

【徹底攻略塾】 VMware Wednesday eXtra

VMware vSAN

Performance Deep Dive

ソフトウェアの進化でパフォーマンスも
大幅進化！？

川満 雄樹

VMware株式会社
クラウドプラットフォーム技術統括部

2022年3月30日

Agenda

ソフトウェアの進化でパフォーマンスも大幅進化！？
VMware vSAN Performance Deep Dive

vSAN HCI これまでの進化

vSAN IO の仕組みと性能影響

vSAN Performance

HCIBench を利用した vSAN IO 性能ベンチマーク

vSAN 6.7u3 vs vSAN 7.0.x Performance Update

自己紹介

サーバー・ストレージ・データ保護系メインの物理・仮想化のインフラエンジニア



川満 雄樹 (かわみつ ゆうき)
VMware株式会社
クラウドプラットフォーム技術統括部
シニアHCIスペシャリスト

@yuki_kawamitsu

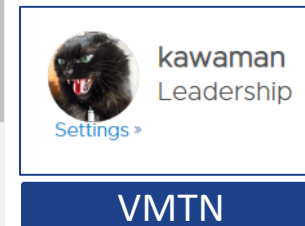
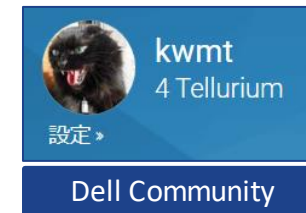
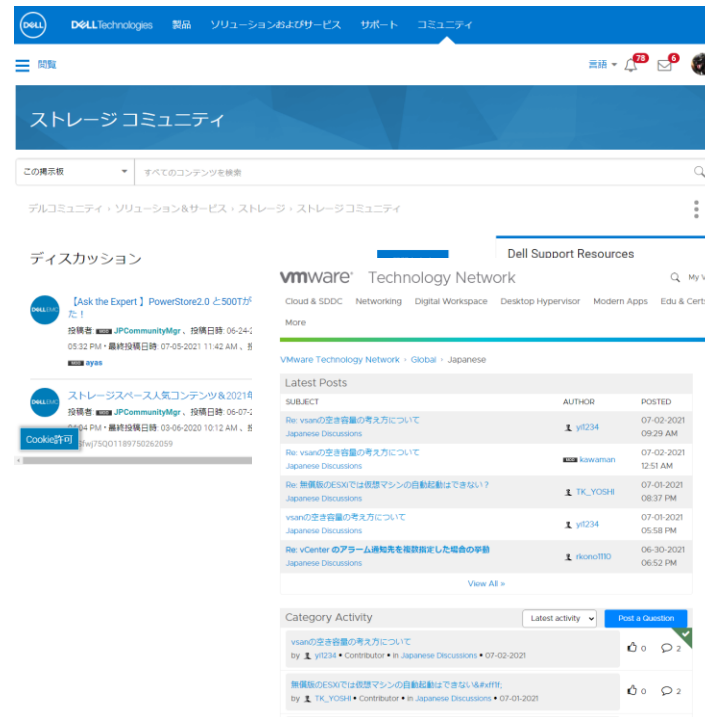
VMware vEXPERT
vExpert 2015 - 2022

<https://vexpert.vmware.com/directory/477>

VMware vEXPERT
★★★★★★

VMware vEXPERT
Cloud Management 2022

VMware vEXPERT
Desktop Hypervisor 2022



業務外でのオンラインコミュニティの回答者活動
(最近 VMTN 日本語フォーラムのモデレータを兼任)

- [VMware Technology Network \(日本語フォーラム\)](https://via.vmw.com/EUZm)
<https://via.vmw.com/EUZm>
- [Dell Community \(ストレージコミュニティ\)](https://via.vmw.com/EUZl)
<https://via.vmw.com/EUZl>

【徹底攻略塾】 VMware Wednesday eXtra : vSAN Deep Dive Series

https://japancatalog.dell.com/c/isg_seminar_vmware_webinar/

2022年 3月 30日 (水) 16:00 - 17:00  本日のセッション

ソフトウェアの進化でパフォーマンスも大幅進化！？
VMware vSAN Performance Deep Dive

2022年 5月 25日 (水) 16:00 - 17:00 (予定)

「支える」アーキテクチャを理解してデザインを考える！
VMware vSAN Architecture Deep Dive

2022年 7月 27日 (水) 16:00 - 17:00 (予定)

vSAN なら運用もこんなにシンプルに！ GUI & CLI vSAN 運用詳細解説
VMware vSAN Management Best Practice & What's Next ?

vSAN 関連お奨め VMWorld 2021 ブレイクアウトセッション

vSAN の最新情報、最新機能のご紹介はこちら

VMware HCI の構成要素 ストレージデバイス

コスト、性能、容量の特性を理解して選ぶ

ストレージデバイス

キャッシュとして利用する **キャッシュデバイス**
データの保管のために利用する **キャパシティデバイス**

デバイス	Queue Depth
IO コントローラ	256 or 512以上 (ReadyNode規定)
NVMe	2048 - 4096
SAS (SSD/HDD)	254
SATA(SSD/HDD)	32

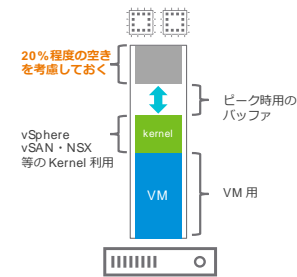
パフォーマンスと価格

容量

vmware © 2021 VMware, Inc.

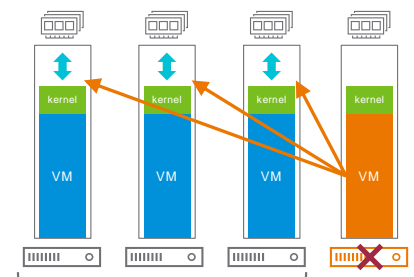
余剰を見ておく（監視する）べき CPU・メモリリソース

CPU サイズング (Max 80% の考慮)



システム負荷のピーク時に VM の負荷と Kernel の負荷の合計が 80% 前後におさまるのがベスト

メモリサイズング (N+X の考慮)



ホスト障害時に HA で起動する VM やメンテナンス時の移行分の VM メモリリソースの余裕を確保する事が推奨 (N+1 や N+2 など可用性を考慮したサイズング)

vmware © 2021 VMware, Inc.

VMware Cloud Foundation, VMware vSAN HCI インフラ設計 Deep Dive

MC11110 (川満 雄樹 担当)

<https://vmworld.jp/content/MC11110/>

安定性と性能向上の鍵はアップグレード！ あなたの知らない VMware vSAN の進化を徹底解説

MC11108 (小佐野 舞 担当)

<https://vmworld.jp/content/MC11108/>

vSphere / vSAN オンラインセミナー #42

アセスメント&サイジングツールを活用した vSphere / vSAN 基盤サイジング手法

vmware IT 価値創造塾

ホーム 入門者向け 課題を解決 導入事例 導入について ビデオギャラリー 資料ダウンロード セミナー

ビデオギャラリー

【ビデオ】 vSphere / vSAN オンラインセミナー #42 アセスメント&サイジングツールを活用した vSphere / vSAN 基盤サイジング手法

2022/03/10

この動画はご好評いただきましたオンラインセミナー「vSphere / vSAN オンラインセミナー #42 アセスメント&サイジングツールを活用した vSphere / vSAN 基盤サイジング手法」(2022年2月9日)のレコーディングです。

vSphere / vSAN 基盤のサイジング等に関して根拠や考慮点がわからないという事も多いのではないのでしょうか?本セッションではご提案内容と比較・精査される方も、ご提案される方もどちらも知っておくと安心なサイジングのポイントを実際にツールを操作しながらご紹介いたします。

アセスメント&サイジングツールを活用した vSphere / vSAN 基盤サイジング手法

vSphere / vSAN オンラインセミナー #42

川島 雄樹 (Yuki Kawamitsu)
ワイエムウェア株式会社
クラウドプラットフォーム技術統括部
2022/2/9

ニュースレターを購読
IT 価値創造塾の更新情報や、VMware の最新情報、セミナーやイベントなどの情報を不定期でお届けします。
ニュースレター購読のお申し込み >

関連する記事
【VMware vSAN 導入事例】ローテック株式会社様
【ビデオ】 vSphere / vSAN オンラインセミナー #41 vSphere / vSAN 最新情報アップデート!
【ビデオ】 vSphere / vSAN オンラインセミナー #40 共有データストアでインフラの可用性を向上、小規模 HCI なら 2 ノード vSAN におまかせ!
【ビデオ】 vSphere / vSAN オンラインセミナー #39 VMware Tanzu on VxRail 環境のコンテナバックアップの検証結果
【ビデオ】 vSphere / vSAN オンラインセミナー #38 これさえ知ってれば大丈夫! HCI の運用管理の基本から応用まで!

関連する資料ダウンロード
アセスメント&サイジングツールを活用した vSphere / vSAN 基盤サイジング手法
オンラインセミナー掲載資料 - vSphere / vSAN 最新情報アップデート!

<https://juku-jp.vmware.com/video/vsphere-vsan-os-042/>

vSAN 容量サイジングの基本的計算式

vSAN Ready Node Sizer 裏の計算方法を理解する事でサイジングミスリスクを排除

① ホスト数	② キャパシティドライブ数 (TB)	③ キャパシティドライブ合計本数 (ホスト通り)	④ 奇数の2倍数換算 (TB → MB)	⑤ スラックススペース (%)	⑥ vSAN 実効容量 (TiB)					
4 台	x	3.84 TB	x	6 本	x	0.909	x	30%	=	58.6 TiB

TB → TiB 換算を忘れずに!

vSAN 7.0u1 以降ではスラックススペース (メンテナンス用空き容量) が従来より可変的に節約できるようになりました。
詳細は次頁を参照

vSAN 容量サイジング : スラックススペースの考慮

vSAN も N+1 のサイジングが安全を考慮した鉄板デザイン

vSAN クラスタではデータのリバランス、ポリシー変更によるデータ再構成、ホスト故障時のデータ再保護・再配置など安定的に利用するために必要なメンテナンス領域(2%~30%)をあらかじめサイジングに含め確保します。

vSAN 7.0u1 以降では操作領域予約: Operations Reserve (OR) 10% + ホスト数に応じた N+1 ホスト障害時予約: Host Rebuild Reserve (HBR) を考慮したサイジングを行います。
※ Host Rebuild Reserve (HBR) を含める事で、vSAN サイジングは N+1 で構成される事と同意となります。
※ 詳細は [VMware vSAN Design Guide](https://core.vmware.com/resource/vsphere-vsan-design-guide) <https://core.vmware.com/resource/vsphere-vsan-design-guide> 参照。

ノード数	OR (%)	HBR (%)	合計予約容量 (%)
4ノードクラスター	10%	25%	約 30%
6ノードクラスター	10%	17%	約 27%
8ノードクラスター	10%	13%	約 23%
12ノードクラスター	10%	8%	約 18%
18ノードクラスター	10%	6%	約 16%
24ノードクラスター	10%	4%	約 14%
32ノードクラスター	10%	3%	約 13%
48ノードクラスター	10%	2%	約 12%
64ノードクラスター	10%	2%	約 12%

予約とアラート | vSAN Designer

vSAN 7.0u1 以降では UI で予約とアラートが設定可能

vSAN HCI これまでの進化

vSAN の歴史ダイジェスト

VMware vSAN これまでの進化

2014年 以来 VMware HCI のスタンダードとして日々進化するテクノロジー



2014 / 03 ~

2015 / 03 ~

2015 / 09 ~

2016 / 03 ~

2016 / 11 ~

2017 / 07 ~

2018 / 4 ~

ストレージ基本機能

- Hybrid vSAN
- All Flash vSAN
- ポリシーベース管理
- RAID1



ストレージオプション機能

- 重複排除 / 圧縮
- RAID5 / 6
- vSAN 健全性モニタ
- QoS
- 2 Node vSAN
- Stretched Cluster



性能・安定性向上・運用機能

- HTML5 Client / vROps 連携
- Update Manager 連携
- Container / Cloud Native Apps
- vSAN Support Insight
- vSAN 暗号化
- Adaptive / Parallel Resync
- デスページ処理最適化
- TRIM / UNMAP

VMware vSAN これまでの進化

2014年 以来 VMware HCI のスタンダードとして日々進化するテクノロジー



性能・安定性向上・運用機能

- HTML5 Client / vROps 連携
- Update Manager 連携
- Container / Cloud Native Apps
- vSAN Support Insight
- vSAN 暗号化
- Adaptive / Parallel Resync
- デステージ処理最適化
- TRIM / UNMAP

