

NS308

【株式会社コミュニティネットワークセンター様】
vSANとNSXをフル活用で実現した、
次世代マルチサイトデータセンターの
中身とは？

株式会社コミュニティネットワークセンター

取締役 技術本部長

石倉 雅巳 様

技術本部 サーバグループ リーダー

ニコライ ボヤジエフ 様

#vforumjp



POSSIBLE
BEGINS
WITH YOU

vSANとNSXをフル活用で実現した、 次世代マルチサイトデータセンターの中身とは？

株式会社コミュニティネットワークセンター

石倉 雅巳
ニコライ ボヤジエフ

2018年11月

Part I.

会社紹介

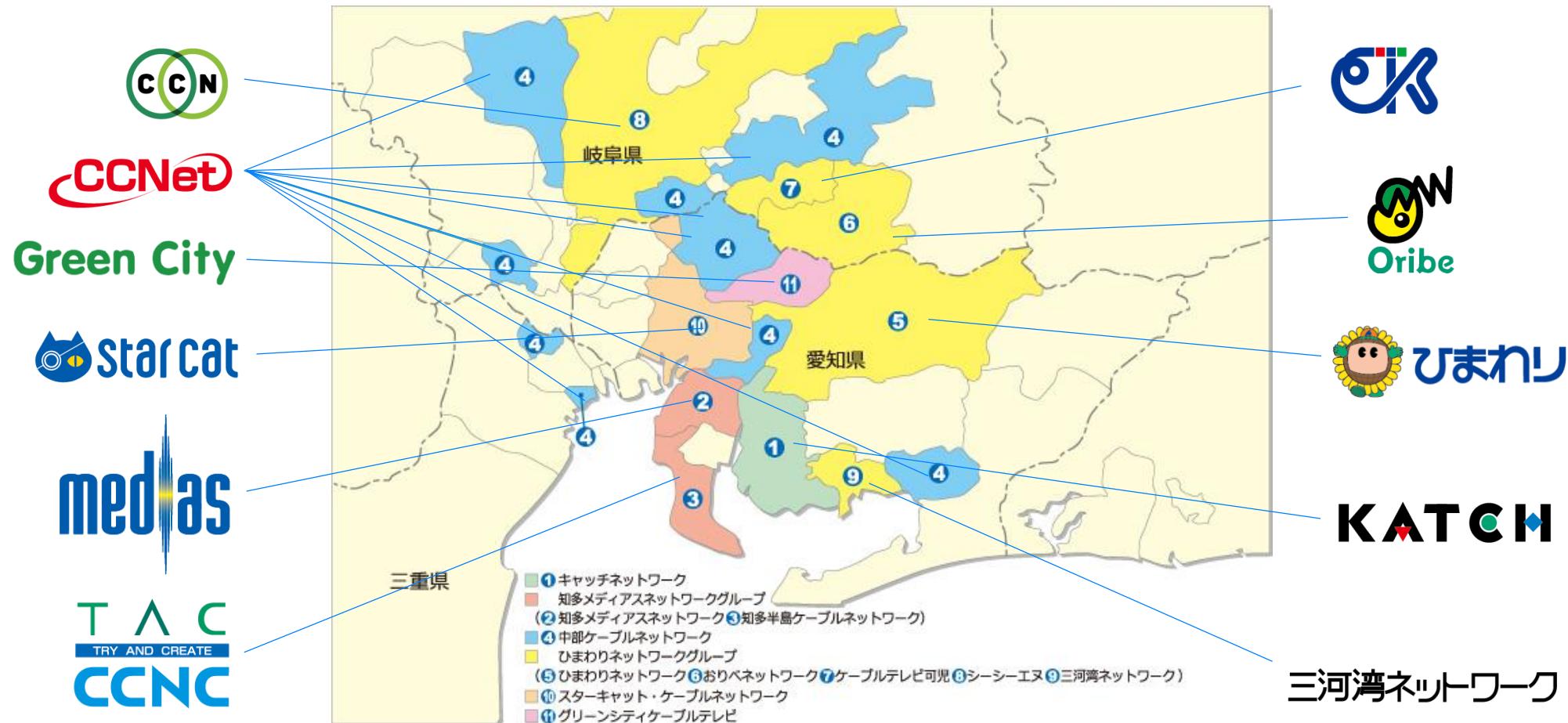
CNCi 紹介

- 愛知/岐阜/三重で活動するケーブルテレビMSO
MSO : Multiple Systems Operator
- グループ ケーブルテレビ局11局



CNCIグループ サービスエリア

CNCI

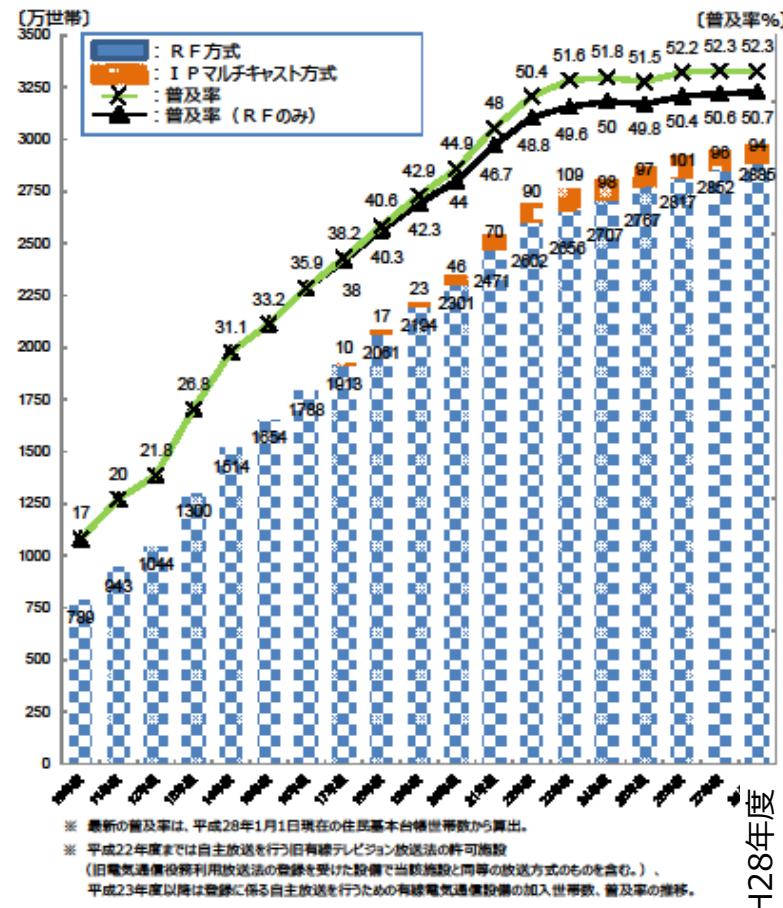


約150万世帯にテレビ || ネット || 電話サービスをご提供中

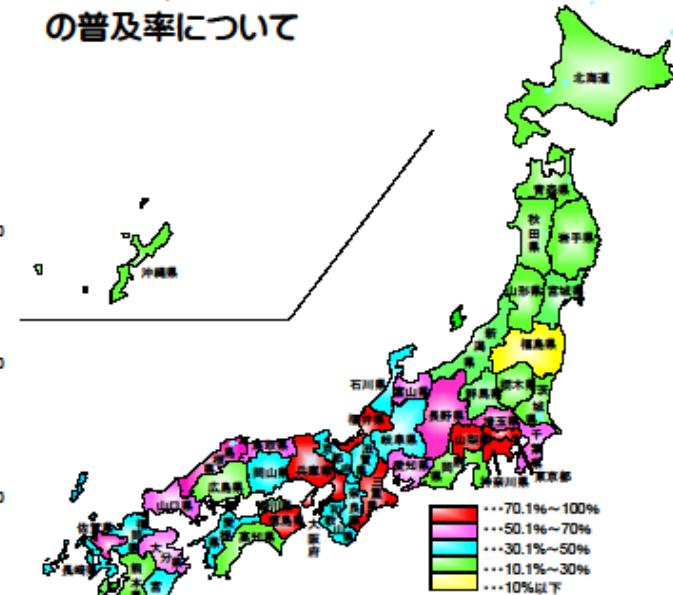
ケーブルテレビの普及率

2. ケーブルテレビの加入世帯数・普及率の推移

□ ケーブルテレビ加入世帯数は年々増加し、平成29(2017)年3月末には2,980万世帯、普及率は52.3%に達している。



3. 都道府県におけるケーブルテレビ(自主放送あり)12 の普及率について



岐阜県	36.7%
静岡県	27.2%
愛知県	54.8%
三重県	75.0%

(東海地区48.2%)

都道府県	普及率	都道府県	普及率	都道府県	普及率	都道府県	普及率
北海道	25.6%	埼玉県	57.5%	岐阜県	36.7%	鳥取県	63.5%
青森県	17.6%	千葉県	55.7%	静岡県	27.2%	島根県	55.1%
岩手県	18.9%	東京都	81.7%	愛知県	54.8%	岡山県	34.1%
宮城県	29.2%	神奈川県	71.7%	三重県	75.0%	広島県	28.8%
秋田県	16.5%	新潟県	22.5%	滋賀県	37.5%	山口県	61.1%
山形県	16.6%	福島県	65.9%	京都府	45.0%	徳島県	89.8%
福島県	3.9%	石川県	43.8%	大阪府	87.4%	香川県	27.8%
茨城県	21.9%	福井県	74.3%	兵庫県	71.6%	愛媛県	37.0%
栃木県	23.0%	山梨県	82.2%	奈良県	46.9%	高知県	24.6%
群馬県	13.9%	長野県	51.0%	和歌山县	37.5%	福岡県	47.3%
						全国	
						52.3%	

● 放送

- 地上波、BS再放送
- 専門チャンネル（STB）
 - 『楽しさ、ヨリドリ、ミドリ。』 (<https://tanoshisa-yoridorimidori.cnci.jp/>)
 - ケーブル 4K (<https://www.cable4k.jp/>)
- コミュニティチャンネル



