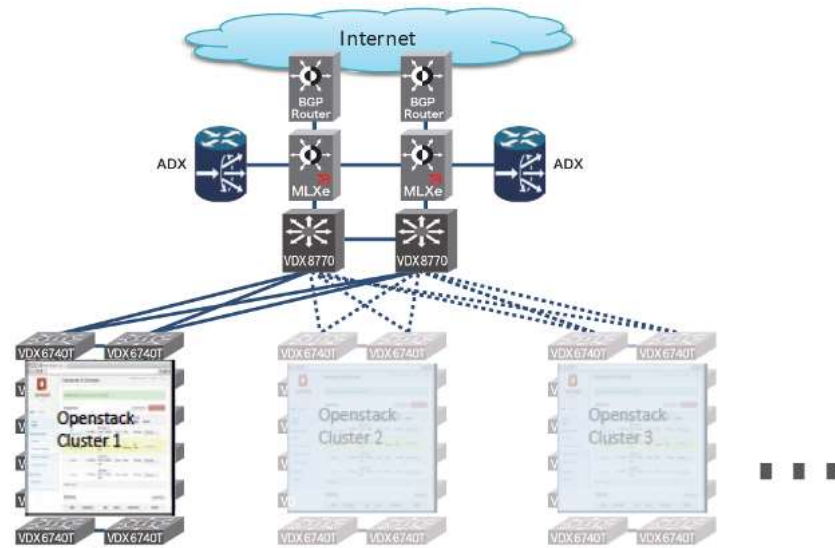


国内最大級のHadoop基盤を構築し
ビッグデータ活用によるサービス強化を加速

Neutronプラグインにより最適化されたネット
ワークを構築し、LBaaS とFWaaS を実現



■ ヤフー株式会社のデータセンター・ネットワーク構成図



■ ヤフー株式会社の OpenStack プライベートクラウド システム構成図

つ・な・が・る

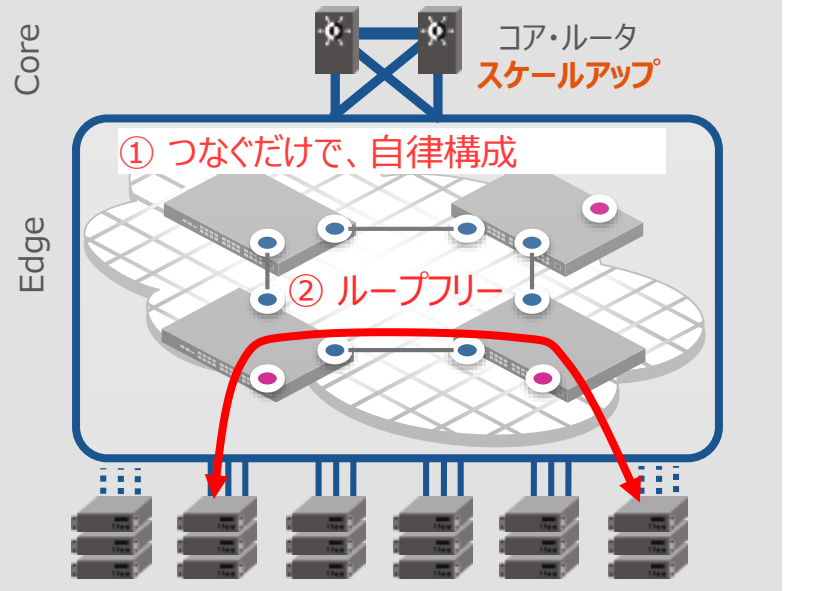


箱を開けて、つなぐだけ



イーサネット
ファブリックスイッチ
VDXシリーズ

イーサネット・ファブリック



- ③ 設定の必要がないポート ④ 最短経路を選択

2 台ではじめる OpenStack!



ファブリック
スイッチ



サーバ

スモールスタートに最適な帯域販売

1. 48ポート1GEからスタート
2. 8ポートだけ10GEにアップグレード
3. ストレージ用にFCポートに変更
4. アップリンクに40GEも追加
5. サービスを止めずにオンデマンド!