Cloud Native / Container Native

- vSphere with Kubernetes
- vSAN File Service
- vSphere Lifecycle Manager
- vSphere Replication 連携強化
- Stretched Cluster 機能強化 (DRS・Resync)
- NVMe Hot Plug

7.x での機能をさらに強化

- HCI Mesh 強化
(Computing Only Cluster)
- vSAN File Service の拡張
- vSphere Native キー管理による暗号化サポート
- Encryption Key Persistence (TPM 2.0)
- vSphere Lifecycle Manager 強化
- パフォーマンス強化 (AMD CPU への最適化)
- 健全性履歴閲覧
- IO Trip Analyzer / Network Analyzer
- vSAN Cluster Sequential Shutdown

vSAN IO の仕組みと性能影響

vSAN の性能を左右する考慮点と
ボトルネック解消のためのポイント

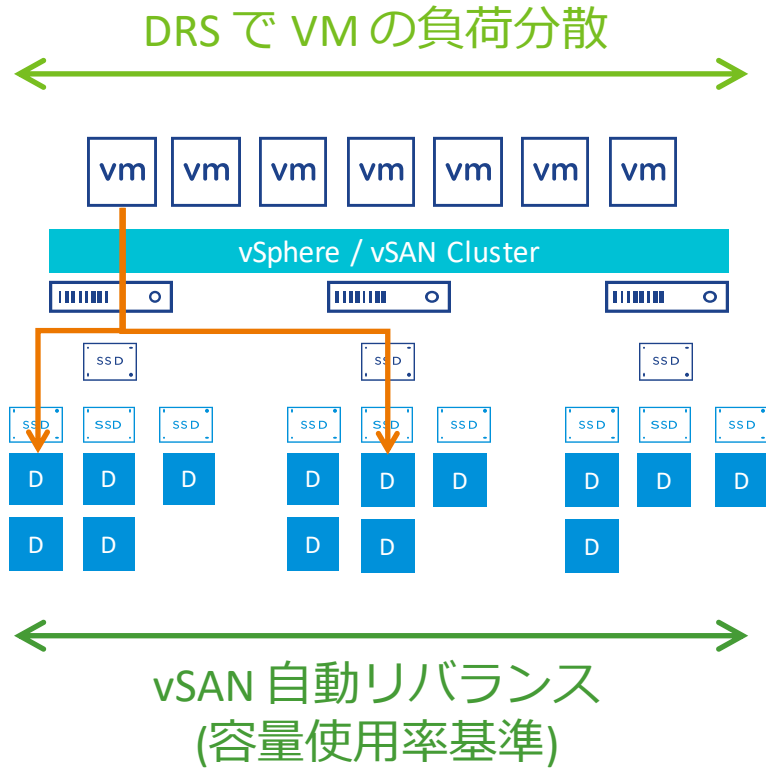
vSAN IO の仕組みと性能影響

vSAN の性能を左右する考慮点とボトルネック解消のためのポイント

- 1. vSAN クラスタデザイン・ディスクグループ構成による性能への影響**
 - 本日のセッションと次回のセッションで詳細解説
- 2. ネットワークはシンプルに構成、MTU9000・広帯域・低遅延ネットワークの利用**
 - 次回のセッションでクラスタデザインと併せて詳細解説
- 3. ドライブタイプ別の性能の違い**
 - 本日のセッションで解説、次回のセッションでさらに詳細検証結果と共に解説
- 4. ストレージポリシーと性能の関係**
 - 本日のセッションで性能検証結果含めて解説
- 5. より最新のvSANバージョンほど性能を高く発揮**
 - 本日のセッションで過去バージョンとの性能比較を解説

必要なのはスケールアップ？ スケールアウト？

vSAN の性能特性を考える



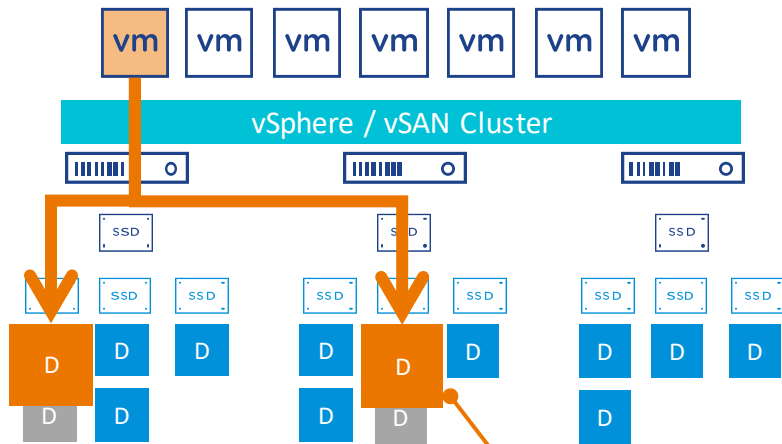
基本の使われ方はクラスタ全体でのリソース負荷のバランスングです

様々なサイズの VM が均等にクラスタ内に分散し、CPU・メモリ・ディスクのリソースを共有しあいながら消費する形が vSAN HCI にとって最も効率の良い使われ方となります

- DRS 機能を有効にして VM の負荷に応じた配置自動制御を推奨
- vSAN 自動リバランス（閾値ベース）は有効化した運用を推奨

必要なのはスケールアップ？ スケールアウト？

モンスターVMをどうカバーするか？



標準のvSAN構成（非重複排除）の場合は実データはランダムにSSDに冗長配置されるので負荷インパクトはそこに集中してしまう

特定のVMがストレージIO含めてリソース増大したらどうなるでしょう？

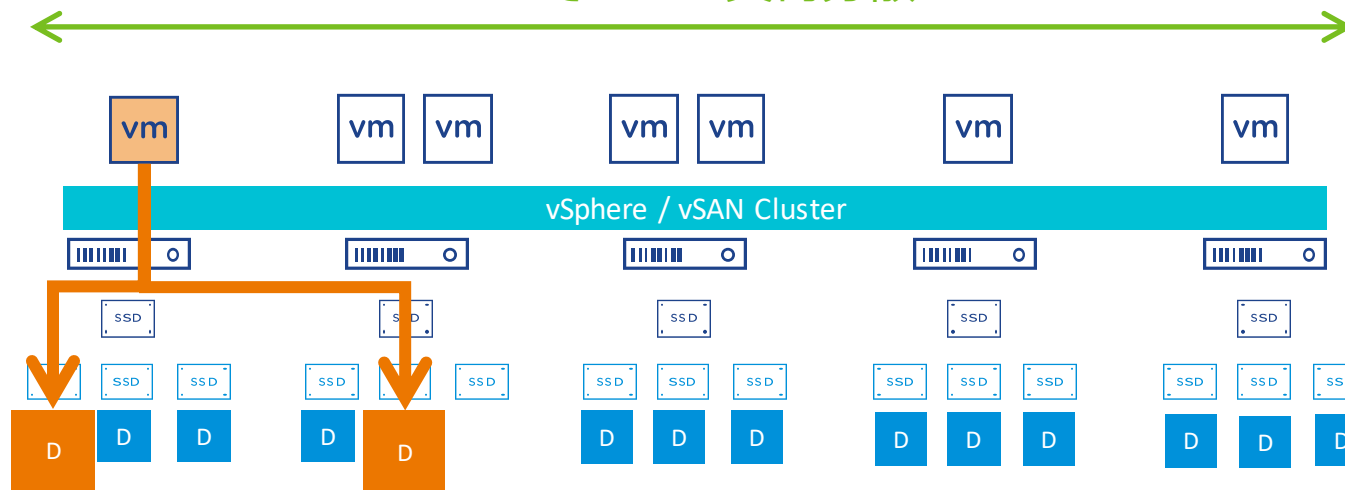
想定を超える高いIOPSやCPU・メモリを要求するVMや、1ESXiに収まりきれない大容量VMの運用が必要な場合、解決方法として何が最適でしょうか？

必要なのはスケールアップ？ スケールアウト？

モンスターVMをどうカバーするか？



DRSでVMの負荷分散



特定のVM (VMDKファイル) に集中したIOはvSANのリバランスで負荷分散されるわけではないので注意

vSAN 自動リバランス (容量使用率基準)

ESXi ホストを追加しスケールアウトでリソースを増強した場合、単純なスケールアウトでは対象VMのIOが分散されるわけではないため、一か所集中した性能問題の解決は難しい

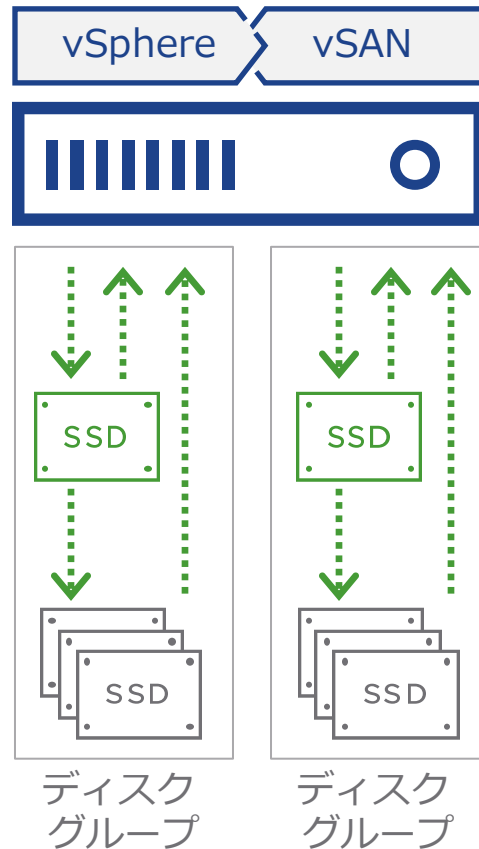
IO 特性の理解が重要

- 基本的には物理SSDの最大IO性能以上を発揮させるのは難しい (ストライプなどの手法は利用可能)
- ストレージポリシーでIO制限を掛ける事も可能ですが、必要なIOを抑えつける事は根本的な解決とはならない