● 通信

- インターネット接続サービス
 - ~2GbpsのFTTHによるブロードバンド接続

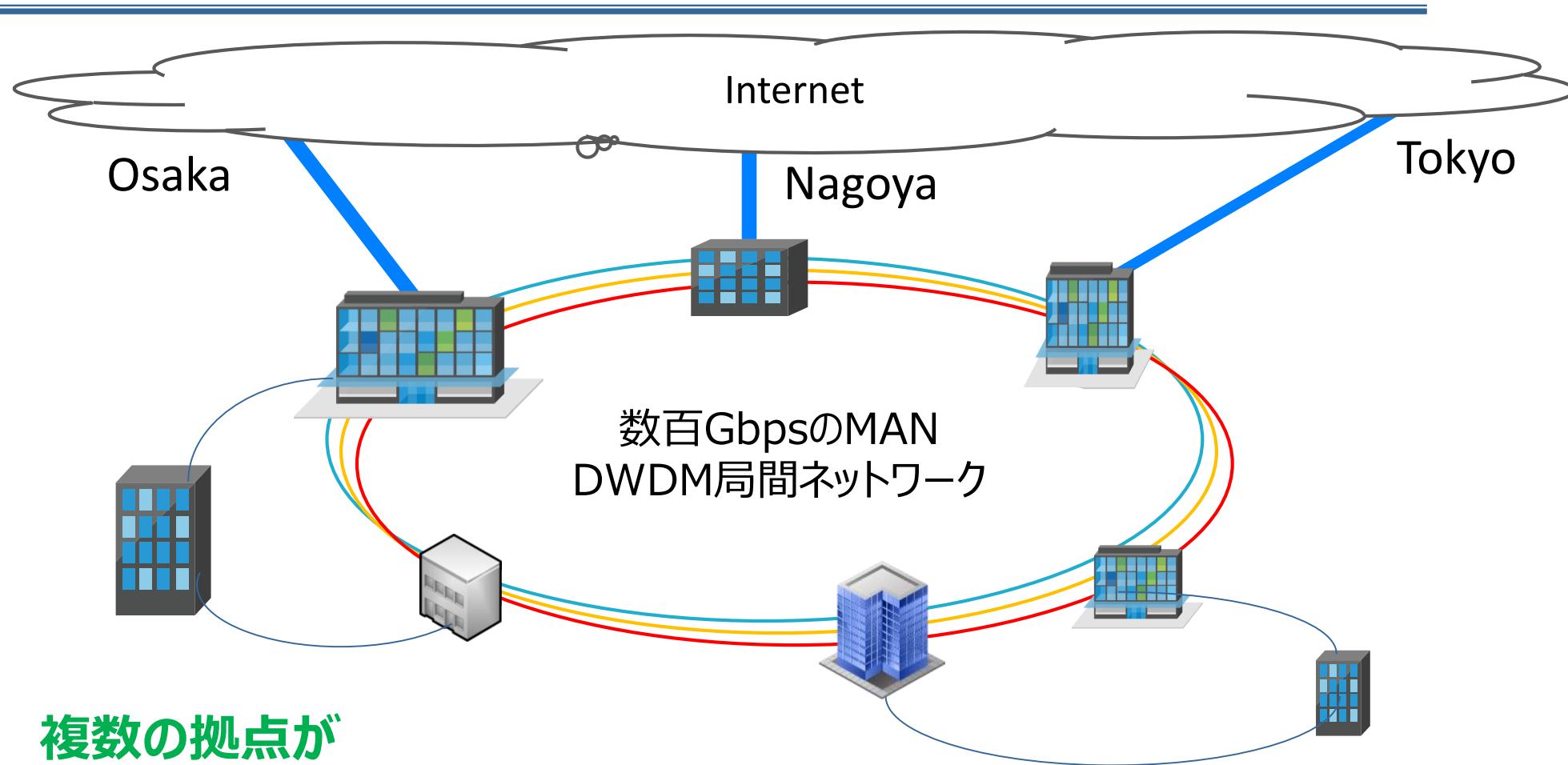
- 固定電話

● その他

- ケーブルスマホ、地域BWA、公衆WiFi、VoDサービス等々

所謂 ISP インターネット接続サービス

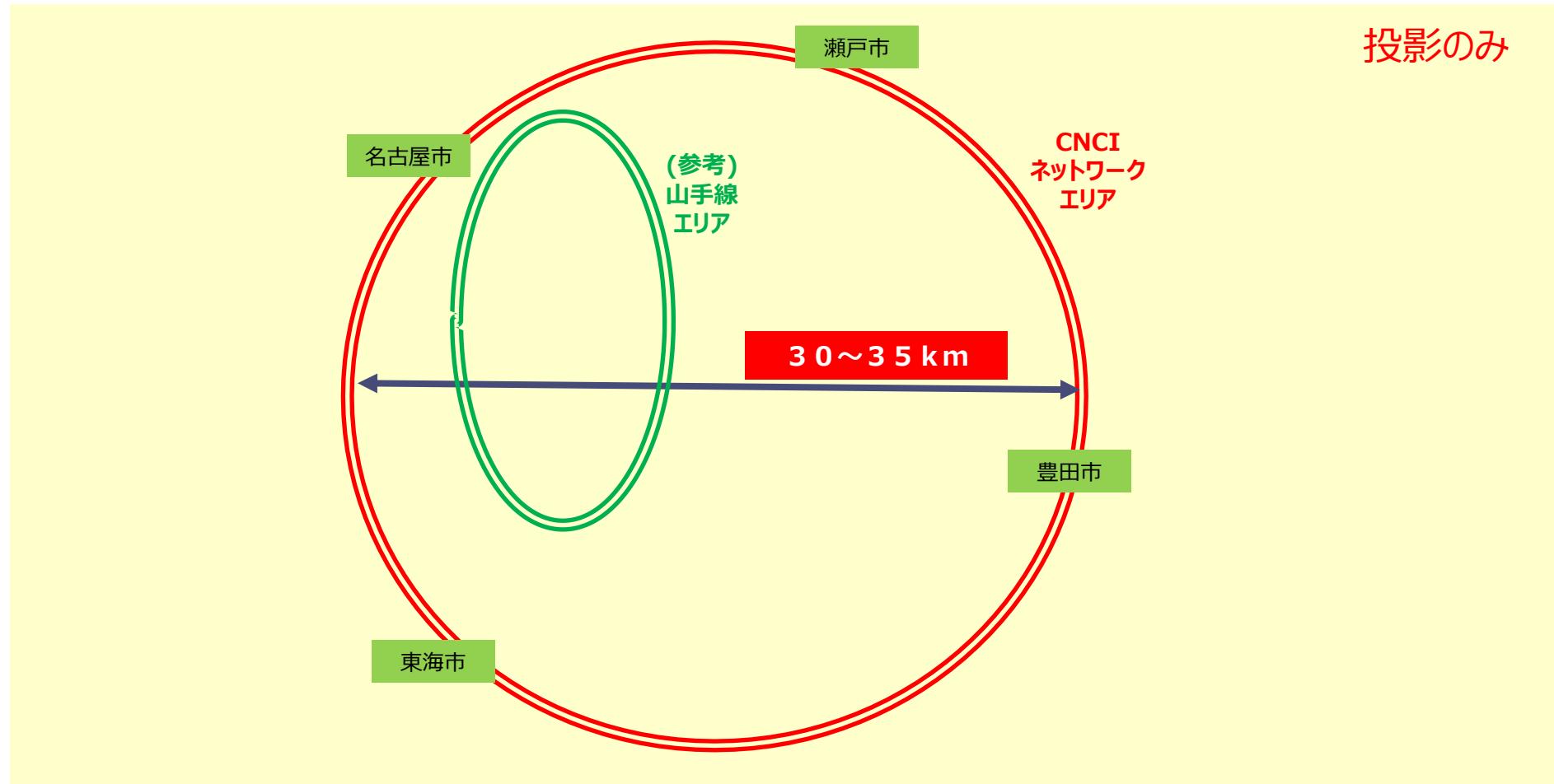
- お客様数は約50万
- ISPとしてメールサーバをはじめとして
多数のサーバー群を運用 (**200台以上**)
- 少数の技術担当で
 - 開発も、構築も、運用も
 - 多種多様なシステムを運営
 - さらにインフラ設備としてサーバ、ストレージ、
アクセス以外のネットワーク (L3/FW/LB)も
すべて守備範囲



複数の拠点が

- ・直径30~35km程度のエリアに分散
- ・超高速ネットワークでつながった
マルチデータセンター環境

参考



Part II.

SDDCの中身

自己紹介

名前：ニコライ ボヤジエフ

出身：ブルガリア 《България》

所属：株式会社コミュニティネットワークセンター
技術本部 サーバグループ

担当：メール、Web、DNS、DB、仮想化基盤…
各種サーバシステムの運用と構築（10年以上）

在日ブルガリア人：

<300人

(投影のみ)