VDX6740T-1G



今後の展望（野望）について
お聞かせください。

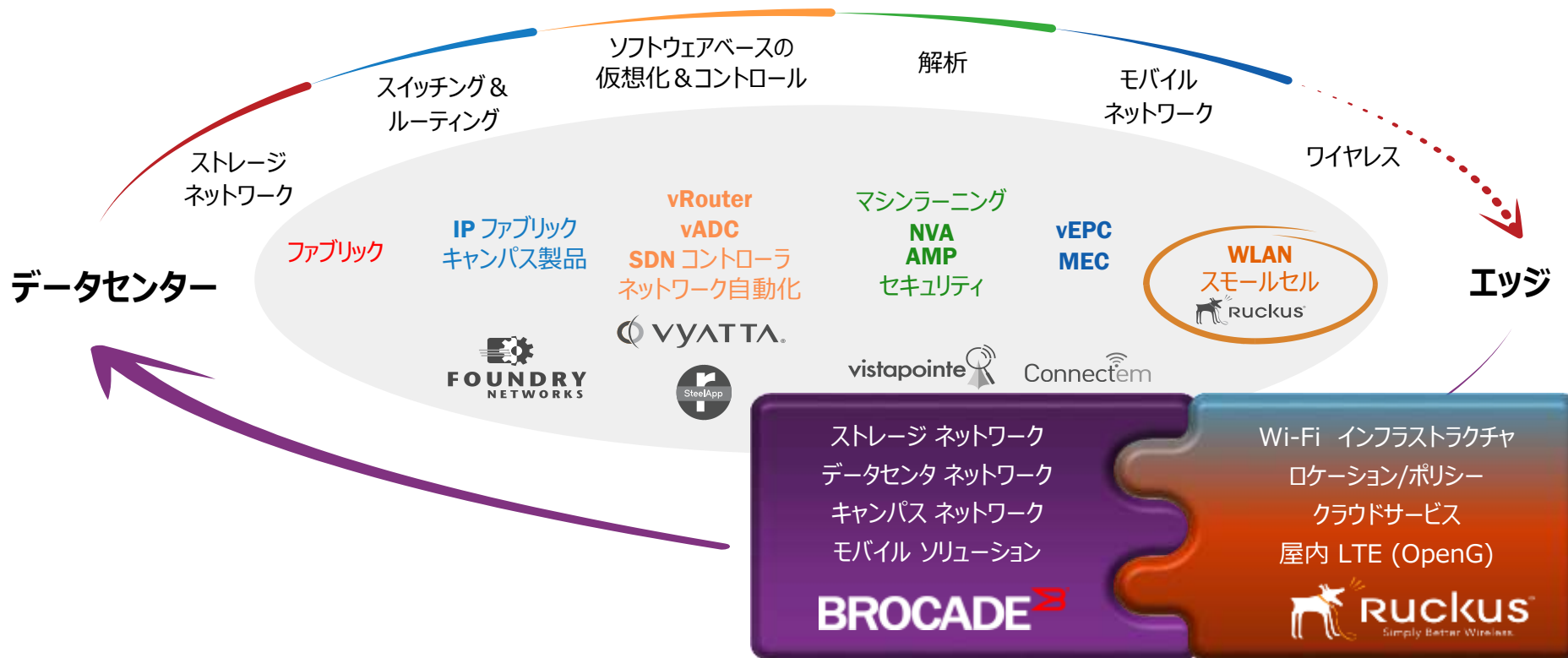
今後の展開

- 社内のすべての環境をOpenStackに載せる
 - ◆ 社内の物理サーバのIronic比率を高める
 - 最終的にはコントローラノード以外は全てIronicにできるはず
 - ◆ 全てをOpenStack化することで同じAPIで物理サーバもVMも使えるようにする
- Ironicにすることでハードウェアのライフサイクルも進めやすくする
 - ◆ サービスアウトしたノードから取り替える

第二部 Brocade の新しい取組み

ユーザーに、より近い場所からサービスを提供

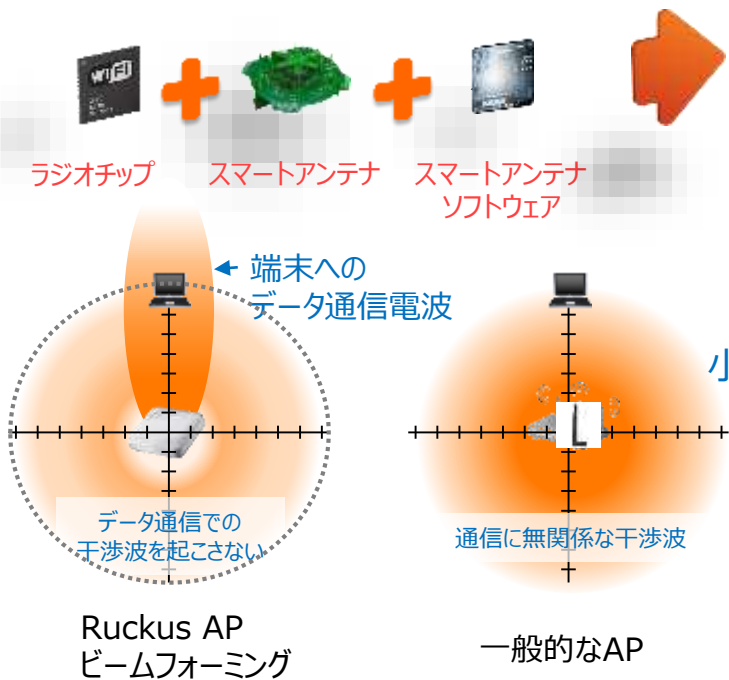
データセンターからワイヤレス・ネットワーク・エッジまで



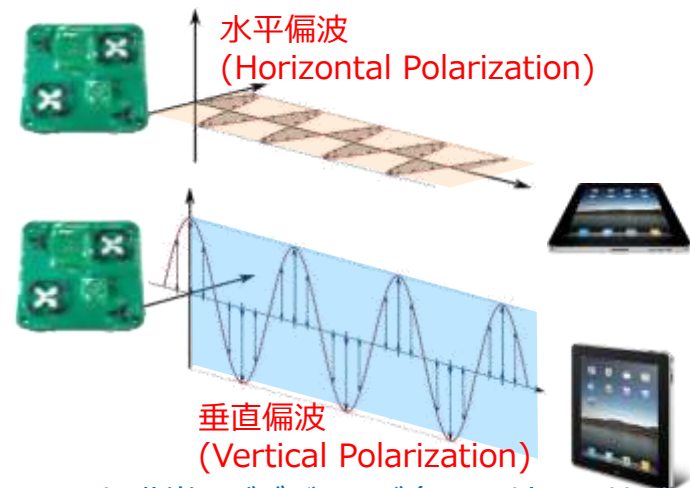
Ruckus 無線LANの特徴 (アンテナ技術)

アンテナ技術の違い (Beam Flex)

- ・スマートアンテナが、端末場所を認識後、データ通信に最適なアンテナパターンを使用



アンテナ技術の違い (PD MRC)