vSAN を構成するキャッシュ層とキャパシティ層の役割

キャッシュ/バッファ層とキャパシティ層からなる2層のシステム

All Flash vSAN



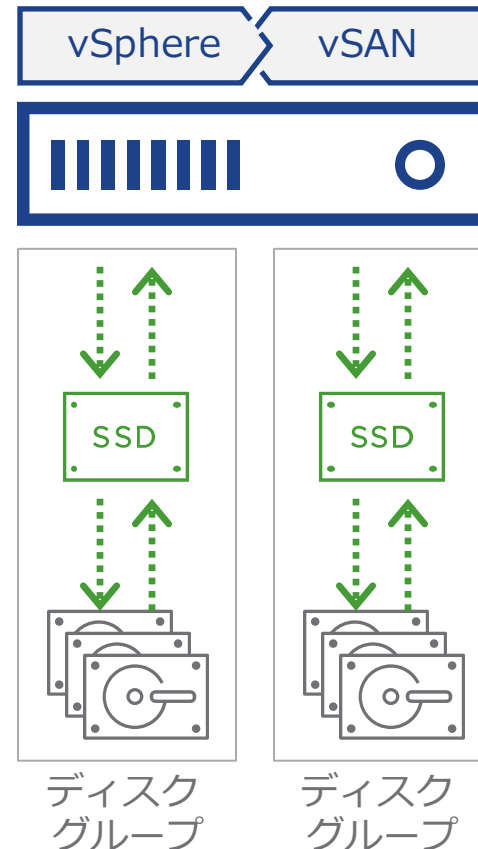
バッファ = 100 %

キャパシティ

ディスクグループ

ディスクグループ

Hybrid vSAN



バッファ = 30 %
キャッシュ = 70 %

キャパシティ

ディスクグループ

ディスクグループ

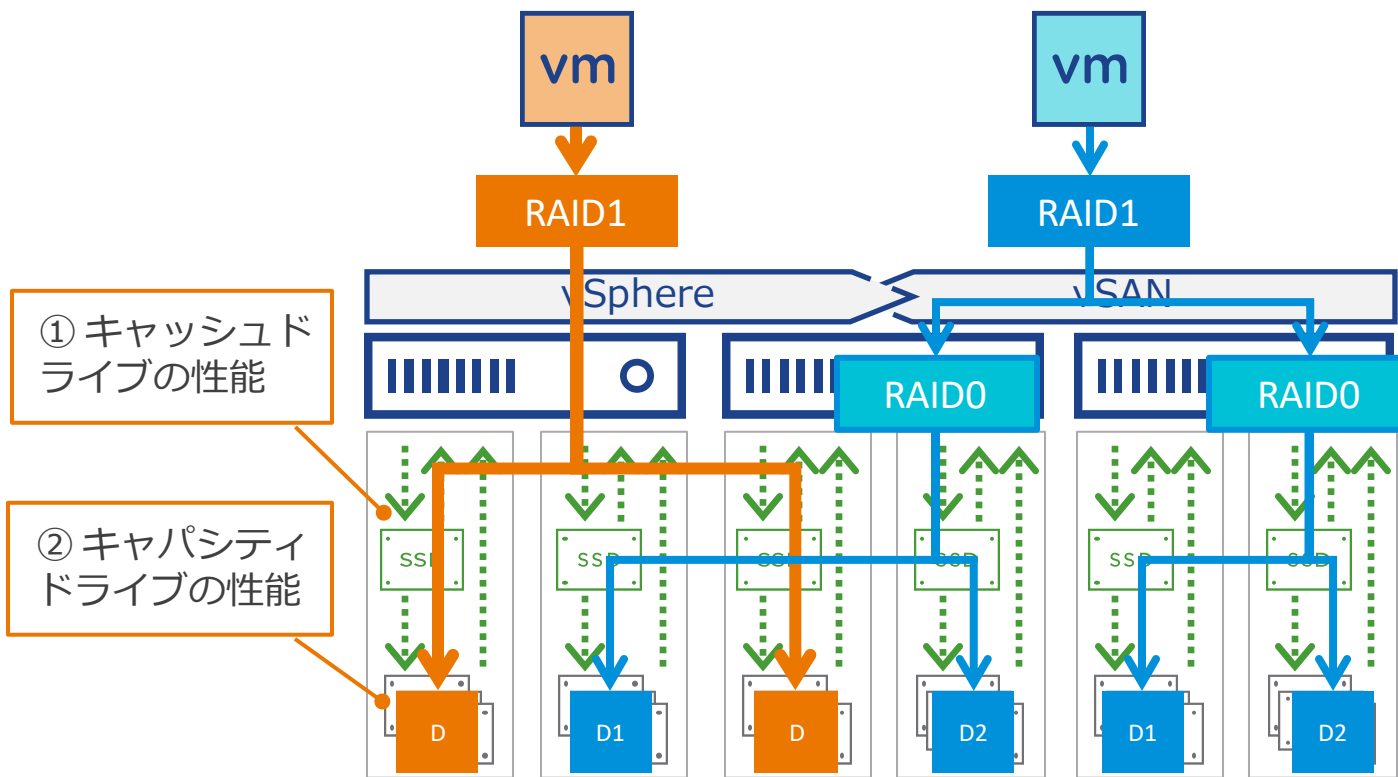
AF vSAN と HY vSAN の違い

- ハイブリッドでは、キャッシュ層の70%が読み取りキャッシュに割り当てられる
- オールフラッシュでは、キャッシュ層全体が書き込みバッファに割り当てられる (デステージ前の書き込みバッファデータも読み取り取得可能)

書き込みは、バッファに到達したときに仮想マシンに確認応答が返され、その後順次キャパシティにデステージされる

ドライブ選定と vSAN 性能の考慮点

ワークロード要件に合わせたモデル選定



① キャッシュドライブの性能

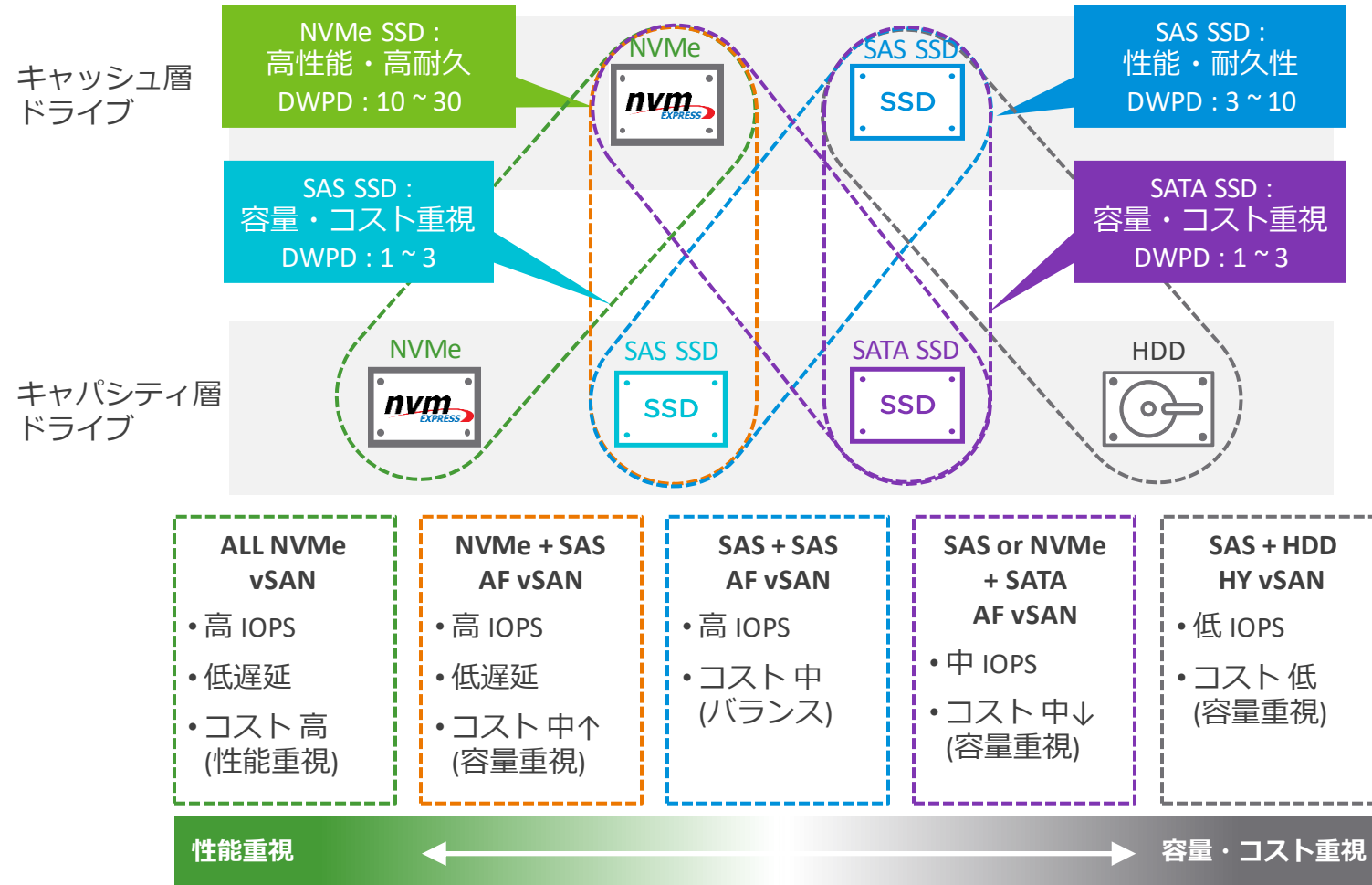
② キャパシティドライブの性能

- ① VM の IO 性能に大きく響くのが **キャッシュドライブの Max IO 性能と耐久性**
- ② キャパシティドライブの考慮で重要なのが沢山の VM の同時 IO (並列 IO) を受ける処理能力があるか否か、**Queue Depth の値に要注意**
- ③ 単一ドライブ以上の性能を求めるときには**ストライプ**のポリシーで IO の分散が効果的
※ RAID ポリシーの違いによる性能の違いは後述の検証結果参照

③ ストレージポリシーによる制御・特性

vSAN ドライブ・ディスクグループ構成

着目すべきはドライブの"耐久性" とインターフェースタイプ



キャッシュ層ドライブの推奨値

- **キャッシュ層ドライブは耐久性 (TDW・DWPD) で容量を計算**
※ 毎日の書き込み容量に対する耐久性でサイジング
 - 10 DWPD 以上の耐久性の Write Intensive (WI) クラス SSD を推奨
 - 3 ~ 5 DWPD の Mix Use (MU) クラスの SSD の場合は容量を多めにサイジング
- **HY vSAN ではキャパシティ容量の 10% の容量でサイジング**

※ DWPD (Drive Writes Per Day) : 5 年間 SSD の全容量を 1 日何回上書きする事が可能か耐久性を示す指標

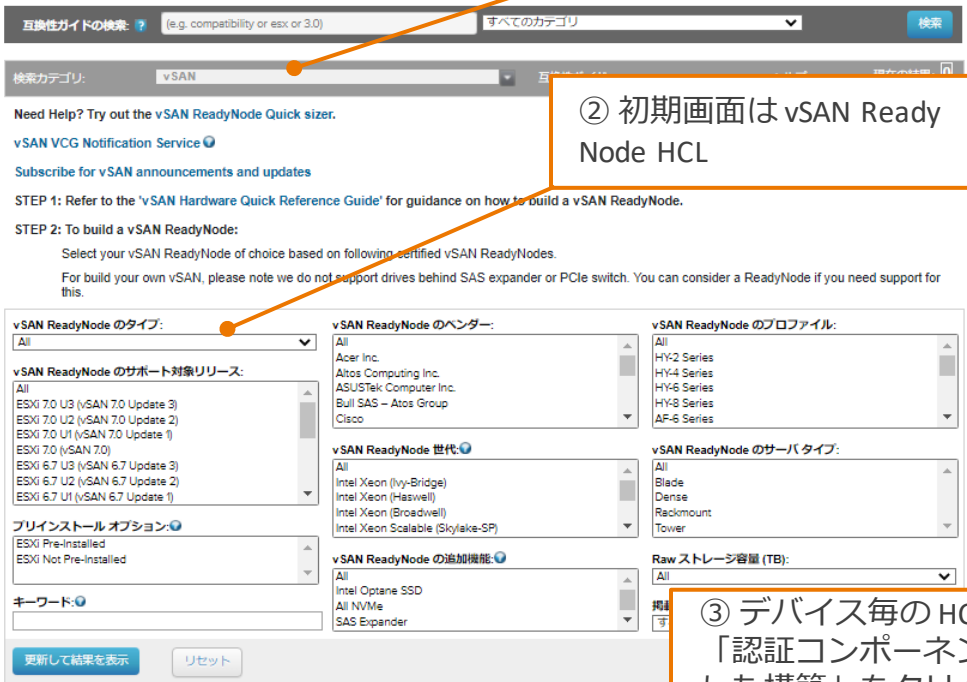
- 例 : 10 DWPD = 1 日 10 回全データを書き換えても5年間書き込める耐久性

vSAN HCL から確認するデバイス性能

<https://www.vmware.com/resources/compatibility/search.php?deviceCategory=vsan>

VMware Compatibility Guide

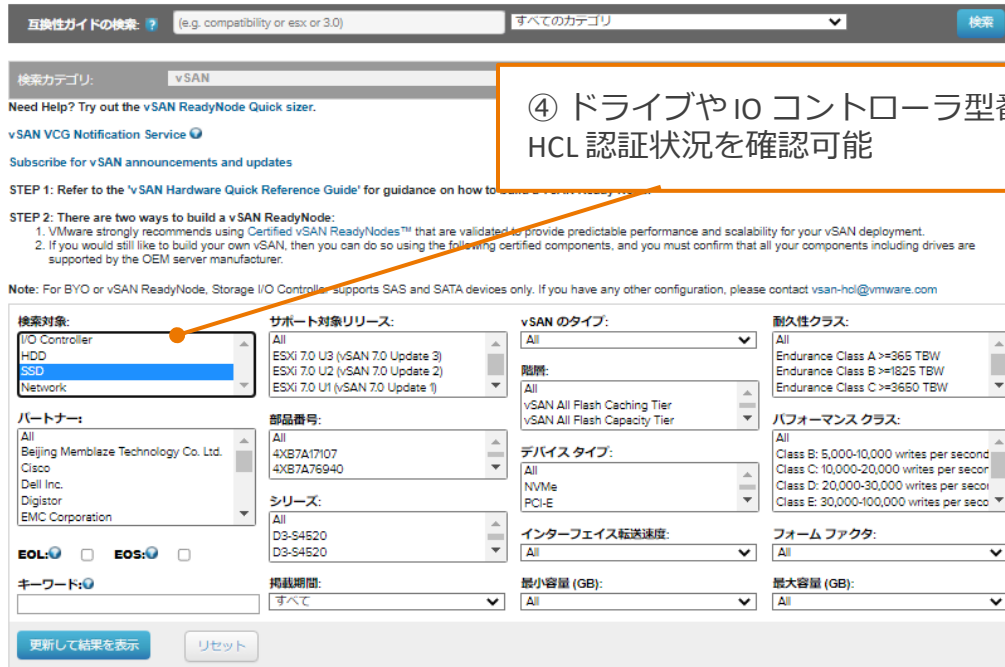
① カテゴリ : vSAN



② 初期画面は vSAN Ready Node HCL

③ デバイス毎の HCL 詳細確認は「認証コンポーネントを基盤とした構築」をクリック

VMware Compatibility Guide



④ ドライブや IO コントローラ型番毎の HCL 認証状況を確認可能

vSAN 健全性チェックのオフライン使用のために vSAN ハードウェア互換性リストを JSON 形式でダウンロードするには [ここをクリック](#) ファイルは使用する前に解凍してください。

VMware 製品のバージョン別の互換性情報: [VMware 製品の相互運用性マトリックスを参照してください。](#)

vSAN のサイジング サポート: VMware パートナーがお客様の環境を評価し、vSAN のメリットを検証いたします。詳細については、[vSAN 評価ツール](#)をご覧ください。

vSAN HCL から確認するデバイス性能

vSAN HCL 認証済みデバイスのIO 性能指標は必ずチェック

耐久性クラス D 以上なので高い負荷の All Flash vSAN のキャッシュで利用可能

400GB の SSD で 7,300TBW の耐久性なので変換すると **10 DWPD**
※ $7,300 \text{ TBW} / (0.4 \text{ TB} \times 365 \times 5) = 10 \text{ DWPD}$

※ TBW は 以下の公式で DWPD に変換可能 (5年製品保証の場合)
 $\text{TBW (Terabytes Written)} = \text{Drive Size} \times \text{DWPD} \times 365 \times 5$
 $\text{DWPD (Drive Write Per Day)} = \text{TBW} / (\text{Drive Size} \times 365 \times 5)$

VMware Compatibility Guide

検索結果に戻る

モデルの詳細

モデル: 400GB Solid State Drive SAS Write Intensive MLC 12Gbps 2.5 inch Hot-Plug Drive

デバイスタイプ: SAS

耐久性: 7300 TBW

耐久性クラス: Endurance Class D >=7300 TBW

部品番号: 5VHHG

容量: 400 GB

シリーズ: PX05SMB

注:

パートナー名: Dell Inc.