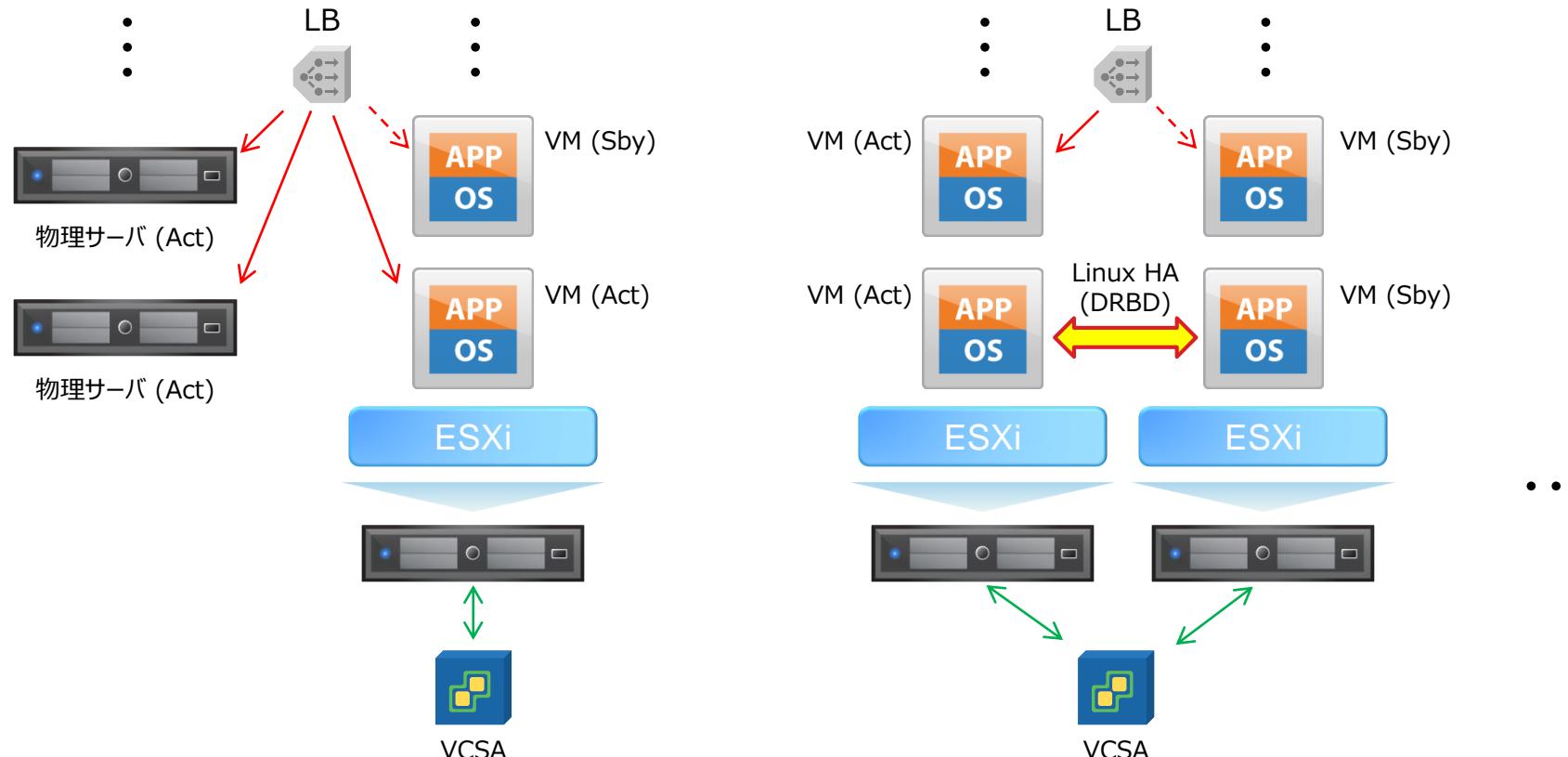
背景

- **一般ユーザ向けサービス**を提供するインフラ (ISP) ※約50万ユーザ
- 「**重大事故**」※ 起こしてはならない
- 障害時、自分たちで対応してあるため、**深いシステム理解度**が必要
- サーバ、ネットワーク、ストレージを含む**幅広い守備範囲**
- システムの**構築も運用も**同じチームでやっている

※3万ユーザに影響があった且つ2時間以上の障害は「重大事故」（総務省報告義務）

今までやってきたVMware製品の使い方

- vMotionなし、HAなし、DRSなし（vSphere Essentials）



- シンプル、直営構築、安価、安定している GOOD
- 管理性よくない、ホストメンテ超面倒くさい、DR考慮されてない BAD

今までやってきたVMware製品の使い方

- お疲れ様、ESXi !



esxi-41.v.cncl.ad.jp

Model: HP ProLiant DL380 G7

Processor Type: Intel(R) Xeon(R) CPU X5670 @ 2.93GHz

Logical Processors: 24

NICs: 16

Virtual Machines: 12

State: Connected

Uptime: 2805 days



7年半無停止！！



出典：ナリタイ < naritai.jp >

```
~ # vmware -v  
VMware ESXi 4.1.0 build-348481
```

※2018年11月15日(木)に消灯式を予定しております。

メールシステムを完全仮想化でリニュアルした

- 2017年3月～ (vSphere Enterprise Plus + vROps, EMC VNX All-Flash + RecoverPoint)



- 非常に安定稼働 GOOD
- DRはかなり複雑（カスタム）、ベンダ依存、お高い BAD

▼DRスクリプト

名前 ▲
サービスチェック.cmd
サイト切り替え.cmd

▼DR訓練スクリプト

名前 ▲
サービスチェック訓練.cmd
サイト切り替え訓練開始.cmd
サイト切り替え訓練終了.cmd

仮想化基盤を作ることになった（2017年度）

- メール以外の物理機器（約40台）の保守切れ時期が近い
- 他部署やグループ会社もシステム更改等でサーバリソースがほしい
- 経営層のクラウドビジョン
- 仮想化基盤を作ることになった



さあ、どう作ろう？

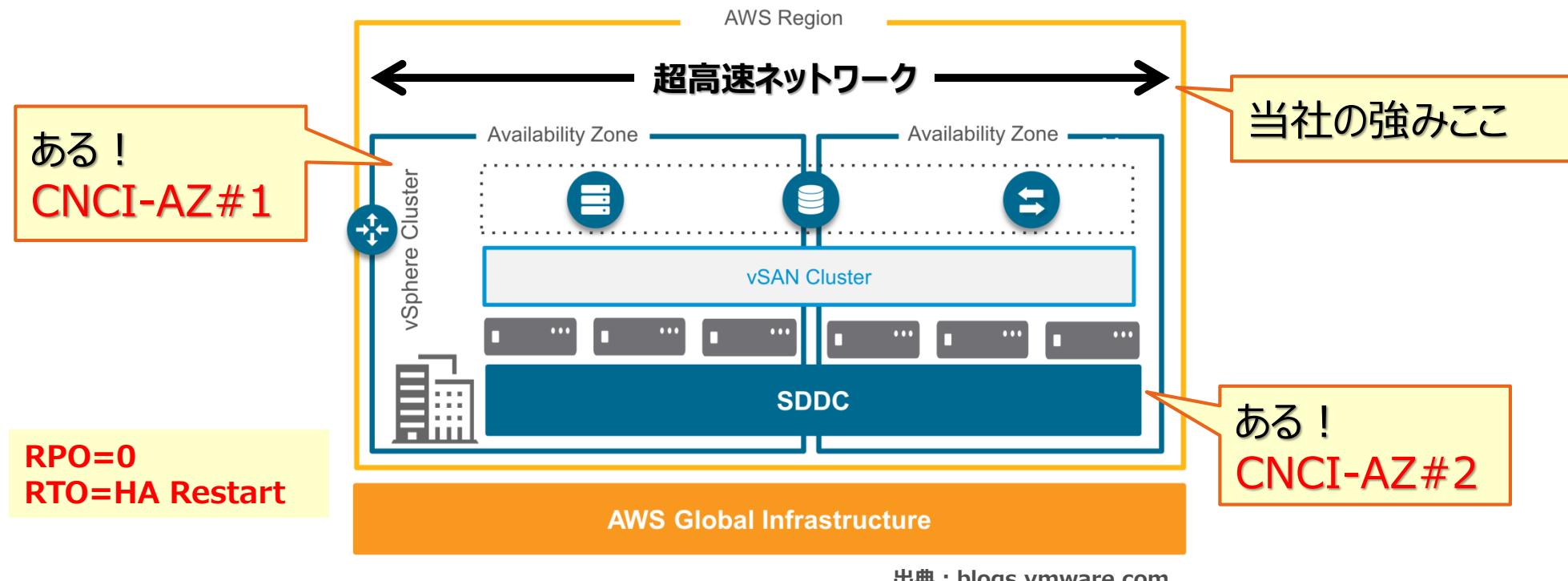


出典：ナリタイ < naritai.jp >

自社の強みを生かした「とんがった構成」に挑戦

- ある憧れから始まった…

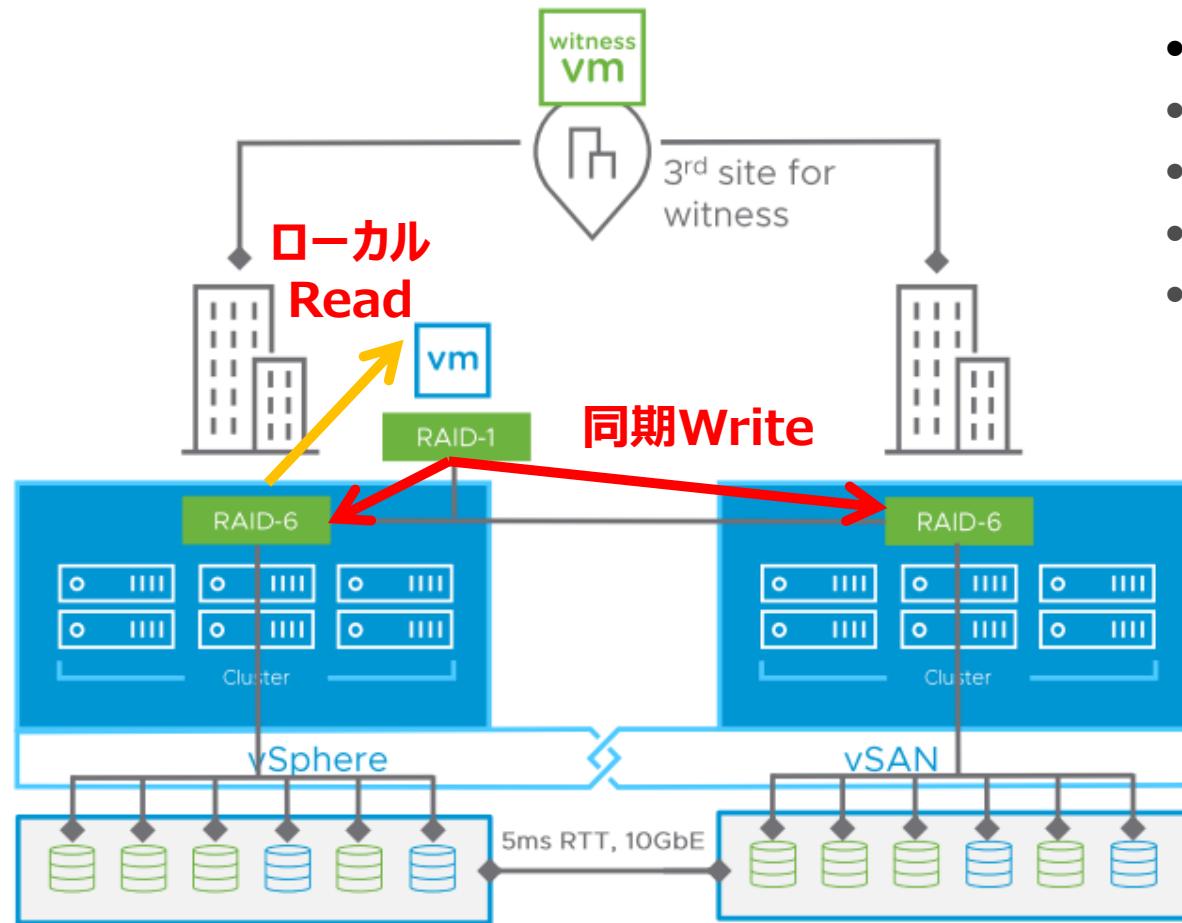
VMware Cloud on AWS Multi-AZ Availability (※当時プレビュー)