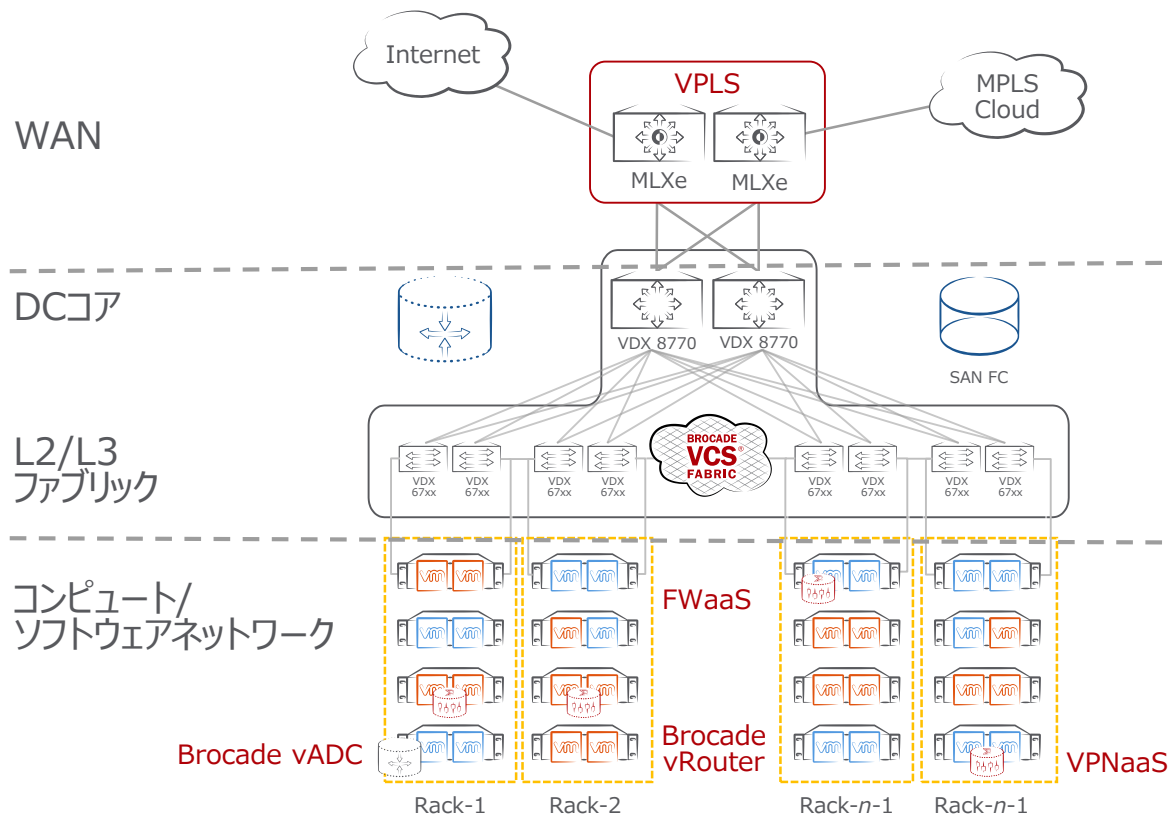


小中学校の教室・オフィスや、ホール・大講堂などデバイスが多い環境でも快適

1. クライアントからのフィードバックにより送信シグナルを最適化
2. デバイスの向きを気にすることなく、最適な通信を提供
3. 通信デバイスの反対方向は最小限の電波干渉



データセンターネットワークの自動化



	Juno Oct 2014	Kilo Apr 2015	Liberty Oct 2015
		MLX ML2 plugin	MLX ML2 plugin
VCS/VDX SVI		VDX and ICX ML2 Plugin VDX SVI & ACLs	VDX: Private L2 extensions
SAN FC		ADC LBaaS SAN FC	ADC LBaaS SAN FC Zone alias & vF
vRouter*		vRouter	vRouter
		VPNaaS	VPNaaS
		FWaaS	FWaaS
vADC* (LBaaS)		vADC (LBaaS)	vADC (LBaaS)

- * Available as private plugin.



次世代プラットフォームのためのネットワーク

BROCADE[®]