ベンダー ID: TOSHIBA

パフォーマンスクラス: Class E: 30,000-100,000 writes per second

フラッシュテクノロジー: eMLC

インターフェイス転送速度: 12 Gbps

フォーム ファクタ: 2.5"

製品 ID: PX05SMB040Y

View History | rss feed

Click here to export this page: Export to CSV

リリースの詳細

VMware 製品名: ESXi 7.0 U3 (vSAN 7.0 Update 3)

Release	Tier	Minimum_Firmware_Version
ESXi 7.0 U3 (vSAN 7.0 Update 3)	vSAN All Flash Caching Tier vSAN Hybrid Caching Tier vSAN All Flash Capacity Tier	AS03

機能カテゴリ	機能
Mandatory	Drive Performance, Drive Reliability, Queue Depth, Surprise Power Removal Protection, Write Cache, Write Failure Notification
Supported feature	vSAN Secure-wipe Capable

SAS インターフェースであることが分かる ※ インターフェースタイプは Queue Depth の把握に必須

パフォーマンスクラス E なので、SSD 1本あたり 30,000 IO 以上の性能が最大で発揮可能

各 ESXi バージョン毎にサポートされるファームウェアバージョン、ドライババージョンが確認可能 (IO コントローラの場合)

vSAN ハードウェア クイックリファレンス ガイド

https://www.vmware.com/resources/compatibility/vsan_profile.html

SSD パフォーマンス クラス	1 秒あたりの書き込み数
B	5,000 - 9,999
C	10,000 - 19,999
D	20,000 - 29,999
E	30,000 - 99,999
F	100,000 - 349,999
G+	350000 +

SSD 耐久性 クラス	SSD 層	書込耐久性 (TBW)
A	VSAN All Flash - Capacity	365
B	VSAN Hybrid - Caching	1825
C	VSAN All Flash - Caching for Medium workloads	3650
D	VSAN All Flash - Caching for High workloads	7300 +

※ vSANハードウェアクイックリファレンスガイドでは vSAN Ready Node の定義、構成の考慮点の詳細が解説されています。
各 SSD ドライブに定義された性能・耐久性指標もここで確認が可能です。

DWPD レベル	SSD タイプ	用途
1	Read Intensive (読込特化)	キャパシティ
3 ~ 5	Mix Use	キャパシティ キャッシュ (HY vSAN)
10 ~	Write Intensive (書込特化)	キャッシュ (AF vSAN / HY vSAN)
30 ~	Write Intensive (超書込特化)	キャッシュ (AF vSAN / All NVMe)

vSAN ドライブ選定時の注意点：インターフェースのキュー深度

キュー深度 (Queue Depth : QD) : IOPS・スループット以外にもIO性能を左右する重要な指標

Queue Depth

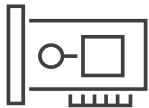


NVMe



PCIe

NVMe デバイスは規格上は最大 **Queue Depth : 65,535 (64k)**、さらに **Queue そのものも 64k 個** で並列処理に強い従来 SCSI コマンドと比べシンプルなIO処理で、HBA を経由せず PCIe ダイレクトにIOを処理するので低遅延・高IOPSを実現可能。近年は3DxPoint SSD など超高耐久、高性能デバイスが登場
※ 実際の ESXi での Max Queue は製品・ドライバにより異なるが 2048 ~ 4096 など設定される事が多く、esxcli コマンドで確認可能



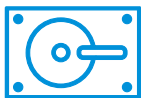
HBA (RAIDカード / PathThroughカード)

vSAN Ready Node の規定は **512 以上の QD** を持つ

2021 年時点の各社 HBA の性能は **1,024 / 2,048 / 4,096 / 9,000 以上** など、かなり高い性能を持つものが多い



SAS SSD



SAS HDD

主流の SAS はデバイス IF 辺りの **Queue は 1 つ**、**Queue Depth は 254**

4000 以上の QD を持つ HBA であれば、16 個以上のドライブを搭載しても十分に性能を発揮する事が可能



SATA SSD



SATA HDD

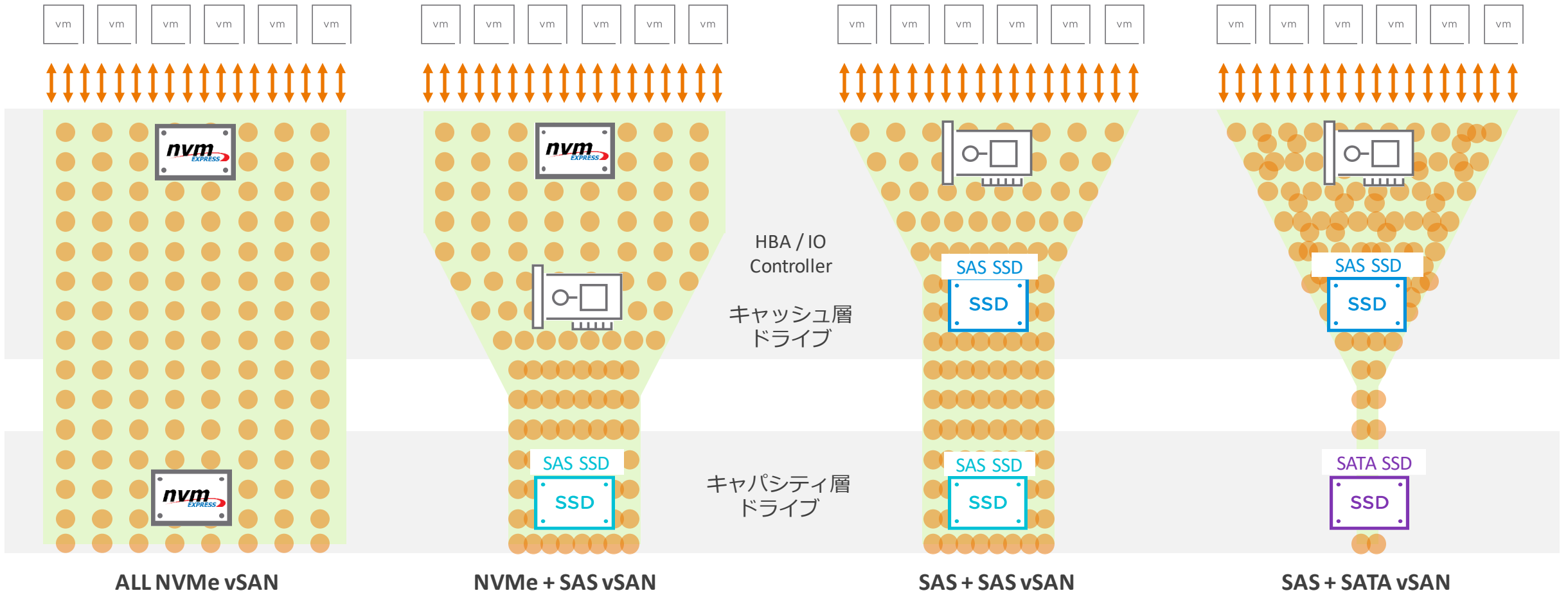
SATA はデバイス IF 辺りの **Queue は 1 つ**、**Queue Depth は 32**

多数の仮想マシンが集約され、多くの並列 IO が発行される環境では最終的な SATA キャパシティドライブの QD の浅さがボトルネックになるリスクがあり、サイジングには注意が必要

PoC 貸出機の SATA 構成や価格勝負になって突っ込んだ SATA 構成での性能問題多発 !

どこがIOのボトルネックになるかを考える

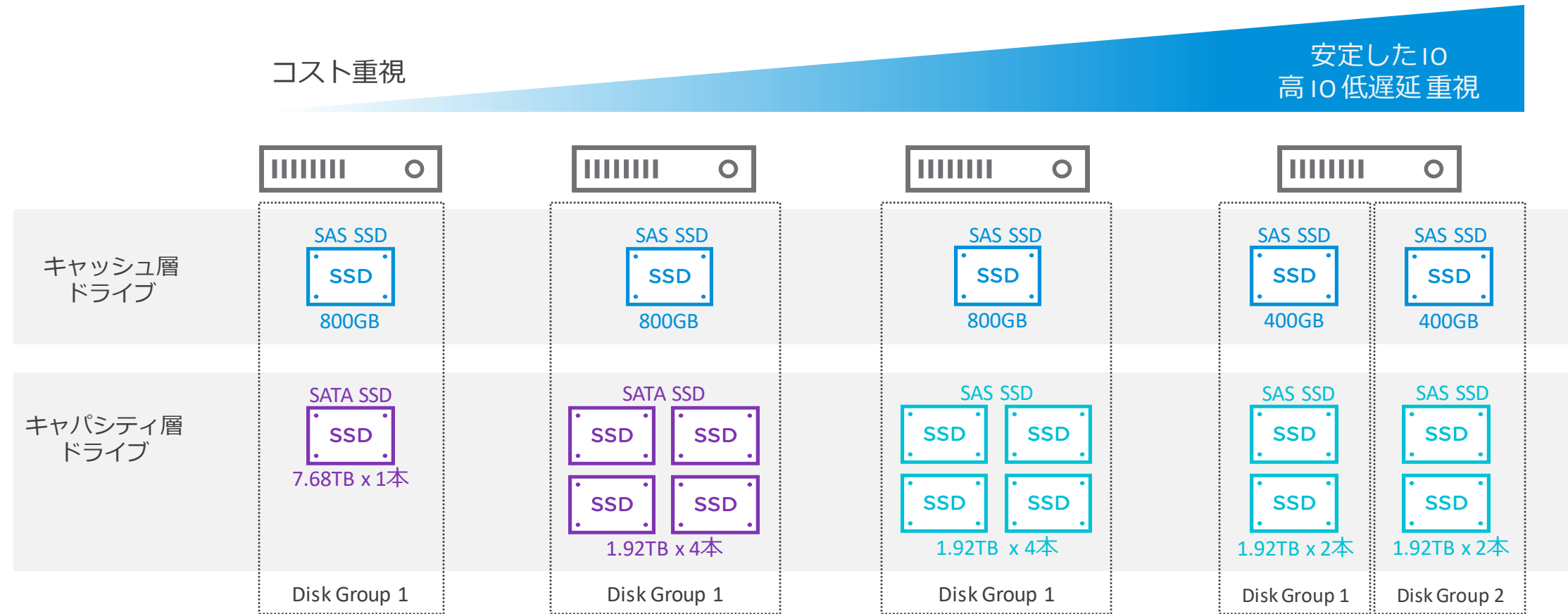
仮想マシン⇔キャッシュ層ドライブ⇔キャパシティ層ドライブ間のIO密度のイメージ



※ この図は説明のためのイメージであり、実際のIOは各ドライブからHBA・PCIe介して双方向通信なのでさらに複雑です

キャパシティの分散・複数ディスクグループ構成の推奨

多数の並列した仮想環境のIO性能を考慮した分散デザインの推奨

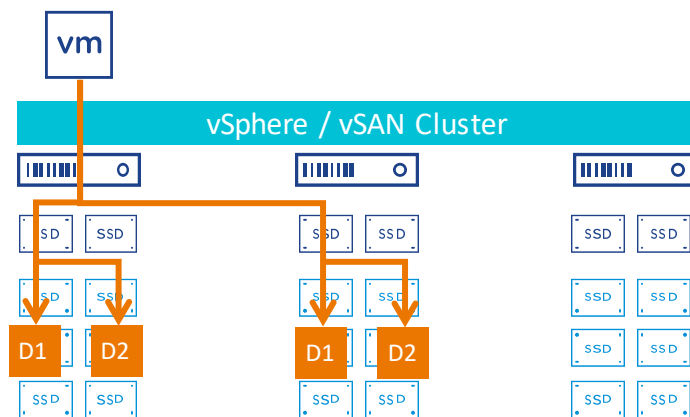


vSANのIO劣化（キャッシュ枯渇）はドライブのIO性能（Queue Depth）が低い場合に発生する傾向が高く、QD：32のSATAドライブではなくQD：254のSASドライブを並列にした構成でバッファ溢れによる性能低下を起こりにくくすることが可能