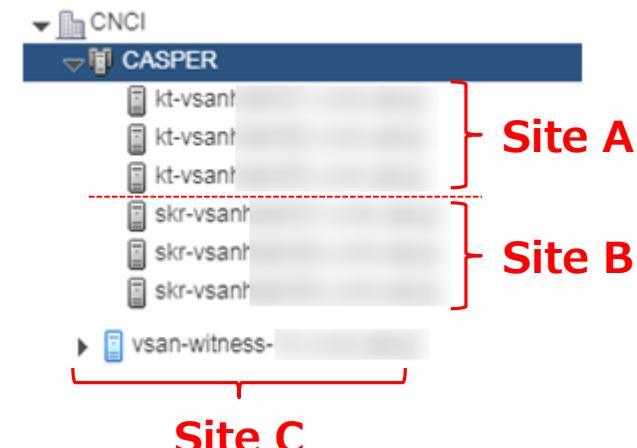
出典 : blogs.vmware.com

- Act-Act (**同期Write**) だから、DRオーケストレーションいらない
- インフラレベル**でAZ間の可用性を実現 (アプリそのまま)
- DCメンテ**や**DC引っ越し**が可能になる (!)

vSAN Stretched Cluster って何？



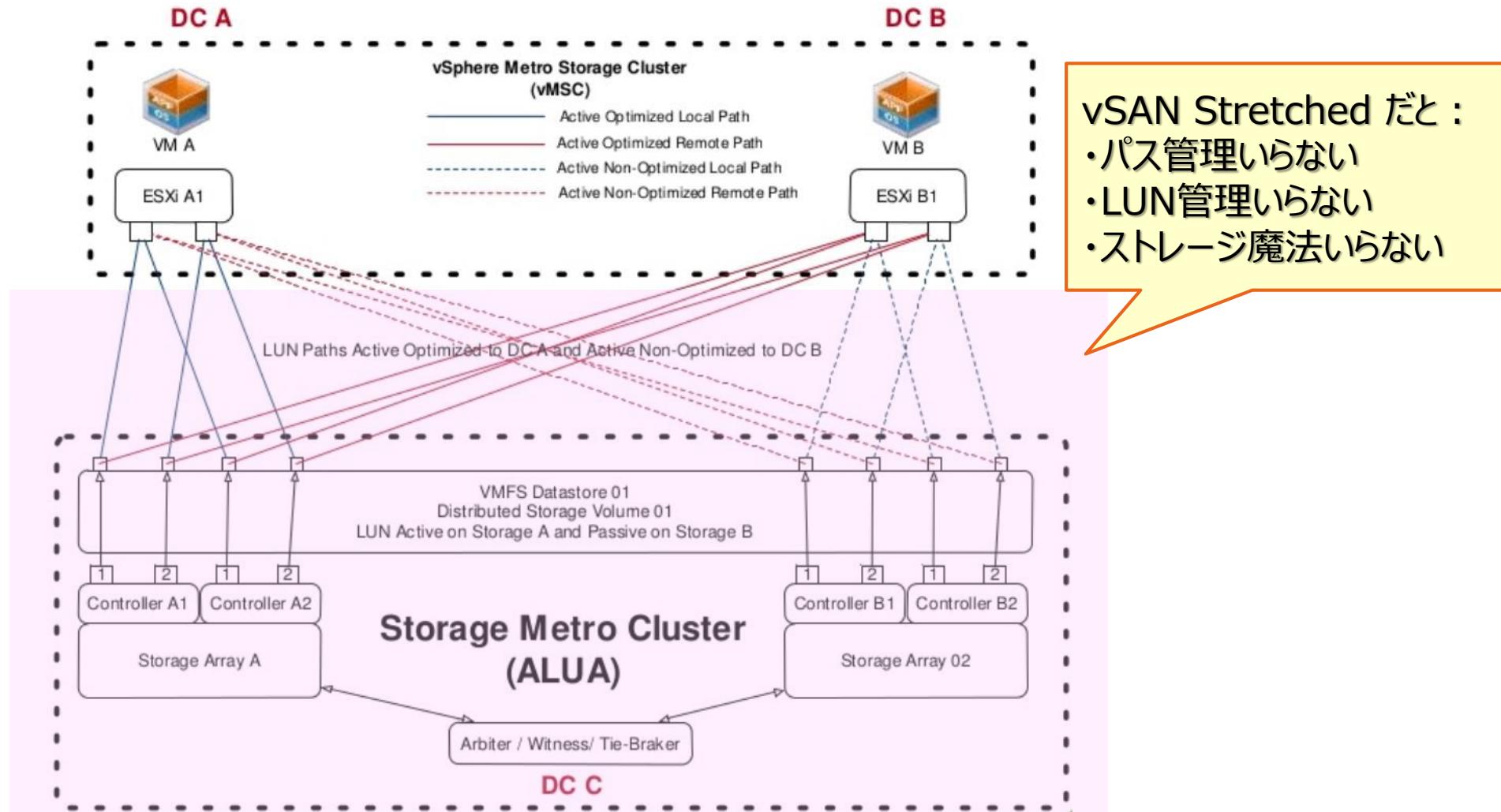
- 3拠点
- **10GbE <5ms RTT 必須**
- Witness は、Nested ESXi のVM
- 単一クラスタ (vMotion/HA/DRS)
- サイト間データ冗長 (PFTT)
- サイト内データ冗長 (SFTT) ※vSAN 6.6～