大きな変化をとげる転換期

The New IP

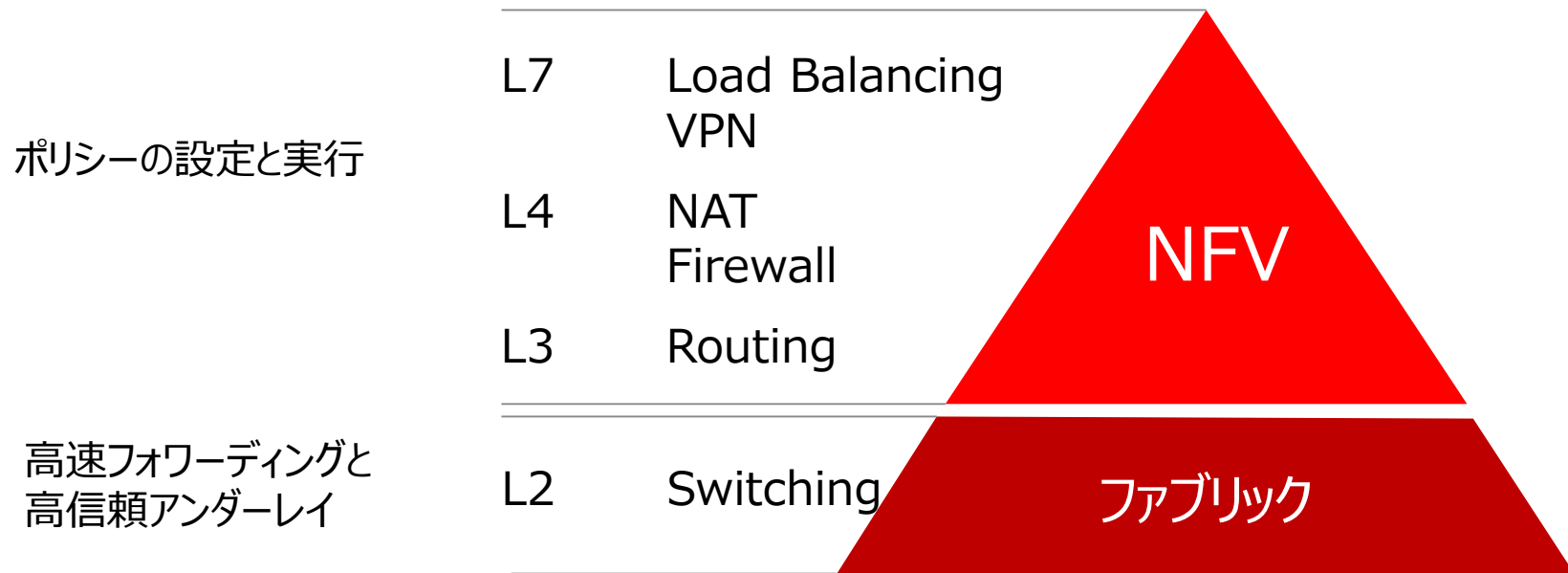


OPEN

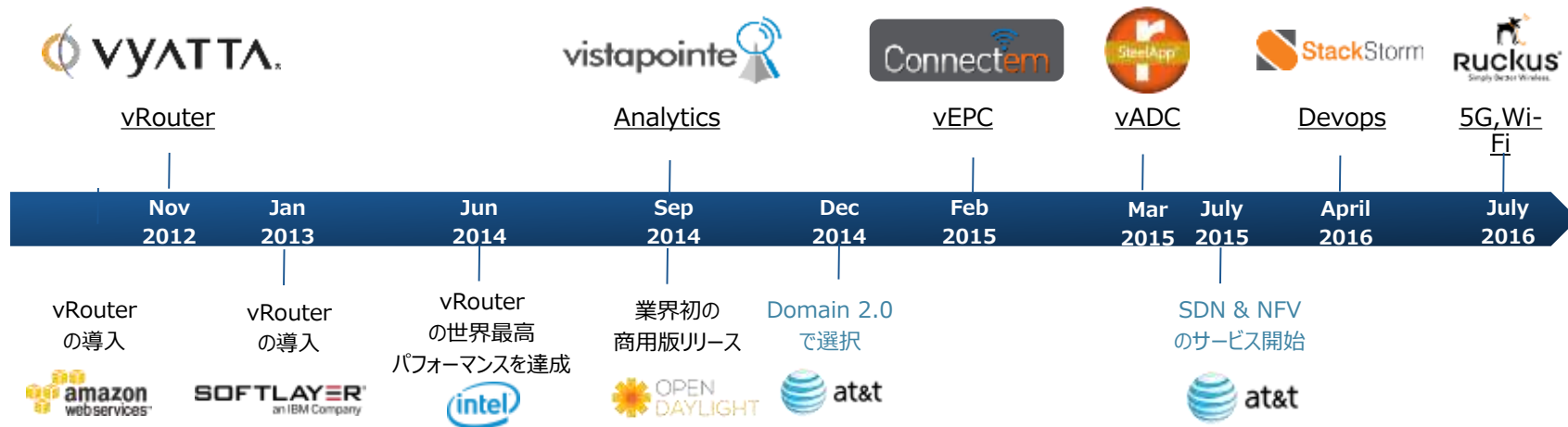
1. マルチベンダのオープンな環境
2. APIにより物理の縛りからの解放
3. 容易なスケールアウトと高速デリバリ
4. ユーザ自身がつくれるネットワーク

The New IP のエッジには NFV が最適

あらゆるお客様、アプリケーション、サービスのニーズに対応



SDI/NFV/Mobile におけるリーダーシップ



なぜ BROCADE なのか?



データセンター
ネットワーク
N.O.2



ソフトウェア
ネットワーキング
リーダーシップ



エンタープライズ
クラウド
NFV



オープン
アーキテクチャ



イノベーション



パートナーとの協業

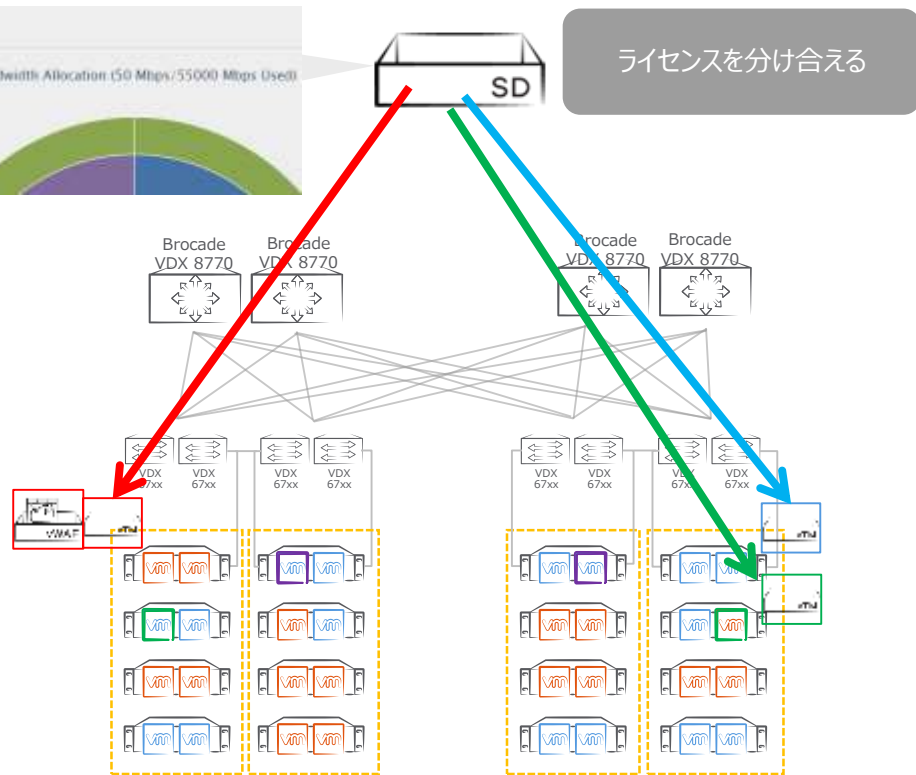


ソリューション

VNFライセンスの統合管理(LBaaS)

vADC Service Director

- Brocade Service Director
- ADCインスタンス
 - インスタンスをオンデマンドで用意
 - 複数Versionを使い分け
- ライセンス
 - 規模に応じたライセンスをオンデマンドで配備
 - WebAppFW/機能もオンデマンドで
- メータリングとロギング