① 性能強化にスケールアップが適切な例

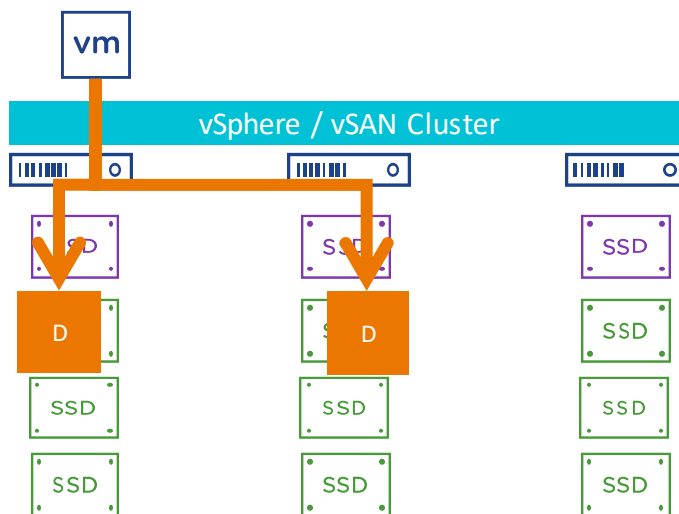
モンスターVMを適切な設計で制御、再配置する事がポイント

ドライブ・ディスクグループ増設



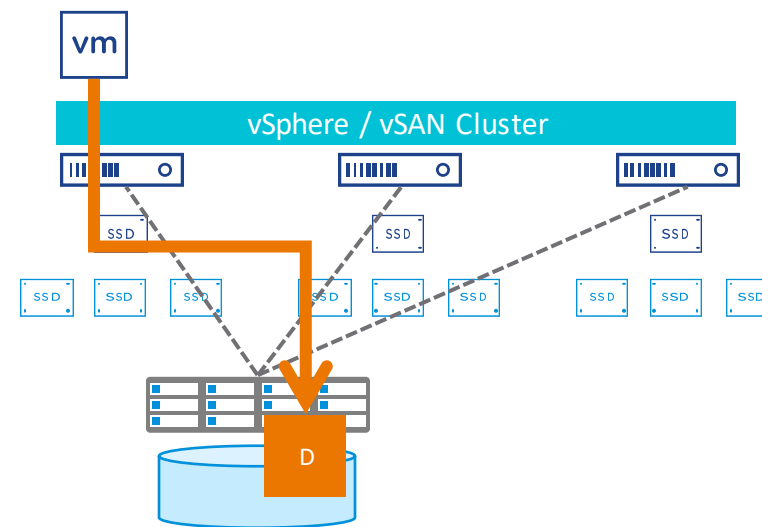
- ディスクグループ増設はIOの分散に有効に働き、ストライプ設定をする事で高IOに対応も可能
- ノード増設不要なのでライセンスの追加コストは発生しません

高性能・大容量デバイスの利用



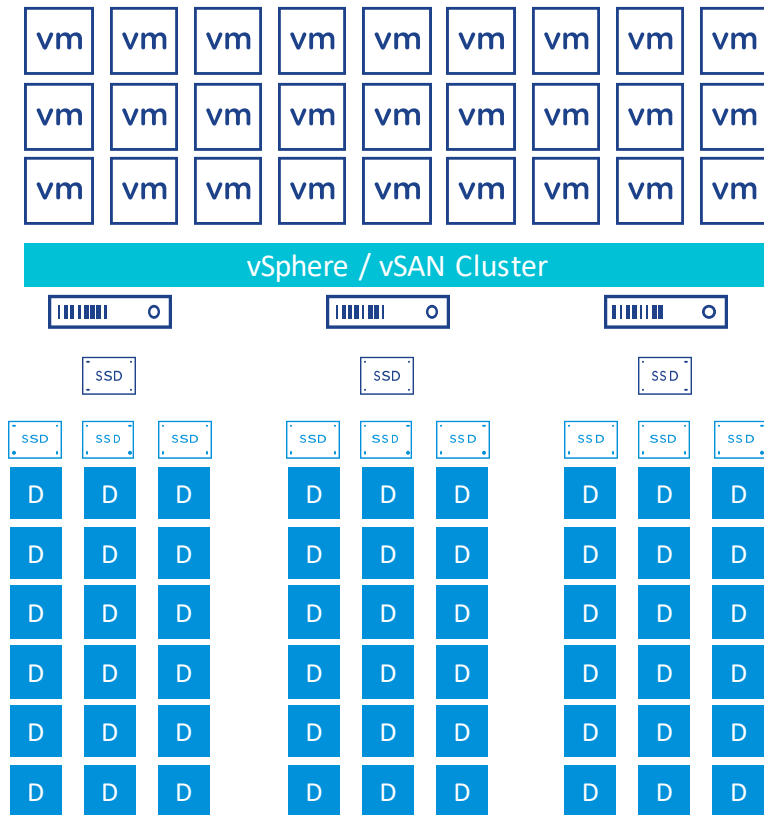
- 超高IO VMでの利用がメインとなる場合は3DxP NVMeドライブの利用などで高パフォーマンスvSAN基盤を組むことが可能
- ワークロードに合わせたクラスタ設計が重要

特定用途向けストレージの併用



- 特定の超高IO VM向けに専用の外部ストレージを利用する、逆にほとんどIOを必要としない大容量アーカイブ用途で外部のアーカイブストレージを利用する等、vSANクラスタは外部ストレージの併用で適材適所の利用が可能

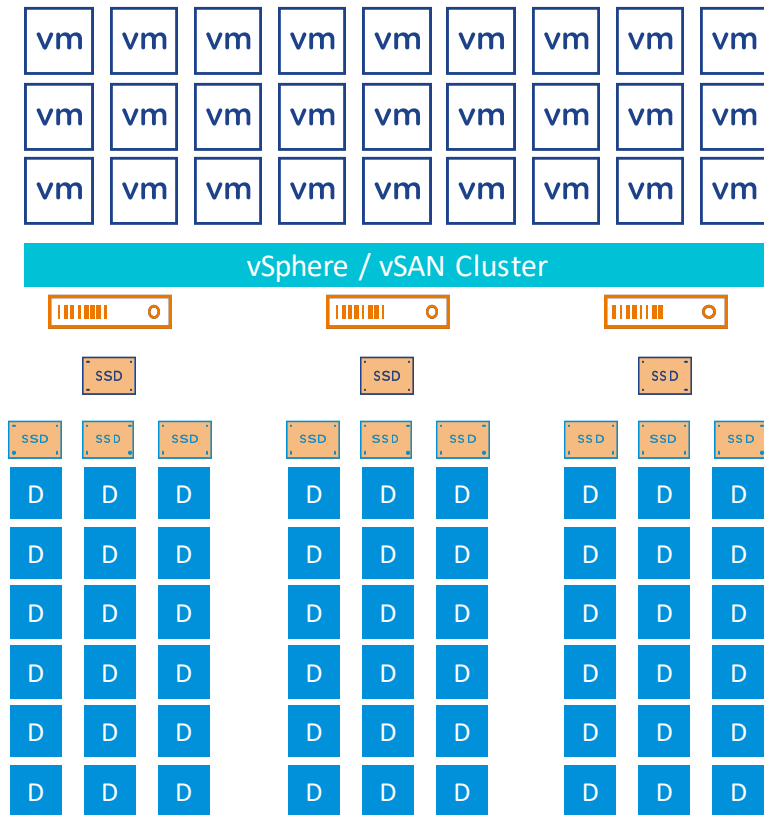
② 性能強化にスケールアウトが適切な例



一般的な仮想化基盤では徐々に稼働するシステム（VM）が増加する

多くのシステムでは導入後に徐々に稼働 VM が増加し、システム全体のリソース利用率も CPU・メモリ・ストレージそれぞれで増加する

② 性能強化にスケールアウトが適切な例

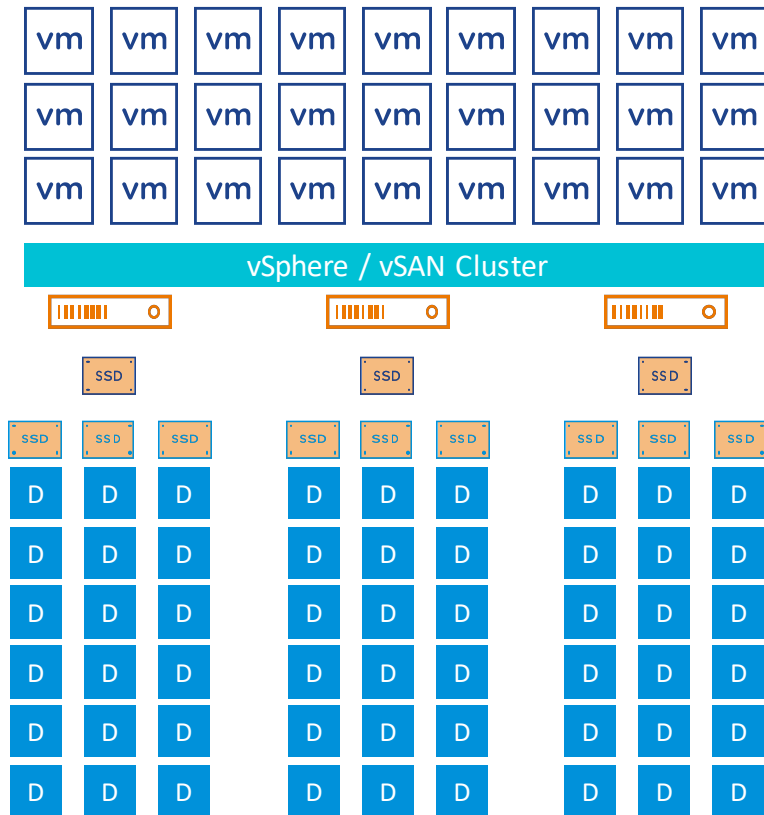


リソース追加に伴い、CPU・メモリ・ディスクのそれぞれのリソースが逼迫した場合は性能低下につながるリスクがある

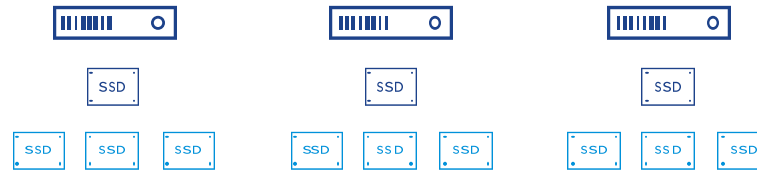
稼働 VM が当初想定より大幅に増えすぎた場合は ESXi ホストのリソースが逼迫し、稼働 VM の性能低下やそれ以上の VM の起動が不可になる場合も