出典 : storagehub.vmware.com

vSAN Stretched Cluster って何？

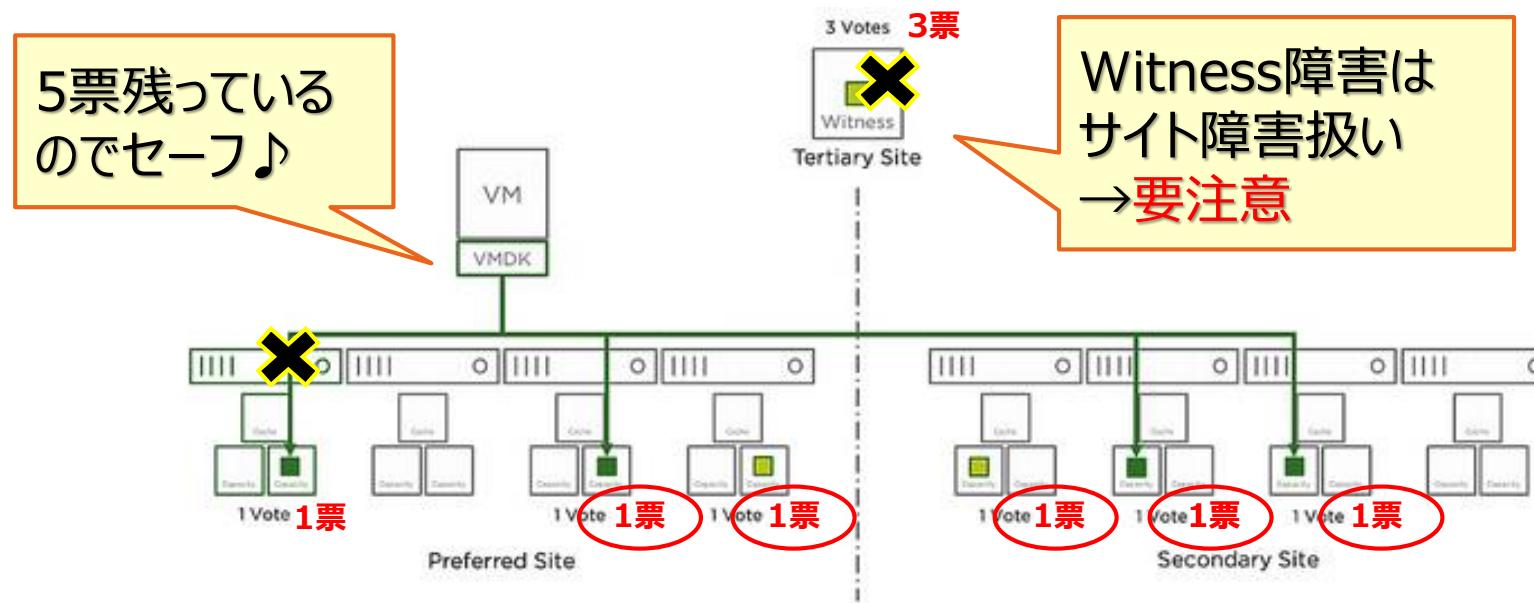
- Active-Active vMSC Cluster との違い



vSAN Stretched Cluster って何？

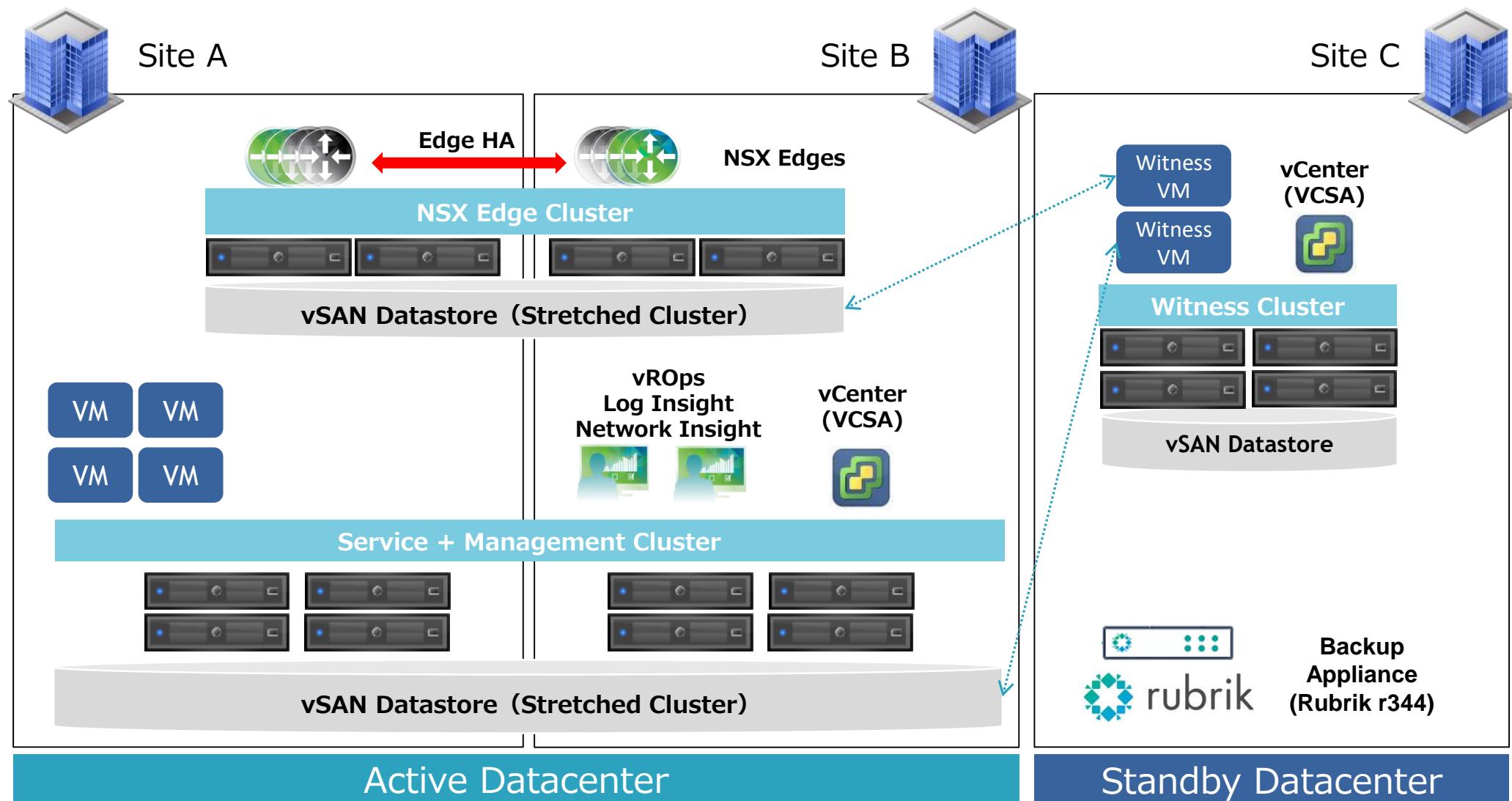
- 定足数(票50%)の仕組みでsplit-brain防止
- 1サイトと1ホストの同時障害まで許容** (PFTT=1/SFTT=1)

例：PFTT=1、SFTT=1(Raid1)、WitnessとSiteAホストの同時障害の場合



- 正常時、 $(1+1+1)+(1+1+1)+3=9$ 票がある
- 定足数のルールとして、総票数(9)の半分(4.5)以下になると、データアクセスできなくなる (split-brain防止)
- この場合、**最低5票**が残ってる必要がある

CNCi の vSAN Stretched Cluster 構成



CNCI の vSAN Stretched Cluster 構成

- ソフトウェア構成

- **vCloud Foundation**

- ✓ vSphere Enterprise Plus (**v6.5U2**) + vSAN Enterprise
 - ✓ NSX Enterprise (**v6.4.x**) (※現在、NSX Datacenter Enterprise Plus)
 - ✓ SDDC Manager (※現在、未使用)

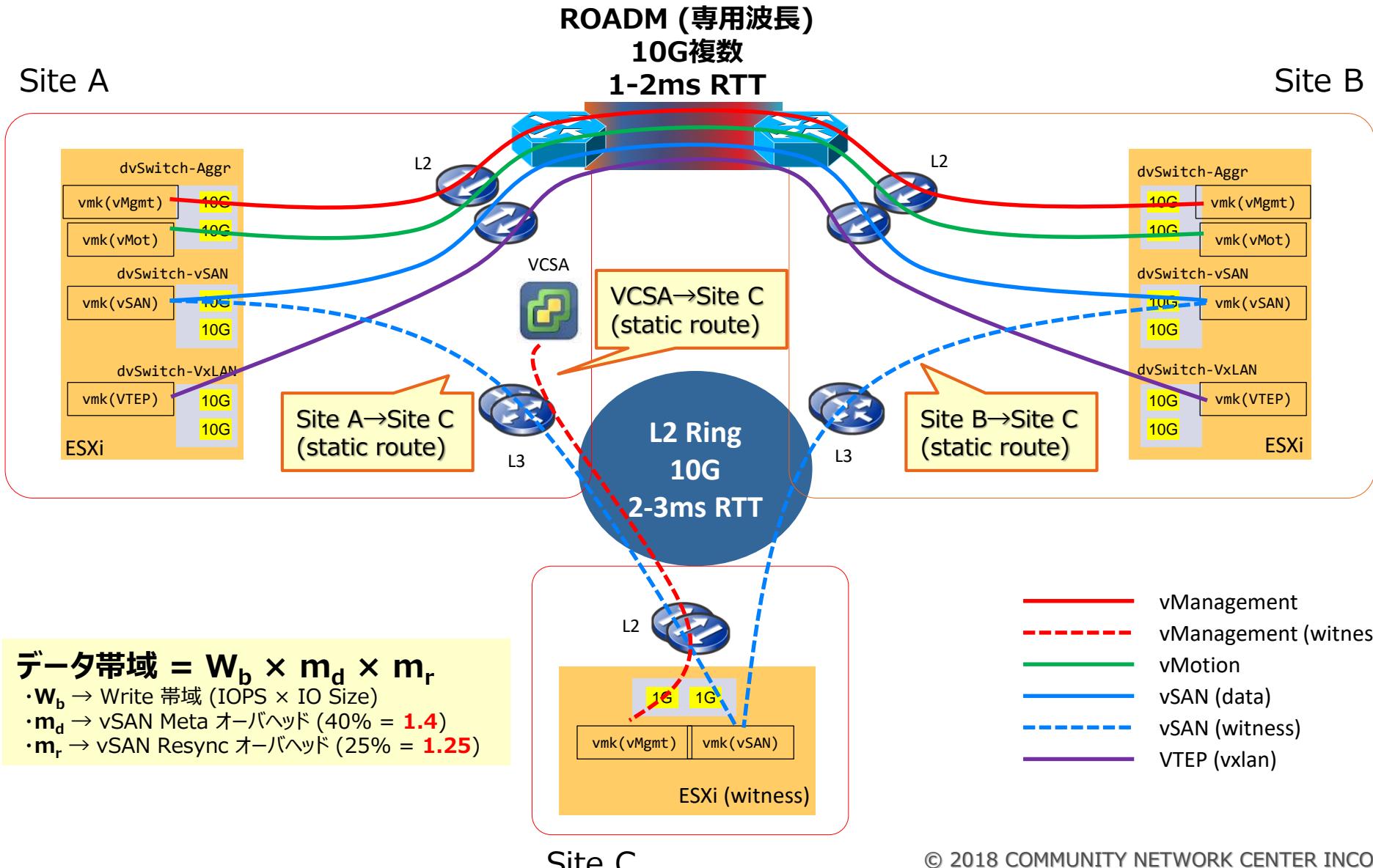
- **vRealize Network Insight** (※現在、NSX Datacenter Enterprise Plus)

- **vRealize Suite Standard**

- ✓ vRealize Operations Advanced
 - ✓ vRealize Log Insight
 - ✓ vRealize Business Cloud Standard (※現在、未使用)