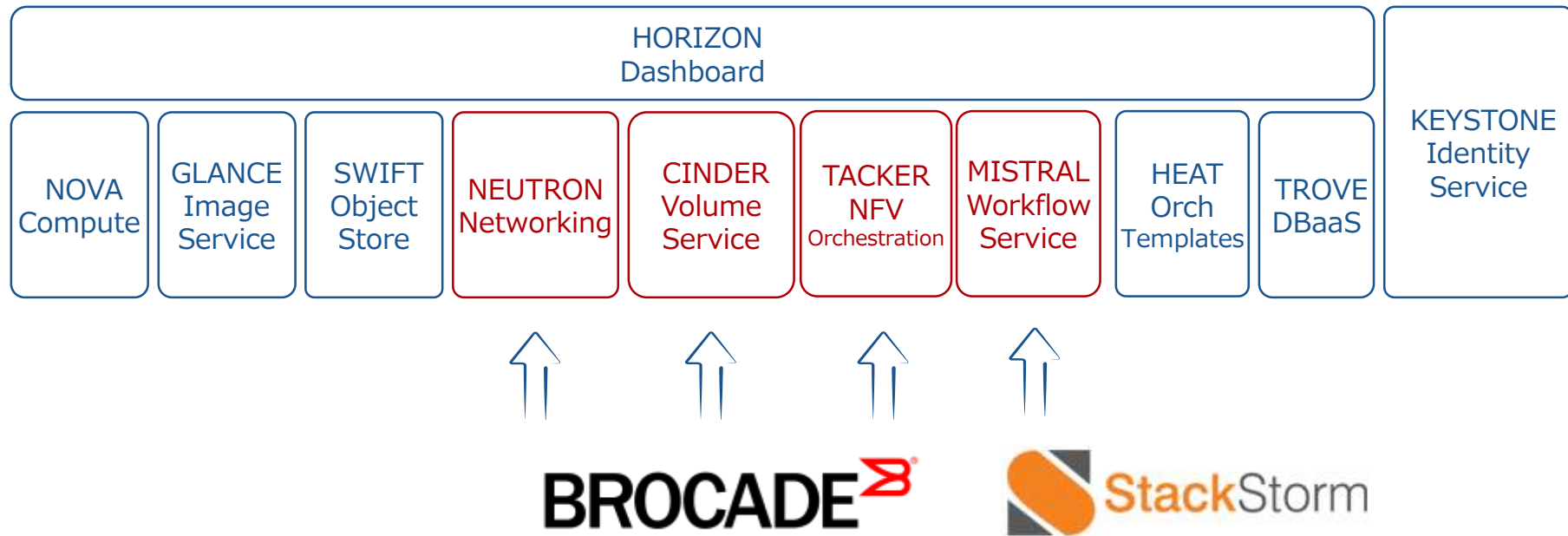
vmware



第三部 OpenStack Tacker OpenStack Mistral

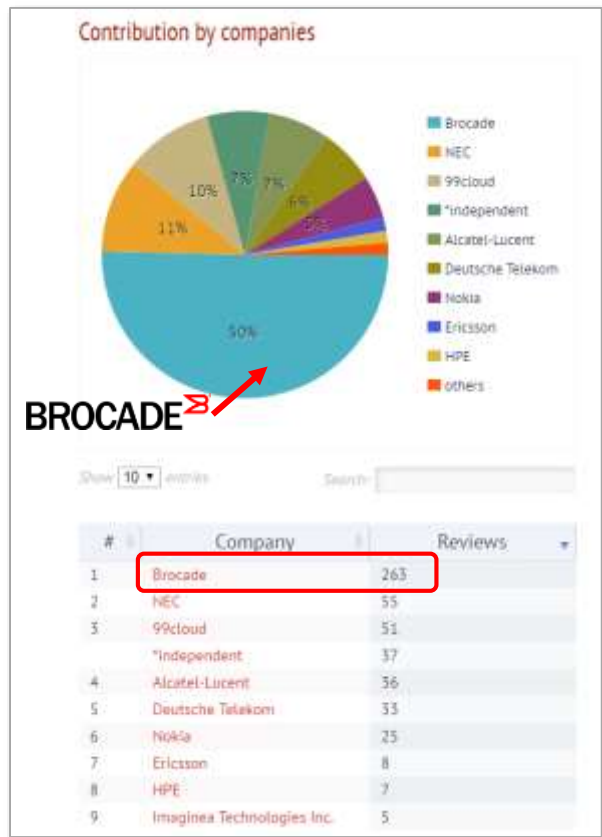
主な OpenStack プロジェクト

Neutron (ネットワーク), Cinder (ストレージ), Tacker(オーケストレーター)
& Mistral (ワークフロー)へフォーカス



オープンソースへのコミットメント

Tacker (NFV Orchestration)



Mistral (Workflow Automation)