これは健全な状態とは言えないので ESXi ホストの増設が必要となります

② 性能強化にスケールアウトが適切な例



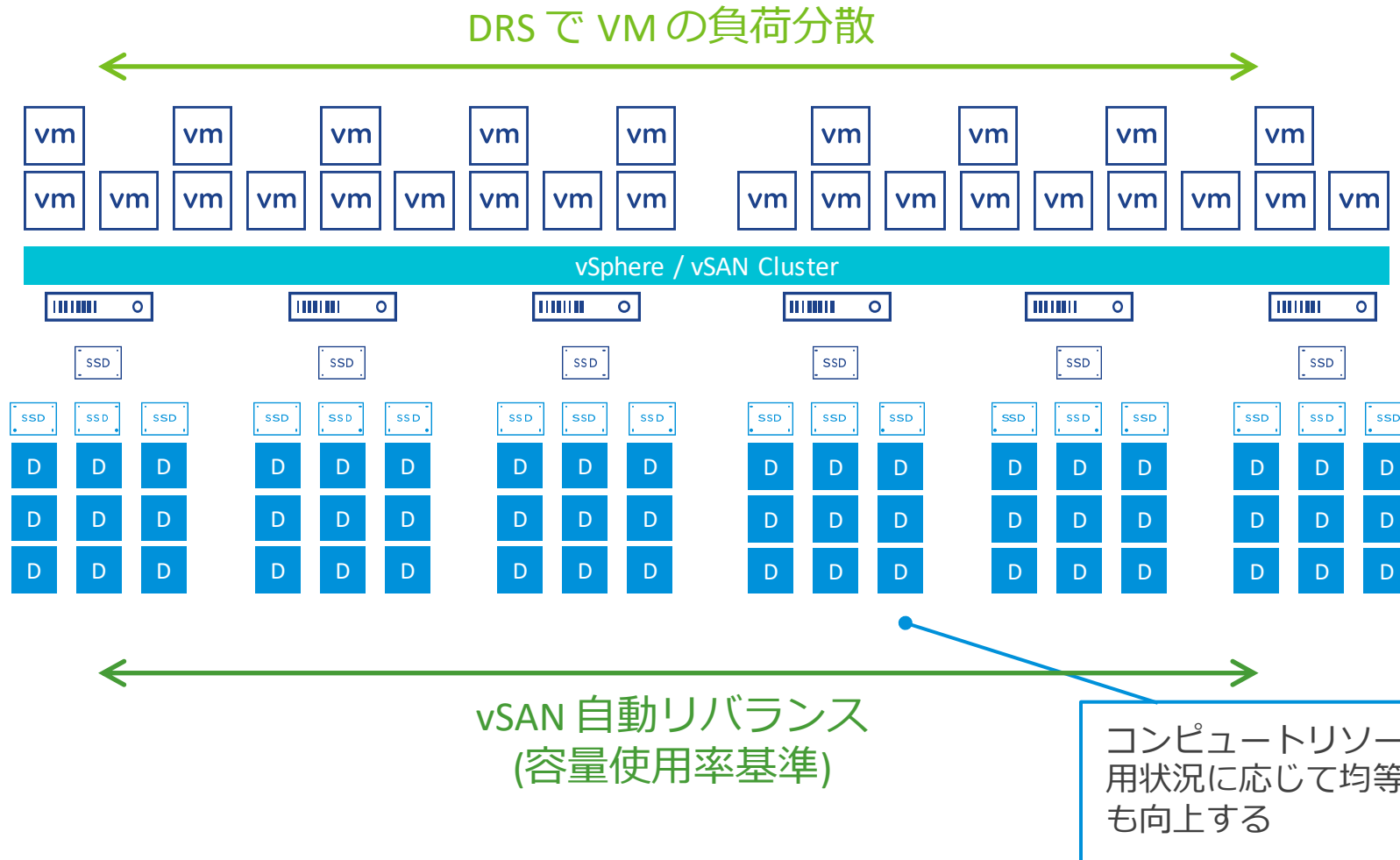
リソース需要に合わせてESXiホストを追加、クラスタをスケールアウトする事で、vSANデータストアも自動で拡張、分散される



vCenter vSphere Clientの「クラスタクイックスタート」機能を利用してESXiホストの一括増設

vDSの設定、vSANディスクグループの組み込みなどの一連の作業をウィザード形式のGUIでシンプルに操作可能

② 性能強化にスケールアウトが適切な例



ESXi ホスト増設後は vSphere DRS (CPU・メモリ利用率基準) による稼働 VM の自動負荷分散と、vSAN 自動リバランス (容量利用率閾値基準) による負荷分散・データ再配置を実施

手作業を極力排除し、自動化でインフラ運用を最適化

vSAN HCI に向く用途、向かない用途

vSAN の性能特性を考えてワークロードを適材適所で配置する事が重要

HCI に向く用途

インフラ全体で比較的均等なリソース要求

- vSAN クラスタ全体に VM とデータを分散し均等に IO を捌く事で vSAN の性能は最大限発揮
- ストレージ容量、IO と CPU、メモリの要求をクラスタ全体でバランスさせて運用

低遅延・高 IO を全体的に要求する環境

- IO は In-Kernel 処理のためオーバーヘッドが最小で処理
- 多数の vSAN 認証取得済みのドライブから性能要件に適した構成で導入可能

HCI に向かない用途

極端なアンバランスワークロード

- 特定 VM のみが大容量・超高 IO を要求する環境
- 1 VM の容量が 1 ESXi に搭載されるディスクグループ容量を大幅に超える大容量
- 殆ど IO を必要としないデータアーカイブ用途

定期的なインフラのバージョンアップが NGな環境

- 健全なライフサイクル運用、既知の不具合リスク回避のために定期的なバージョンアップを強く推奨
 - VM Service を停止せず ESXi ホストのローリングアップデートを vLCM で実施するのでバージョンアップ時の VM 停止は通常は不要

vSAN Performance

HCIBench を利用した vSAN IO 性能ベンチマーク

標準的な AF vSAN 4ノード構成での性能評価
vSAN HCI の最新性能情報の一例を紹介します

※ vSAN 6.7u3 → 7.0u2 & 7.0U3 での遅延と IOPS 傾向

どんなストレージ性能検証ツールを使っていますか？



IOMeter ? FIO ? VDbench ? CrystalDiskMark ?

IOMeter

<http://www.iometer.org/>

1998年リリースの老舗のIO 負荷計測ツール

fio

<https://github.com/axboe/fio/>

ベンチマークの他ストレステストやハードウェアチェックにも使われるツール

VDbench

<http://www.oracle.com/technetwork/server-storage/vdbench-downloads-1901681.html>

高機能なIO ベンチマークツールで詳細オプションで様々な負荷をシミュレート

CrystalDiskMark

<https://crystalmark.info/software/CrystalDiskMark/index-e.html#Summary>

主に単体 Windows でお気軽にIO 性能測定可能なツール

HCIBench

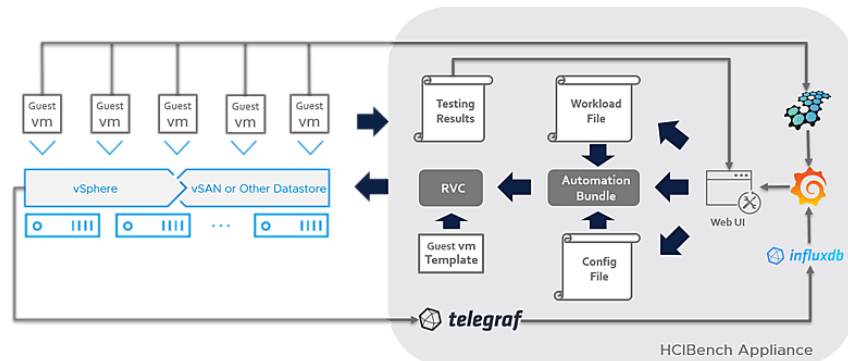
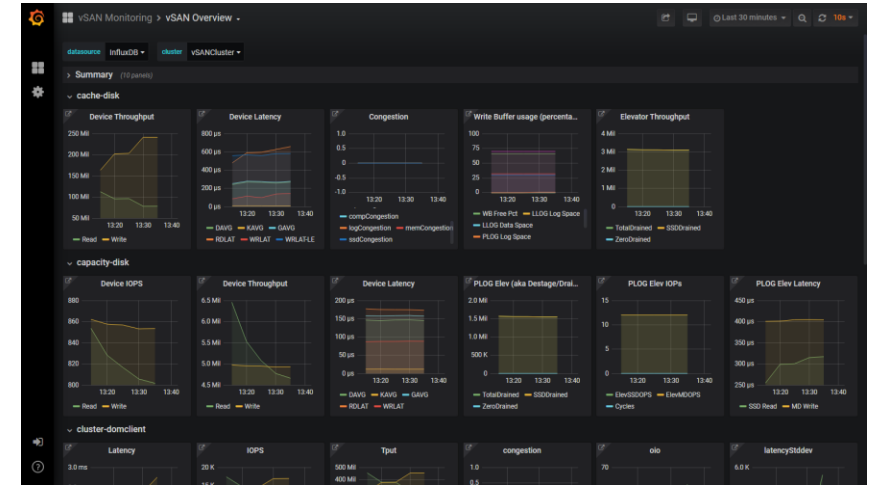
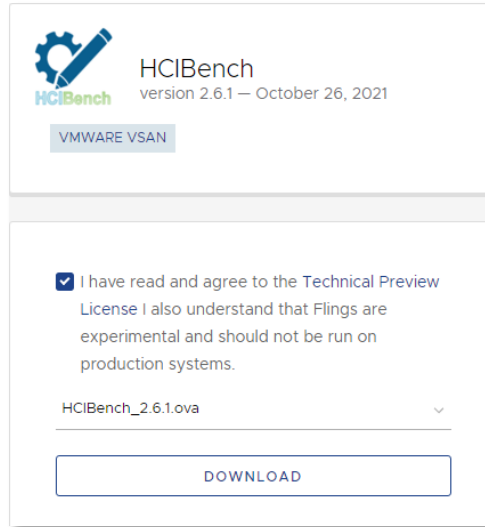
<https://flings.vmware.com/hcibench>

vSphere 基盤のストレージ向けに特化したツール
fio・VDbench をワーカーとして自動レポート化

HCIBench とは？

VMware Flings で公開されている無償のストレージ負荷テストツール

<https://flings.vmware.com/hcibench>



HCIBench は VMware Flings で無償公開されている vSphere 環境で Vdbench や FIO を用いたストレージ負荷テストのジョブ作成・実施・レポート化するツールです。

vSAN など HCI に限らず、従来の外部ストレージを利用したデータストアでも同条件で検証可能なのでぜひお試しください。

※ 使い方については次頁参照

HCIBench を利用した詳細検証の実施方法

IO のトップスピードよりそこに至るまでのIO 遅延から傾向を把握する事が重要

調べたこと 試したこと

サーバー、ストレージ、仮想化関連を中心に調べたこと、試したことを書き留めています。

VMworld 2021 Japan 登録受付中

ゼネラルセッションに CEO ラグー・ラグラム登壇。100以上のブレイクアウトセッションを配信 ヴィエムウェア

2019年6月16日 日曜日

HCIBench を利用したベンチマークテストのポイントと効率化 (2)

前回、前々回で紹介した HCIBench 2.0 の紹介に続いて、HCIBench を利用した VDBench の各種パラメータのカスタム、レポートの自動生成について紹介します。

前回までの記事は以下参照ください。

- HCIBench 2.0 のリリースと機能強化のご紹介
- HCIBench を利用したベンチマークテストのポイントと効率化 (1)

※ HCIBench の公式情報や質問は本家のサイトにコメントすると開発者の方が即返信してくれます。
<https://labs.vmware.com/flings/hcibench>

なお、本投稿内のグラフ、IOPS 性能値などは特定の機器について述べたものではなく、サンプルとして数値に手入れをしたものを利用しています。

|| HCIBench で生成する パラメータファイル の保存先

Select a Workload Parameter File
fio-10vmdk-100ws-4k-70rdpct-
ADD REFRESH DELETE

Upload a Parameter File
ファイルを選択 選択されていません
UPLOAD

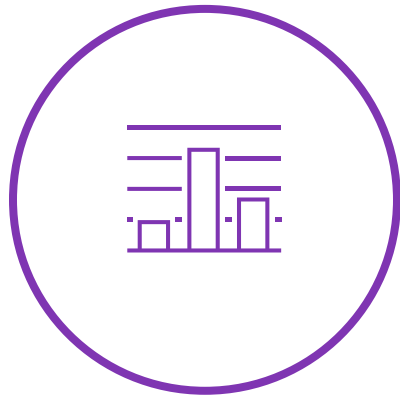
2019年に検証方法を整理して公開していますので、「ストレージ検証を実施したいがどうするのがベストか分からない」といったときの参考にさせていただけると幸いです

- HCIBench を利用したベンチマークテストのポイントと効率化 (1)
<https://kwmtlog.blogspot.com/2019/05/hcibench-performance-test-tips1.html>
- HCIBench を利用したベンチマークテストのポイントと効率化 (2)
<https://kwmtlog.blogspot.com/2019/06/hcibench-performance-test-tips2.html>