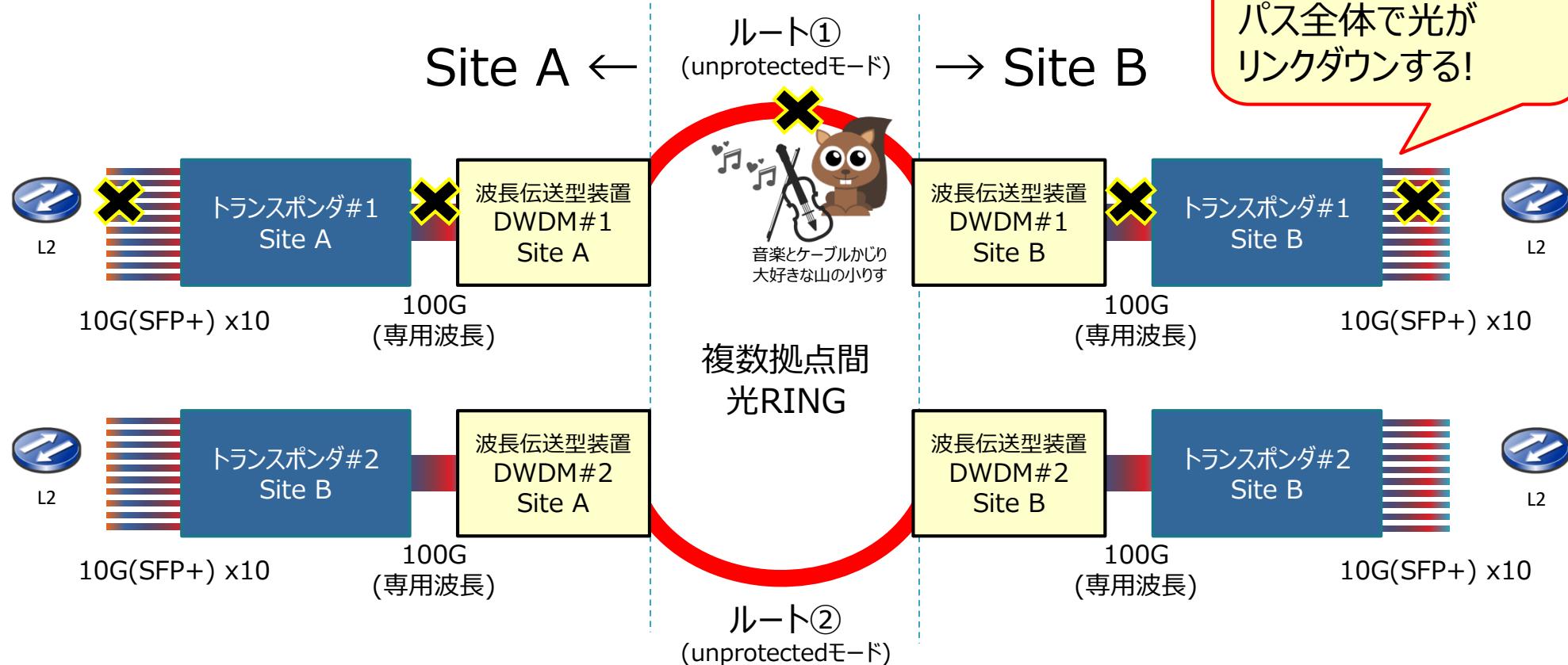
- **vCenter Standard**

CNCI の vSAN Stretched Cluster 構成



vSAN Stretched Cluster ネタ① : ROADM

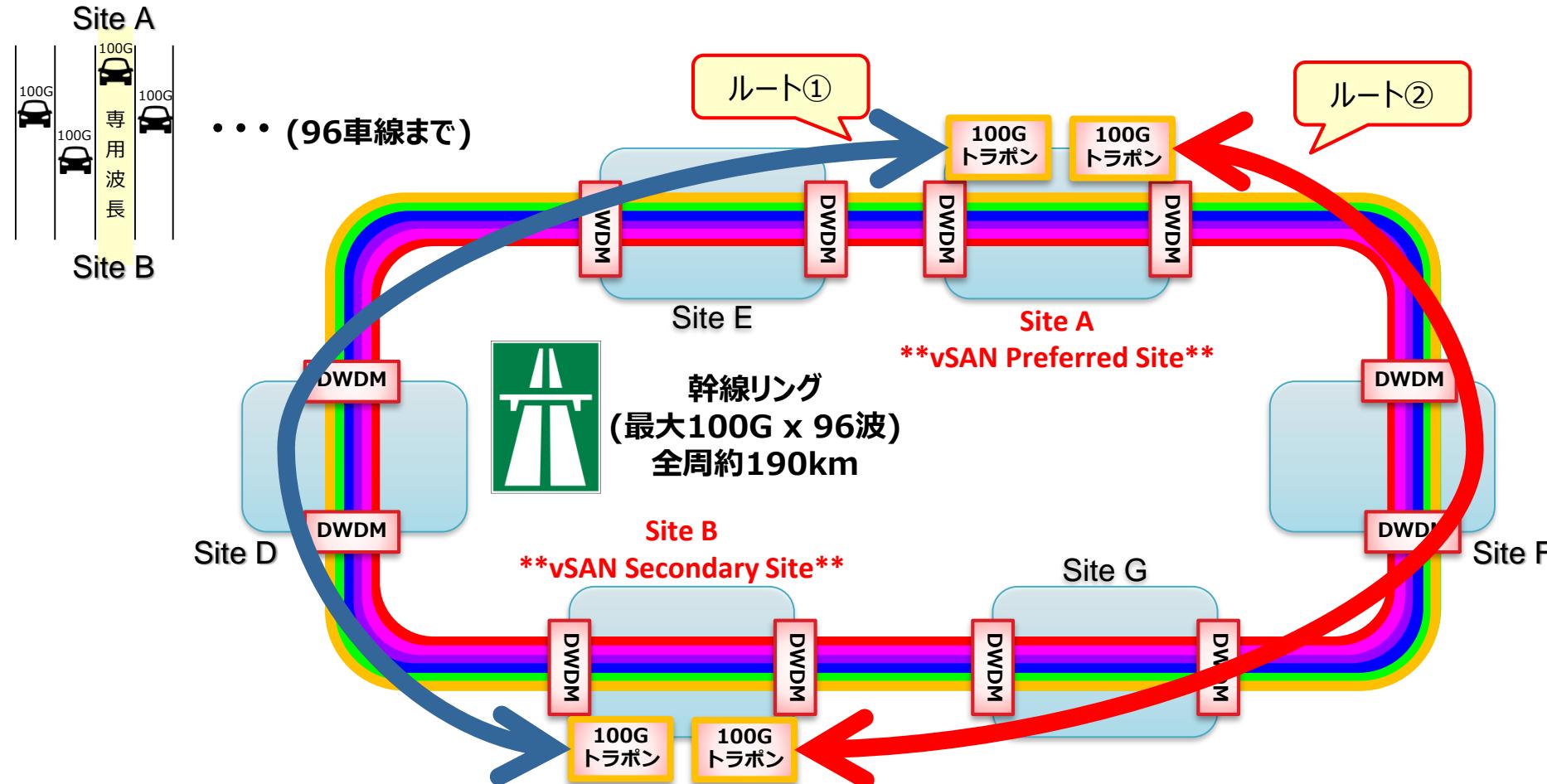
- 拠点間はROADM(専用波長)を使っている



- 冗長 L1 Trunk なのでやり放題 (MTU、マルチキャスト、L2トポロジ等)

vSAN Stretched Cluster ネタ① : ROADM

- 100G専用波長は、「光Autobahn」のたった1車線



vSAN Stretched Cluster ネタ②：サーバ

- サーバ構成

ESXi ホスト (サイトA・B各4台/計8台)



Hewlett Packard
Enterprise



HPE DL380 Gen10 / 1台あたりのスペック

(投影のみ)

Point: 1xCPU

※NUMA配慮不要
※ライセンスコスト抑制

Point: SSD Capacity は SATA

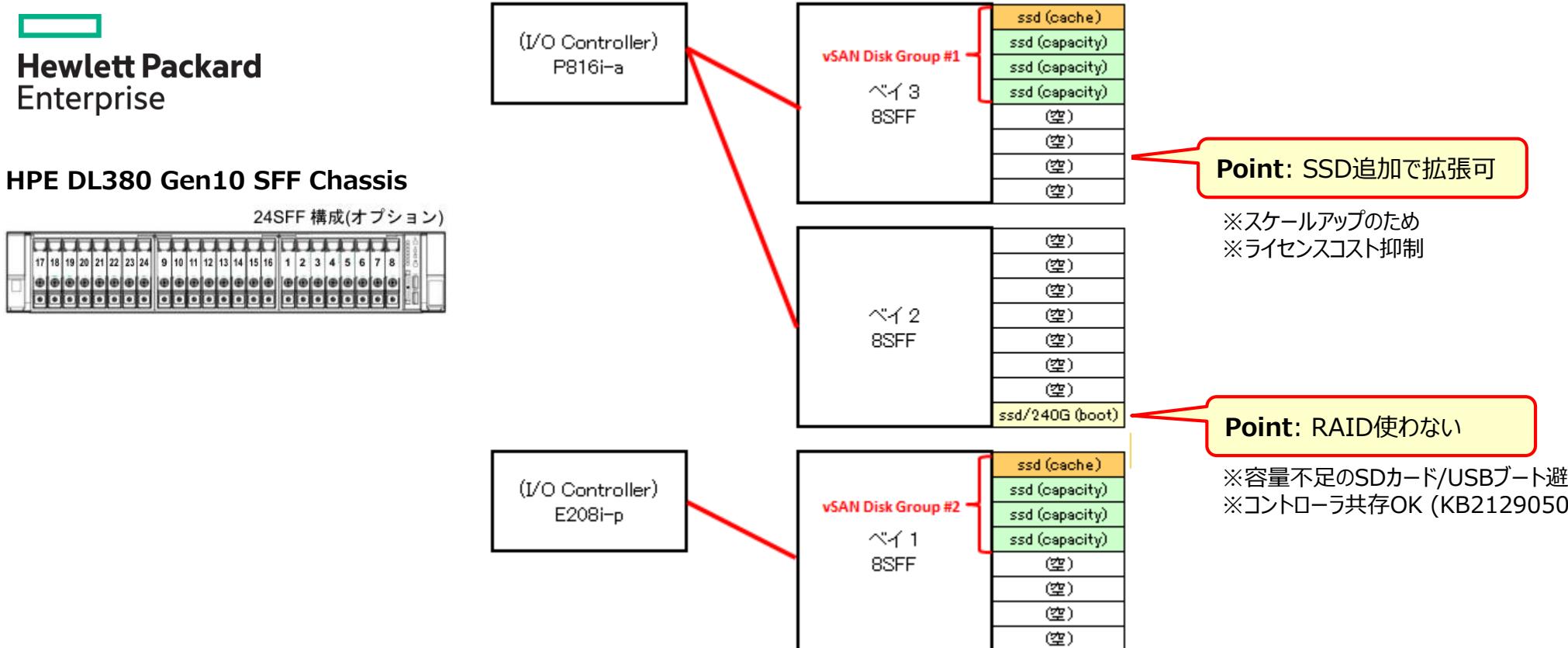
※ハードウェアコスト抑制

Point: 10G Base-T

※汎用L2SW使うため

vSAN Stretched Cluster ネタ②：サーバ

- Disk Group 構成



▼参考

Best practices when using vSAN and non-vSAN disks with the same storage controller (2129050)