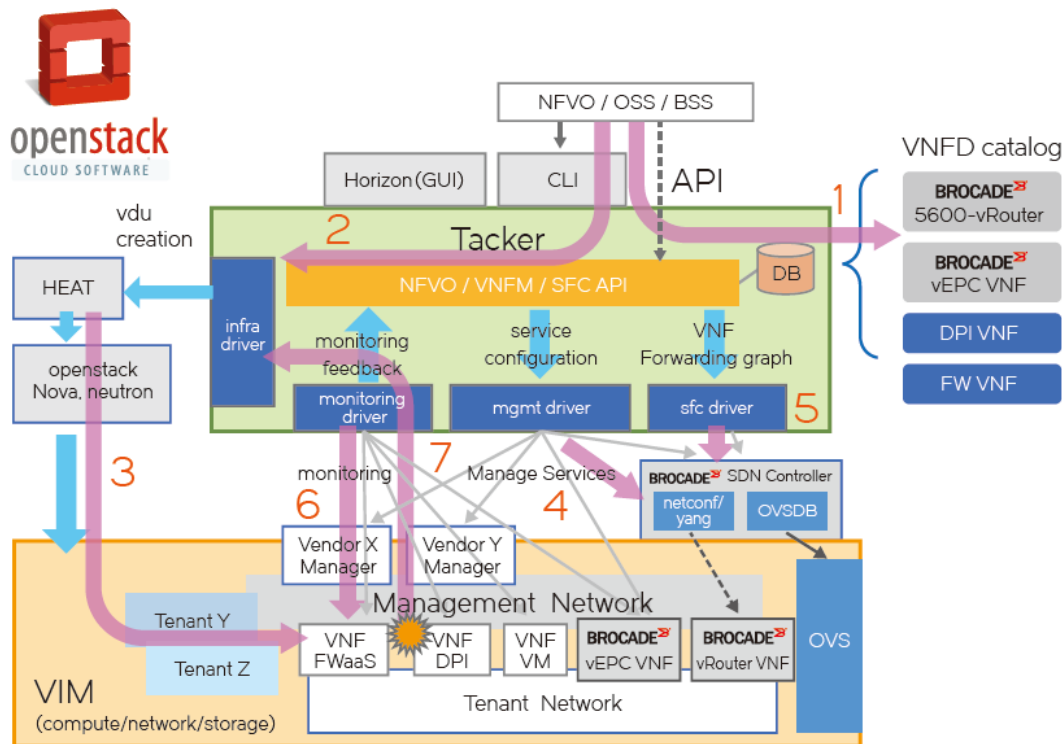


ETSI MANOベースの OpenStack VNF管理

Tackerで提供される機能

- VNFカタログ
TOSCA テンプレートを展開
- VNF ライフサイクル・マネージメント
- TOSCA VNFD テンプレートによるパラメータ指定
- VNF ユーザデータの挿入
- VNF の設定投入
インストレーションやアップデート
 - ✓ SDN Controller から NetConf/YANGを利用
 - ✓ カスタム・マネージメント・ドライバ
- ロードバブルKPIとヘルスマニターのフレームワーク