ストレージ性能検証のポイント

高IO 負荷を掛ければ良いというのではなく、正しい想定負荷を安定して処理するかを検証

最大瞬間性能の計測はあまり意味がない。性能検証で何のためのデータを取りたいのかを明確に！



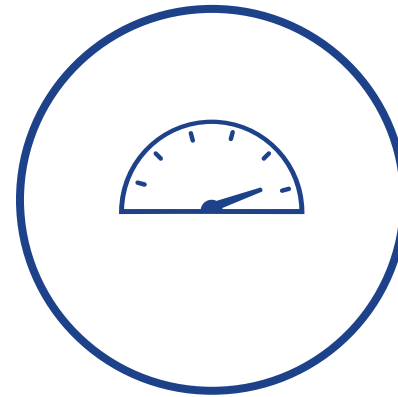
IOPS ・ Throughput

ストレージの性能の重要な指標だが他の指標・条件との照らし合わせ比較が必須



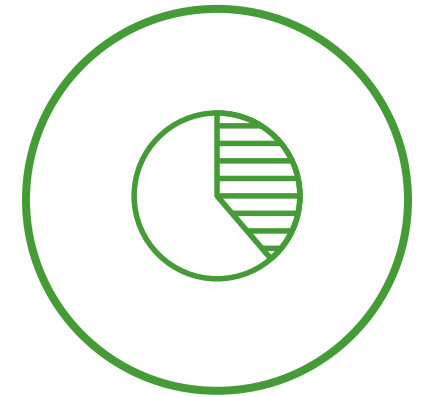
Latency

低遅延で処理が捌けるほど優秀。負荷パターン※1にもよるが、1ms ~ 2ms 以内の応答時間が理想



System Usage

性能が良くてもコントロールの限界ギリギリの負荷では常用に耐えられないので、必ず負荷の余裕を確認する



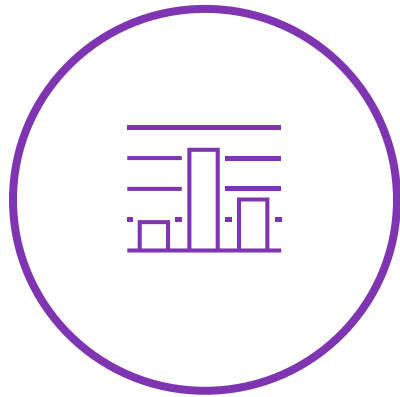
Assumption

検証結果から想定する利用率、ワークロードのパターンを定義し、サイジング設計に活用

ストレージ性能検証のポイント

高IO 負荷を掛ければ良しというのではなく、正しい想定負荷を安定して処理するかを検証

最大瞬間性能の計測はあまり意味がない。性能検証で何のためのデータを取りたいのかを明確に！



IOPS ・ Throughput

ストレージの性能の重要な指標だが他の指標・条件との照らし合わせ比較が必須

多くの製品紹介や性能アピール資料、構成サイジングツールでは IOPS や Throughput がアピールされ、本質的なストレージ性能の誤解が生じ易い

低遅延で処理が捌けるほど優秀。負荷パターン※1 にもよるが、1ms ~ 2ms 以内の応答時間が理想

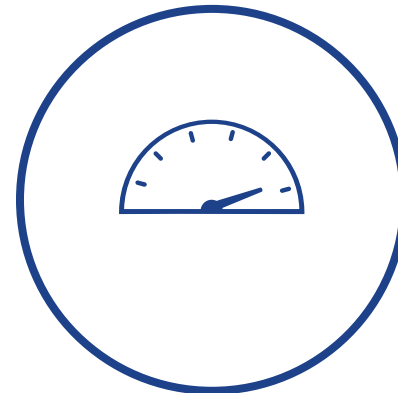
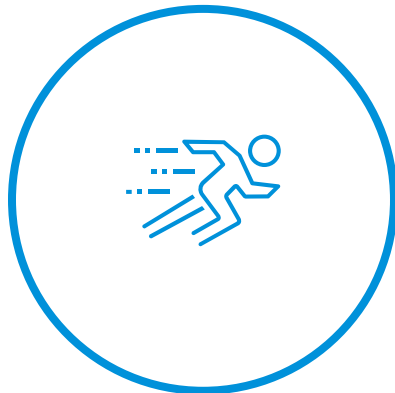
性能が良くてもコントロールの限界ギリギリの負荷では常用に耐えられないので、必ず負荷の余裕を確認する

検証結果から想定する利用率、ワークロードのパターンを定義し、サイジング設計に活用

ストレージ性能検証のポイント

高IO 負荷を掛ければ良しというのではなく、正しい想定負荷を安定して処理するかを検証

最大瞬間性能の計測はあまり意味がない。性能検証で何のためのデータを取りたいのかを明確に！



特定の負荷を余裕をもって提供できる事が重要

IO 負荷が健全な状況で捌けているか？

の重
指
合

わせ比較が必須

Latency

低遅延で処理が捌けるほど優秀。負荷パターン※1にもよるが、1ms ~ 2ms 以内の応答時間が理想

System Usage

性能が良くてもコントロールの限界ギリギリの負荷では常用に耐えられないので、必ず負荷の余裕を確認する

Assumption

検証結果から想定する利用率、ワークロードのパターンを定義し、サイジング設計に活用

ストレージ性能検証のポイント

高IO 負荷を掛ければよしというのではなく、正しい想定負荷を安定して処理するかを検証

最大瞬間性能の計測はあまり意味がない。性能検証で何のためのデータを取りたいのかを明確に！



**ストレージの性能特性と顧客環境のIO 特性の双方を把握して適切なサイジングを行う事が重要
特にアセスメントは必須！！**

IOPS ・ Throughput

ストレージの性能を測る重要な指標だが、顧客環境との照らし合わせ比較が必須

~2ms 以内の応答時間が理想

Latency

これもコントロールが難しいので、必ず負荷の余裕を確認する

Assumption

検証結果から想定する利用率、ワークロードのパターンを定義し、サイジング設計に活用

vSAN 6.7 vs vSAN 7.0 Performance Update

最新版でさらに高性能となった
vSAN 性能検証の結果報告

ベンチマーク試験環境 概要

今回はかなりスモールスタート構成(1CPU・1DG・4Node)で実施



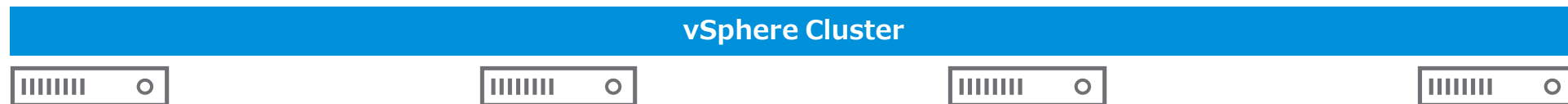
vCenter : 7.0 U2, 17958471



HCI Bench : 2.5.3

VD Bench : 5.04.07

vSphere Cluster



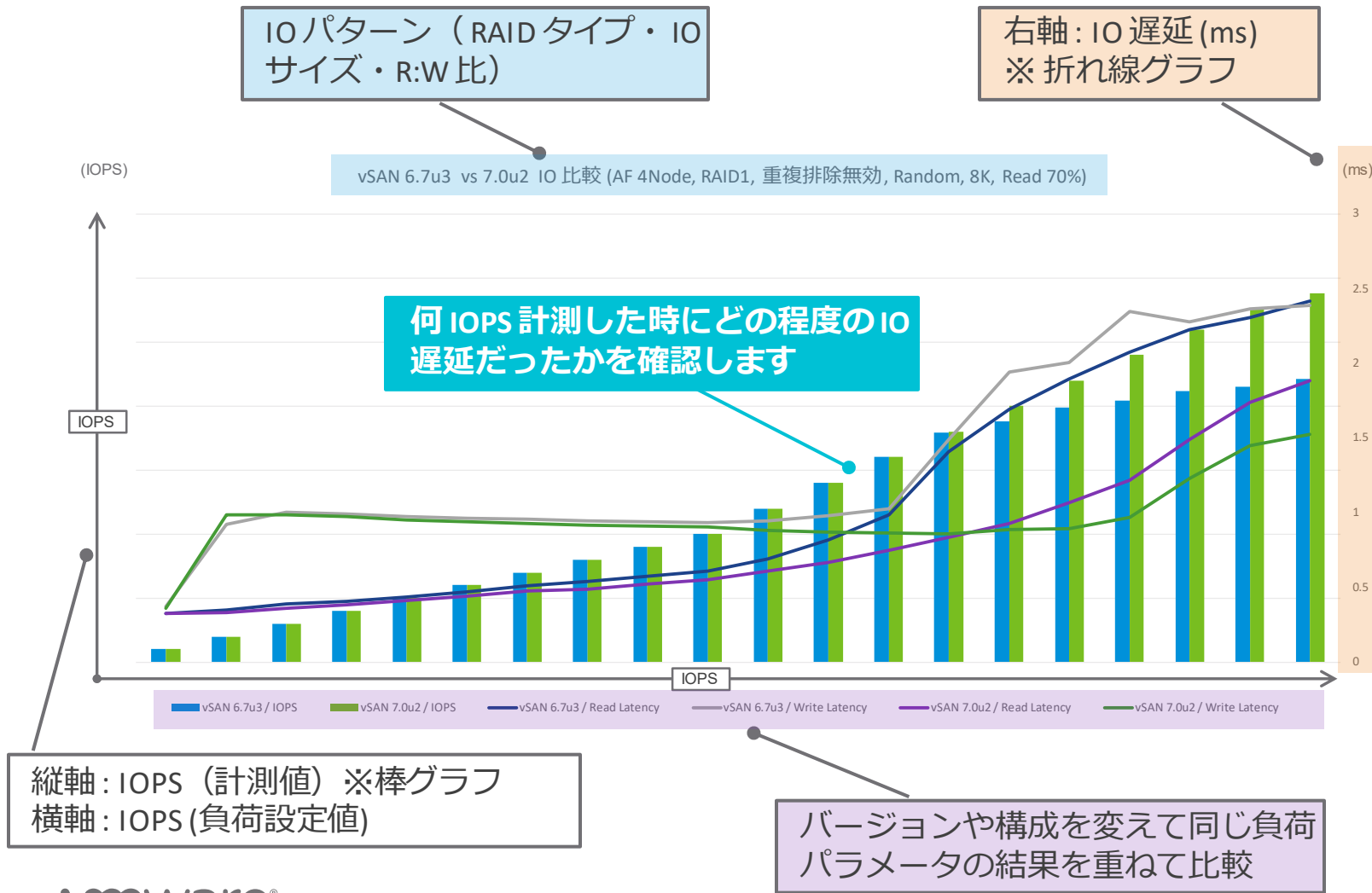
【利用ハードウェア】 : DELL PowerEdge R640 * 4 Node

- CPU Xeon 6138 : 2.0 GHz 20 core * 1 socket
- Mem 96GB (16GB * 6)
- 2port 10G BASE-T
- Boot領域 M.2 PCIe接続 SSD (DellBOSS)
- キャッシュSSD : PX05SMB040Y(10 DWPD) 400GB * 1本
- キャパシティSSD : MZILS1T9HEJH0D3 (1 DWPD)1.92TB * 2本

【重要】 負荷試験時は実際の利用量を想定した領域に均等に負荷を掛ける事 (実効容量の 60% ~ 80% が一般的)

ストレージ負荷試験のシナリオ・グラフの解説

ストレージ負荷を徐々に増加させた際のIO遅延の推移を計測、安定（低遅延）域を確認



次頁以降で紹介する vSAN データストアベンチマークテスト結果の読み方

棒グラフ (IOPS)