<https://kb.vmware.com/kb/2129050>

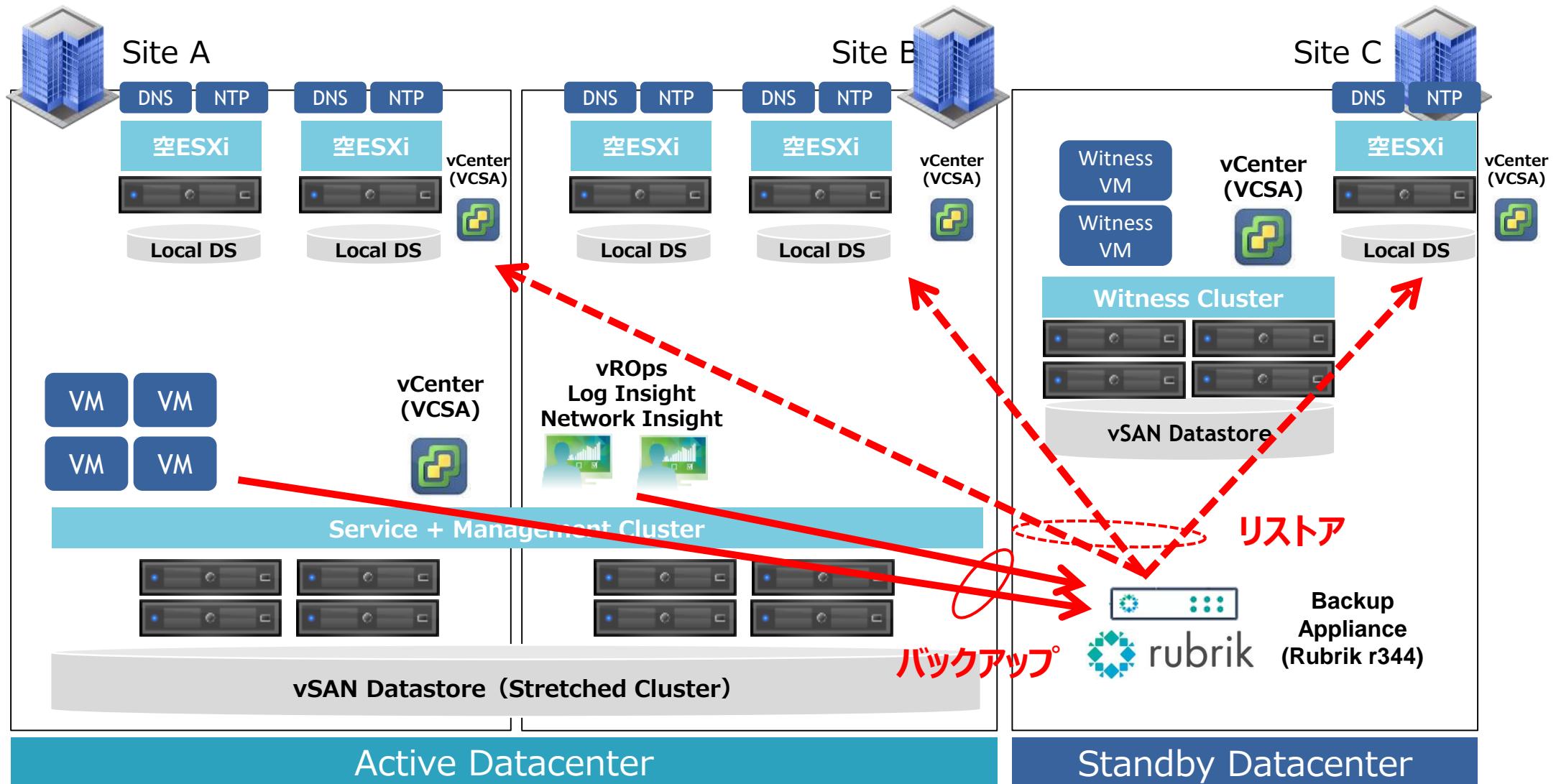
vSAN Stretched Cluster ネタ②：サーバ

- SATA SSD で大丈夫？？



(投影のみ)

vSAN Stretched Cluster ネタ③：バックアップ[¶]



vSAN Stretched Cluster ネタ④：ワークロード

- ・ どのようなワークロード？

(投影のみ)

- VDIなし
- ストリーミングなし
- **大容量ファイルサーバなし** (HPE Nimble + Zertoで個別構築)

【理由】

- ・コスト (¥/GB) (15TB以上)
- ・vSANネイティブバックアップまだない



nimble
storage

Zerto

vSANってどう？

- Stretched Cluster構成が簡単、ちゃんと動く
- HCIだから良い
 - スモールスタート可、簡単拡張、**ストレージ管理不要**
- 純粹ソフトだから良い
 - 専用ハード不要、**汎用ハード**で構成できる（★コスト削減）
- VMwareだから良い
 - サードベンダーなし、**サポート一本化**
 - vSphere慣れてるので、**新しい製品覚えなくていい**
- 高パフォーマンス
 - (投影のみ)
 - Resyncトラフィックの自動流量制限機能あり（手動流量制限も可）
- 可視化できる
 - ヘルスチェックがいい、vCenterでパフォーマンス詳細情報確認できる、**ブラックボックスではない**
 - 詳細情報全てAPI経由で取得可能
- シンプルでありながら賢い
 - ディスクの**障害予兆検知**できる！（次項参照～）

やはり、**ネットワークボトルネック！**
※Stretched Cluster 6-host 実NWで試験

具体的な例：vSAN予兆検知

- こんなこと、経験ありませんか？

2-3年前、従来型ストレージのRAID-10アレイ(FC接続/SAS-Disk)で、1つのHDDのみで異常が発生し、性能低下しているが、障害検知されなかった

(投影のみ)

アレイ全体のLatencyが
上昇し、障害発生

解決まで：1ヶ月以上

(投影のみ)

具体的な例：vSAN予兆検知

- 先月、vSANクラスタで初めてのディスク障害 (SAS SSD)
- キャッシュSSDの障害検知し、ディスクグループ全体が障害になった

vSAN Health (Last checked: Today at 8:58 AM)						
Test Result	Test Name					
✖ Failed	Physical disk					
✖ Failed	Operation health					
Disk		Overall operation health	Metadata	Operational	In CMMDS/VSI	Operational State
Local ATA Disk (mpx.vmhba0:C2:T0:L0)		✖ Failed	✔ Passed	✖ Failed	Yes/Yes	Permanent disk f...

- サーバ側の監視ツール(iLO/ssaci)で、異常検知なし

Storage |  OK

- VMwareサポートの解析で、**SCSIエラー**および**高レーテンシ検知**履歴見つかった
- ハードウェアベンダー (HPE) に依頼し、**SSDの交換してもらった**

具体的な例：vSAN予兆検知

- vSAN Degraded Device Handling (DDHv2) 機能の詳細

参考URL : <https://cormachogan.com/2016/03/08/vsan-6-2-part-10-problematic-disk-handling/>

■Capacityデバイス

- Capacityデバイスは、400分(6-7時間)の間で、500ms超過のWrite遅延がランダムに4回発生した場合に対象デバイスを自動的にアンマウントする

■Cacheデバイス

- デフォルト設定では、CacheデバイスのWrite遅延検知しても、自動的にアンマウントしない
- **/LSOM/IsomSlowTier1DeviceUnmount**を1に設定することで、Write遅延を検知した場合に自動的にアンマウントされるようになる

- デフォルト設定では、今回発生したCacheデバイスのエラーで自動アンマウントされなかつたが、
「/LSOM/IsomSlowTier1DeviceUnmountを必ず1に設定するように」というアドバイスをもらっていた

具体的な例：vSAN予兆検知

- アドバイスをもらったのは、**VMware PSO チーム**
- vSAN Stretched Cluster の設計支援に PSO サービスを利用していた
- 国内で珍しい「**とんがった構成**」だったおかげで、**経験豊富なPSOメンバー**が集まっていた
- 大手他社の案件でCacheデバイス障害で**困った実績から、アドバイス頂いた**
- お陰様で、ディスクエラーの**予兆検知ができる、早期対応ができた**

vSAN Tips

- **All Flash推奨**
 - IOPSバッファーができるので、障害時のResyncやバーストI/Oによる性能低下が吸収できる
 - Erasure Coding (RAID5/6) 使える（※性能低下注意）
 - Read Cache使わないのでStretched Clusterにおけるサイト間移動時の心配がいらなくなる
- **ディスクグループを2つ以上** (Fault Domain増やす、パフォーマンス向上)
- **スペアホスト** (ホスト障害時のリビルド先確保、メンテ時完全データ退避可能)
- **vSAN用に専用10GbE NIC、専用vDS** (※vSAN用L2ネットワーク完全独立時、HAの挙動要注意)
- **バランス構成を保つ** (※主にDisk Group)
- **Witnessの可用性**とバックアップを考慮 (Stretched Cluster時)
- ベンチマークは**HCI Bench**で <https://labs.vmware.com/flings/hcibench>
- **HCLの厳守** <http://vmwa.re/vsanhcl>
 - コントローラFWとドライバの組み合わせが重要
 - SSD firmware (記載バージョン以上であればOK)

Why NSX?