Tacker アーキテクチャとワークフロー

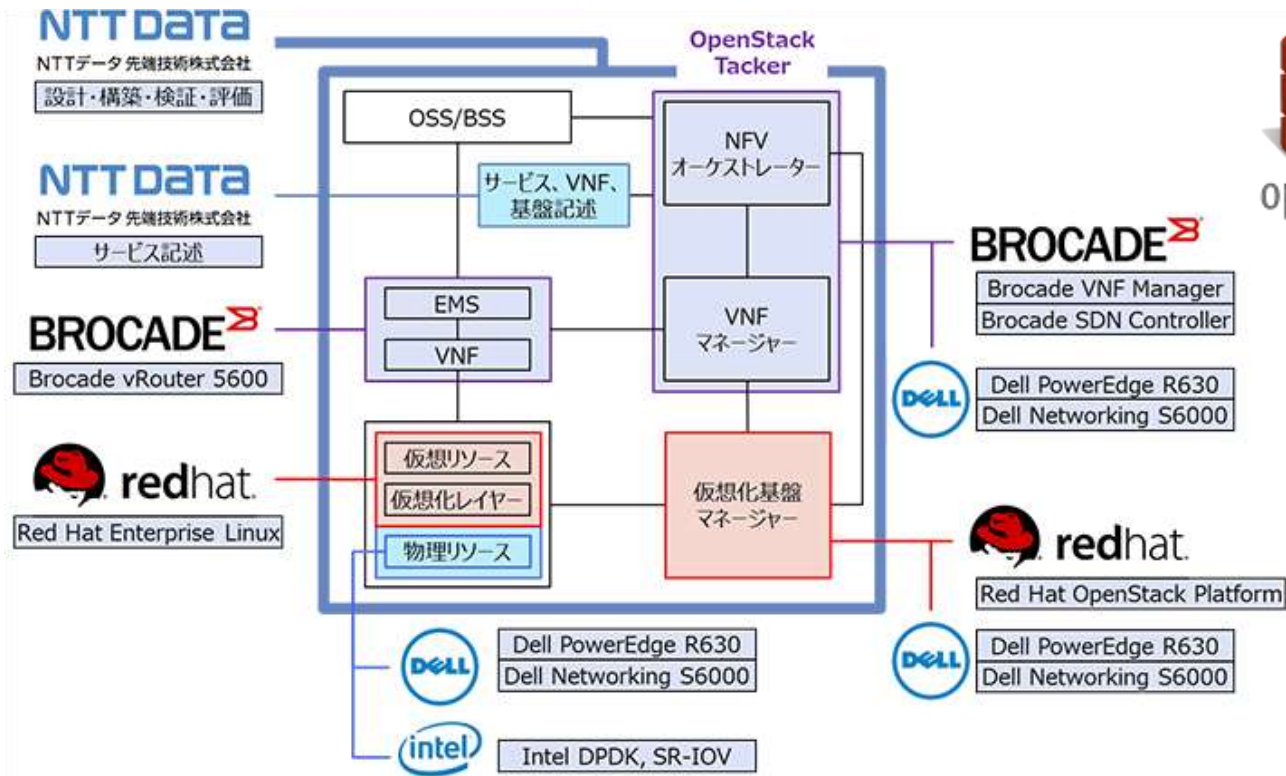


「OpenStack Tacker」による、VPNサービスの運用自動化 およびNFVの障害自動再構成による保守自動化

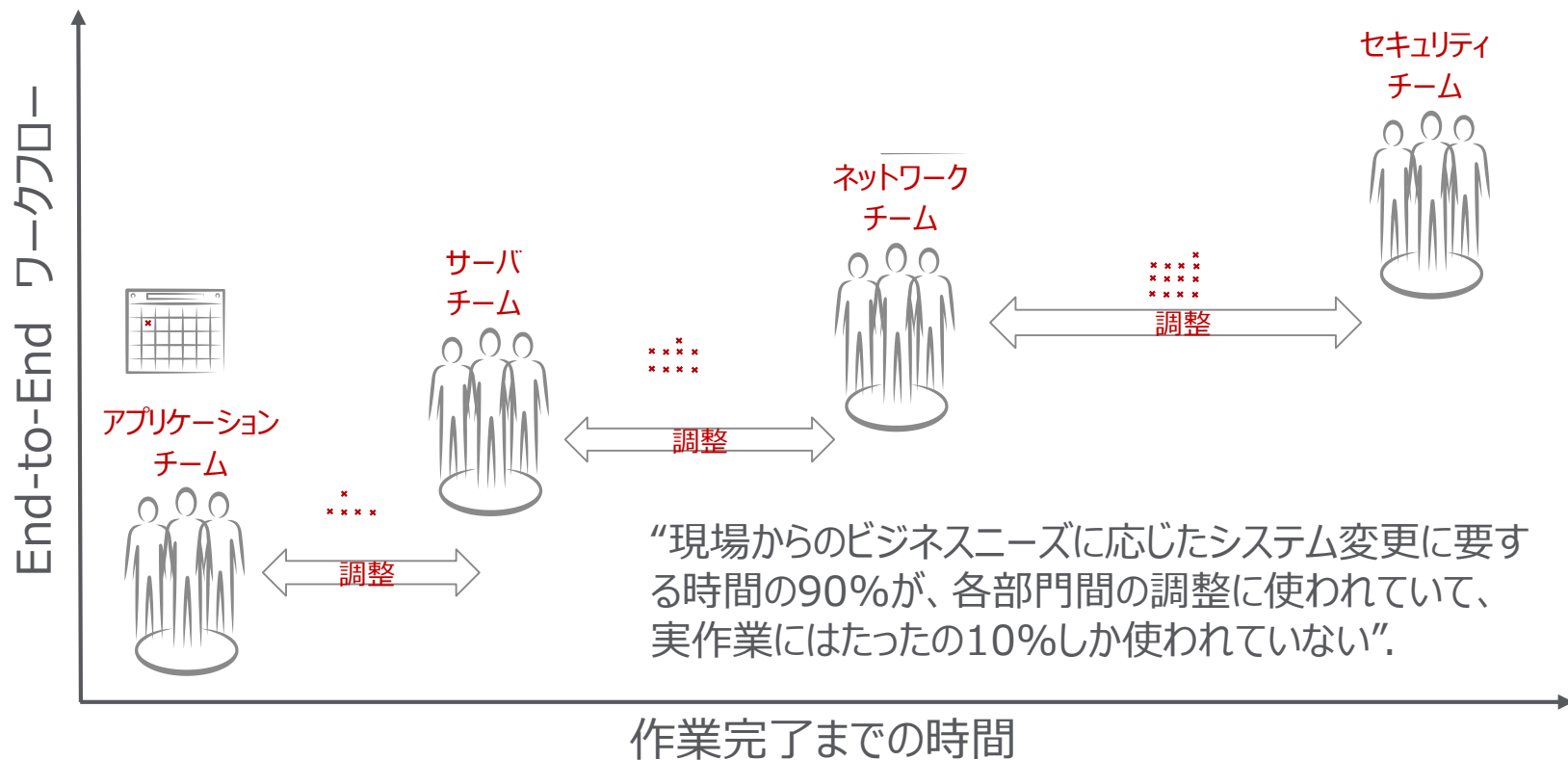
2016年6月6日：プレスリリース



SDI/NFV部門
グランプリ受賞

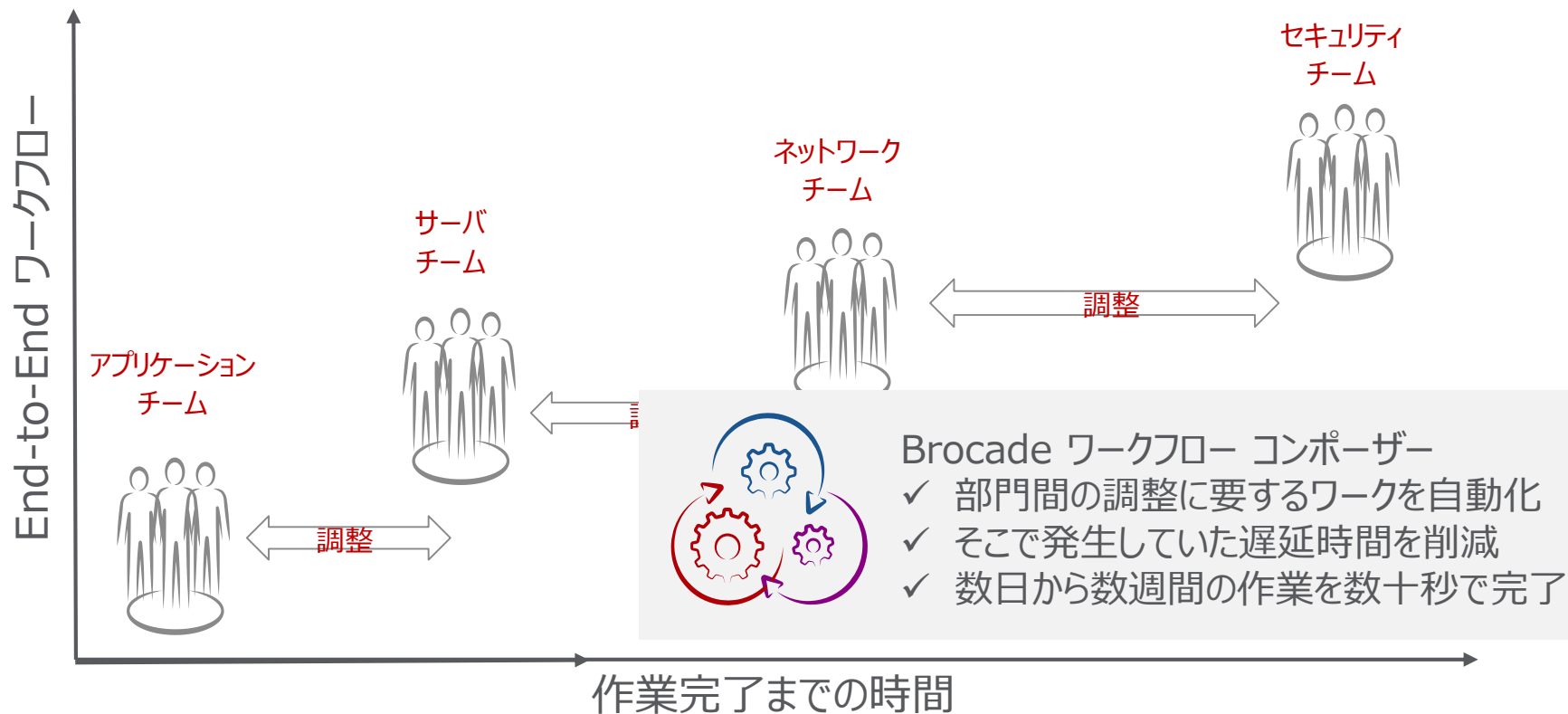


サイロ型の自動化だけでは足りない

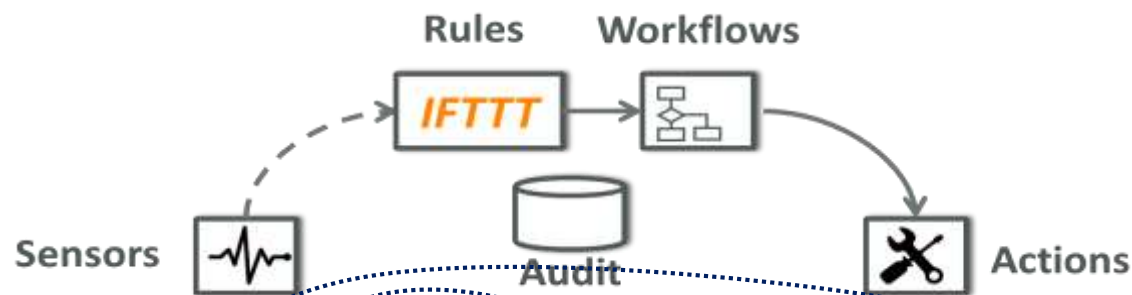


部門間をまたがった自動化が必要になる

クロスドメイン・オートメーションこそが真の自動化

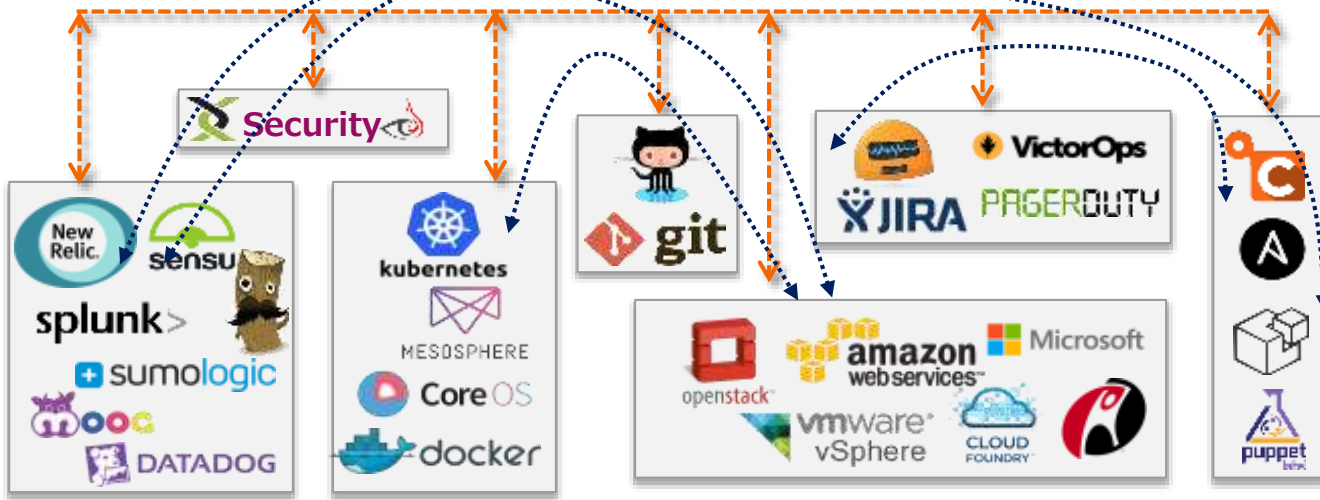


ネットワークの境界を越えて究極の自動化の世界へ



 StackStorm

OpenStack Mistral



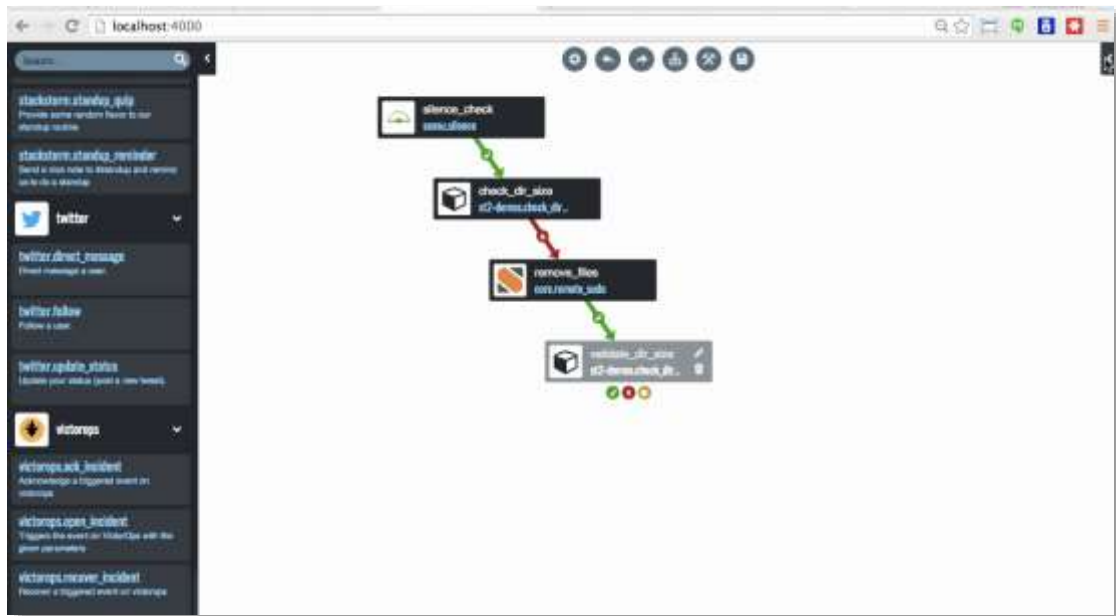
仮想障害を起こして耐障害性を日常的に検証



人工的にシステム障害を引き起こす、
逆転の発想のツール「Chaos Monkey」
をNetflixがオープンソースで公開

遅延猿、お掃除猿、あばれゴリラ。。

Amazon EC2のインスタンスをランダムに落とし、サービスに対して仮想的な障害を引き起こす。
これにより、サービスがきちんと冗長化され、耐障害性を持つように作られているかを日常的に検証。



NEW WORLD. NEW NETWORK.

Brocade: Your Partner for the New IP

ありがとうございました