- Act-Actマルチサイトで**ネットワークの自動フェールオーバ**必須だよね…
- VMwareシステムと**相性**が良くて**実績**がある
 - » サードパーティ仮想アプライアンスは、「動くかもしれない」レベルの話しばかり
- 今までできてなかった**ネットワーク運用改善**のチャンス
 - » **可視化** (vRealize Network Insight)
 - » **セキュリティ向上** (マイクロセグメンテーション)
 - » **自動化**
- VMwareだから良い
 - » サードベンダなし、**サポート一本化**
 - » 但し、vSANよりハードルが少し高い…
- 一回導入すれば使い道増えていくはず (#NSXmindset)
- **組織的なバリアがない** (自分たちでネットワークも見てる)
- 今後のHybrid Cloudの可能性
 - » NSXで構成していればもっと簡単にできるはず

NSX概要図（絶賛、PSOでデザイン中！）

NSXがIPv6の動的ルーティング未対応のため、外部Firewallと動的ルーティング用L3SWを物理機器で上位に配置する

外部FW（物理）

NSX Edge の使い方：
・ルータ
・Firewall
・LB（一部）
・VPN/NAT

分散ルータ（DLR）は
当初使わない
※East-Westトラフィックが
10Gbps超過しそうになつたら
導入する

VXLAN-VLAN変換
実施しない構成

分散Firewall

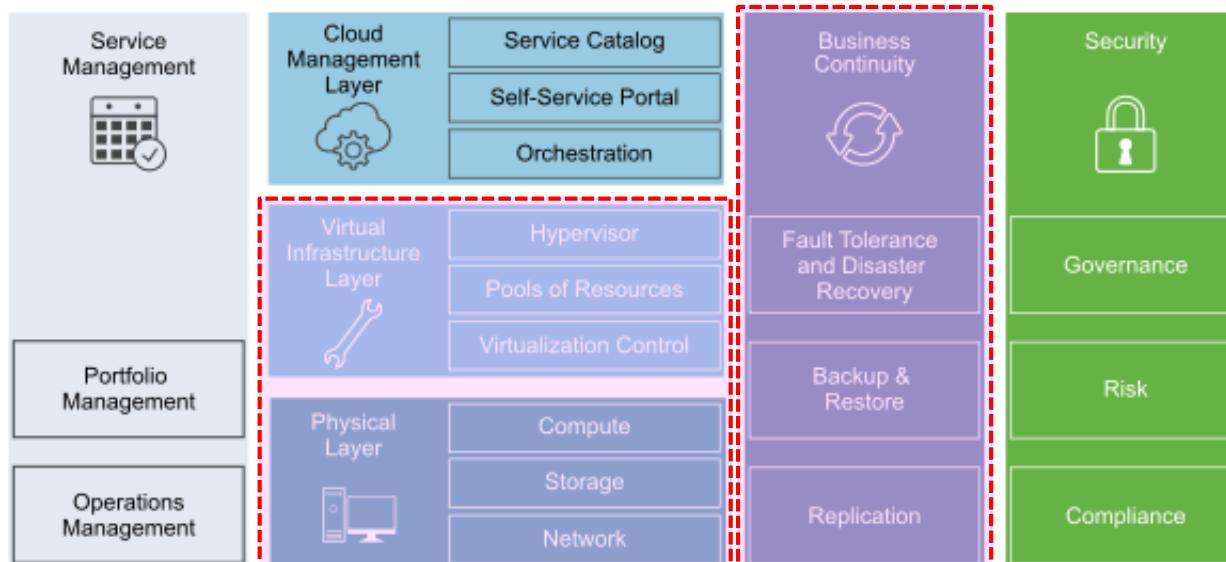
（投影のみ）

今後について

- VMware Validated Design for Software-Defined Data Center (SDDC)

<http://vmwa.re/vvd>

Architecture Overview



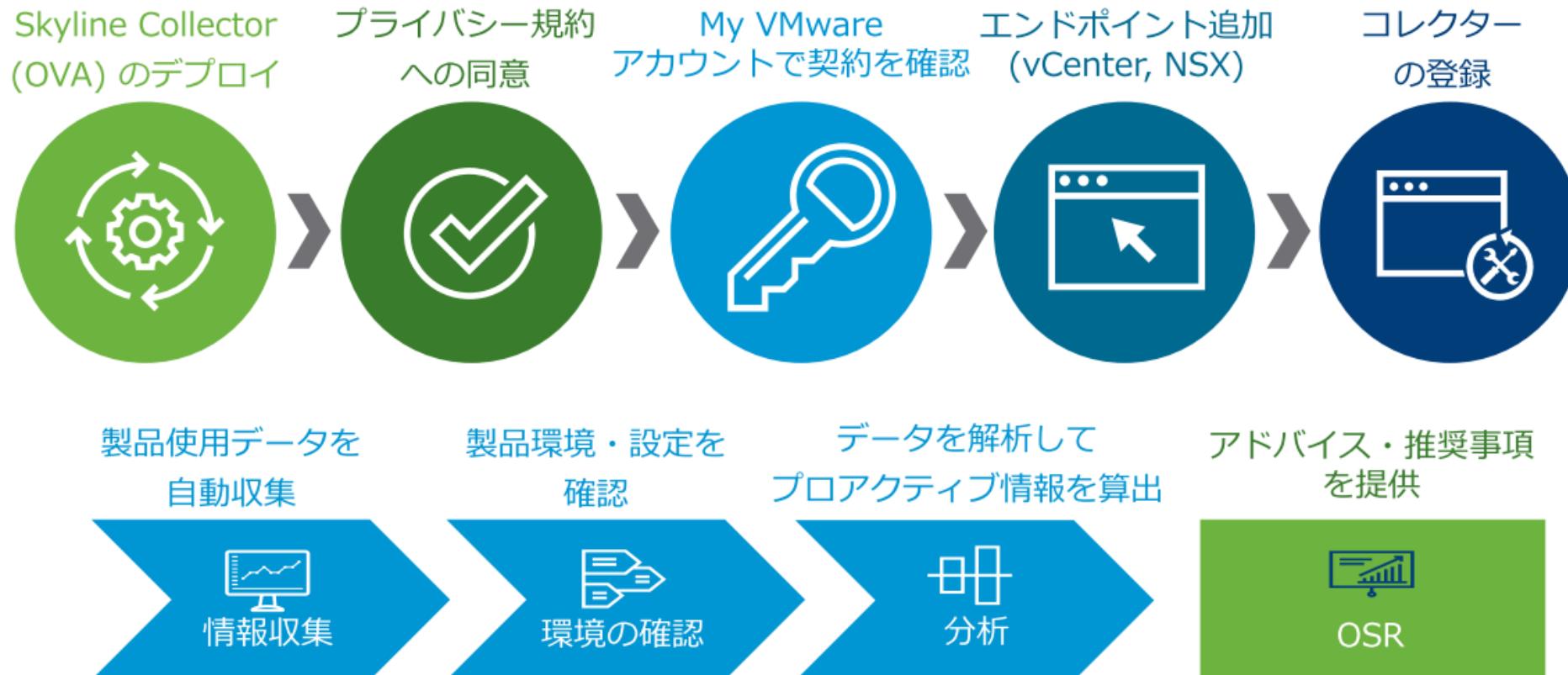
出典 : docs.vmware.com

- 「仮想インフラレイヤ」、「物理レイヤ」、「事業継続レイヤ」は、今回の基盤構築で検討した
- 引き続き、「サービス管理」、「クラウド管理」および「セキュリティ」に注目して行きたい

今後について

- VMwareサポートの新しいサービス「**Skyline**」を導入

・別料金不要
・Premier契約以上

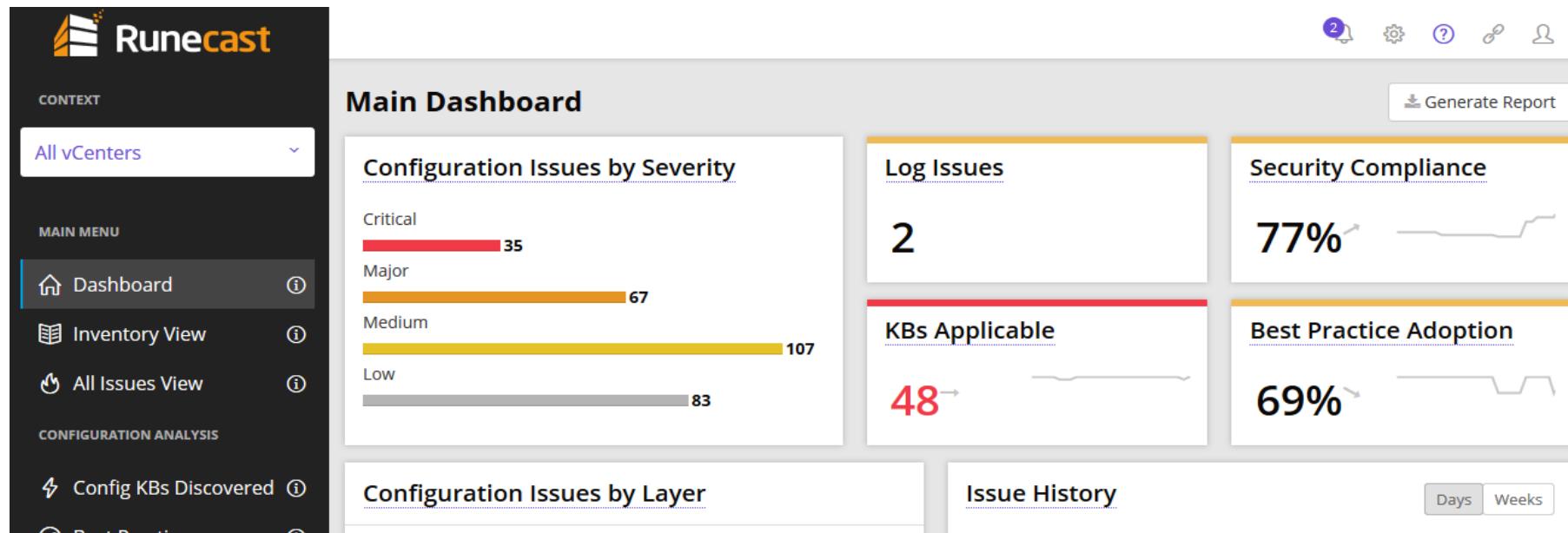


- **Proactiveサポート** (環境分析→推奨情報提供)
- **サポート品質向上** (環境情報GSS側にあるため、迅速化が期待できる)
- 今後、**ログやサポートバンドル収集の自動化**できるようになる予定

出ました！ 2018/11/08

今後について

- Runecast Analyzer



- オンプレ仮想アプライアンス (環境情報収集→定義ファイル使って**オンプレで分析**)
- ログメッセージとKBの紐付**ができる
- KB重要度** (4段階) はRunecast社独自で設定
- セキュリティコンプライアンス** (DISA STIG/PCI DSS/HIPAA) 準拠分析
- ベストプラクティスガイド**の準拠分析

ご清聴ありがとうございました！

ご質問等ございましたら、ご連絡ください。

ニコライ ボヤジエフ
boyadjiev@cnci.co.jp