- 左から徐々にIO負荷を増加させながらパターン試験を実施
その際に計測されるIO遅延を計測

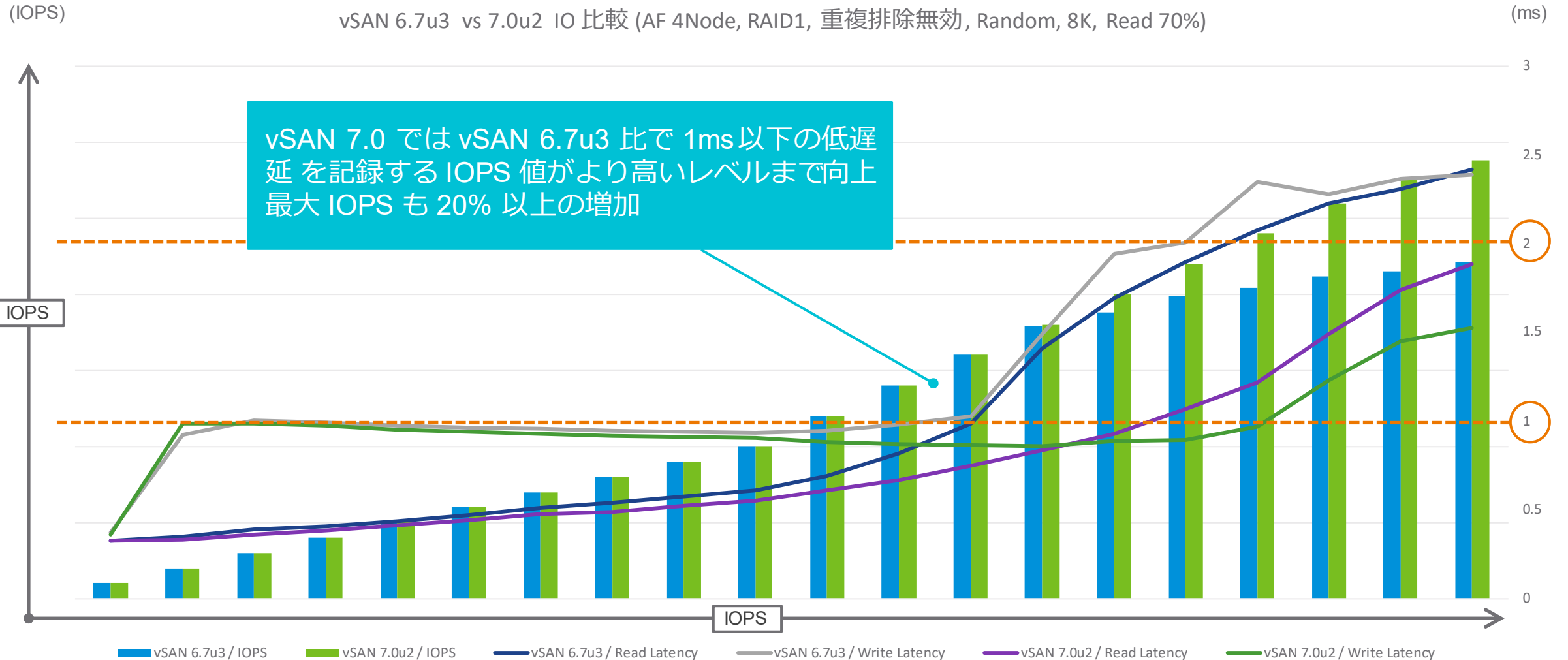
折れ線グラフ (IO遅延)

- 指定のIOPSをベンチマークツールで負荷掛けした際のIO遅延遅延が増加する(過負荷になる) IOPSを見極める

※ VMwareの性能情報公開ポリシーの関係で詳細IOPS値はマスクしての公開となります。

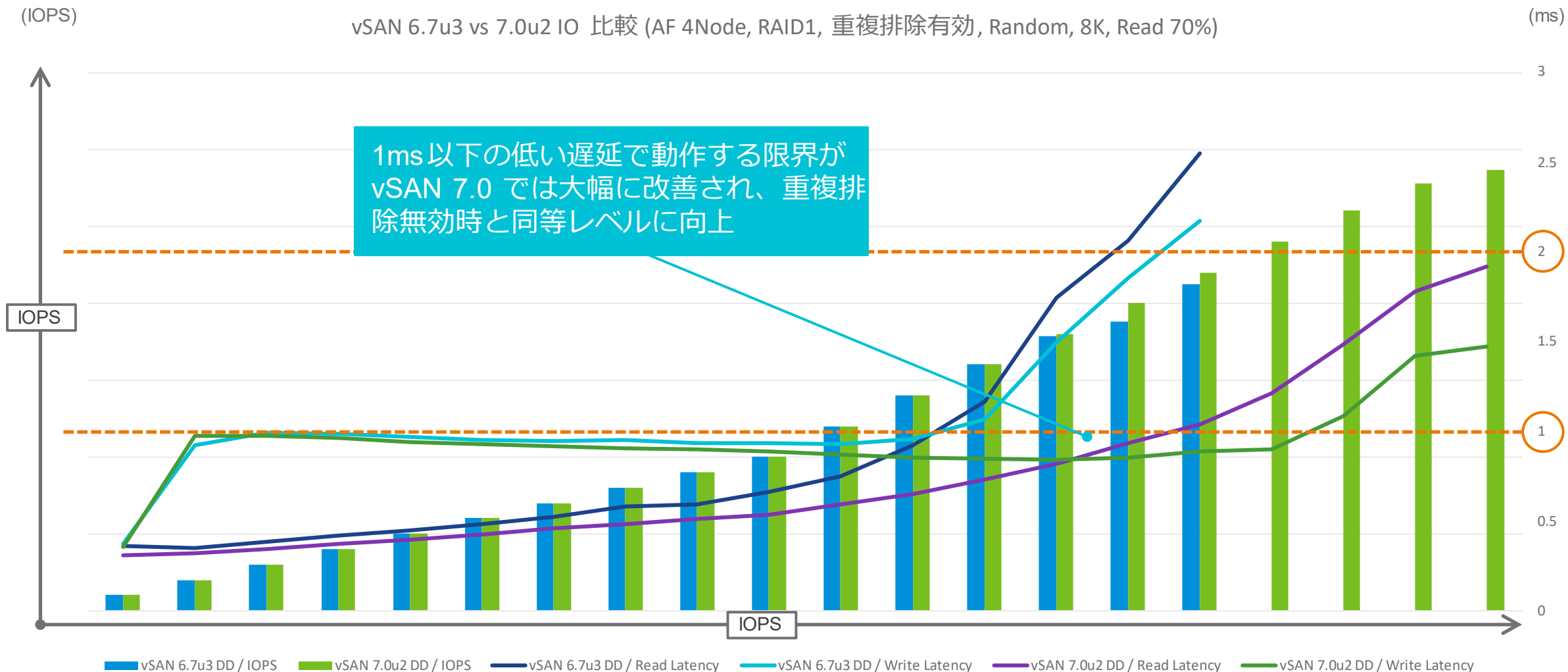
vSAN 7.0 ではアルゴリズム改善による大幅な性能向上を実現

同一ハードウェア利用時における vSAN 6.7u3 と vSAN 7.0u2 性能比較 (重複排除無効時)



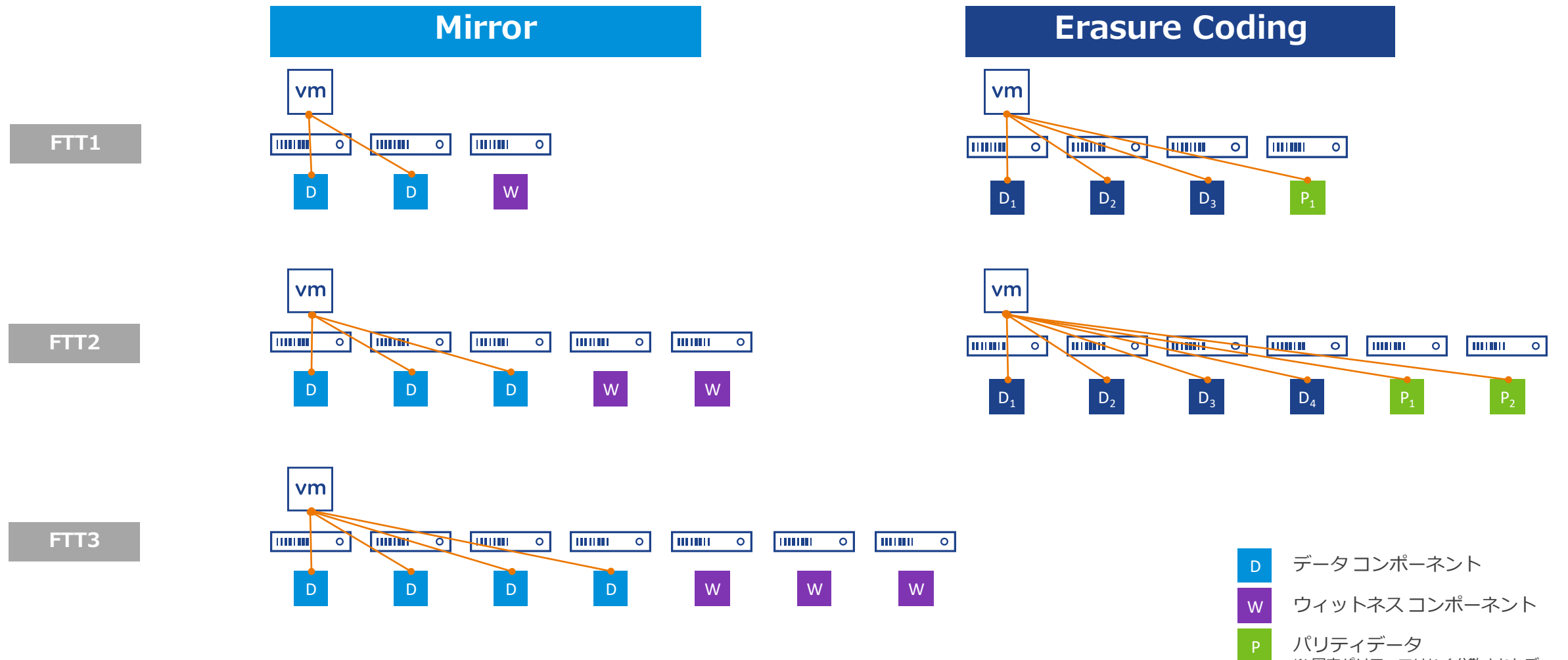
vSAN 7.0 では特に重複排除有効時の性能向上効果が顕著

同一ハードウェア利用時における vSAN 6.7u3 と vSAN 7.0u2 性能比較 (重複排除有効時)



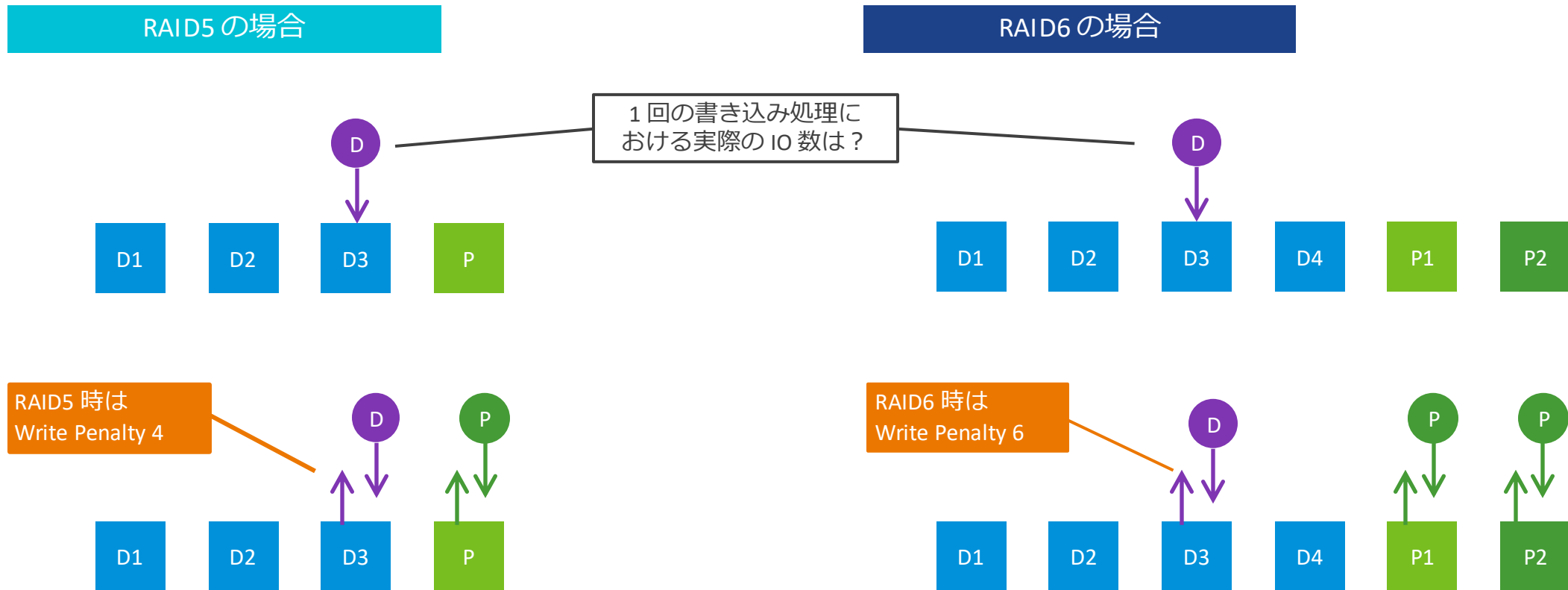
vSAN 保護レベルとデータ配置の違いによる性能差

ホスト障害・デバイス障害からの回避のため複数のESXiホストにデータを冗長配置



vSAN 保護レベルとデータ配置の違いによる性能差

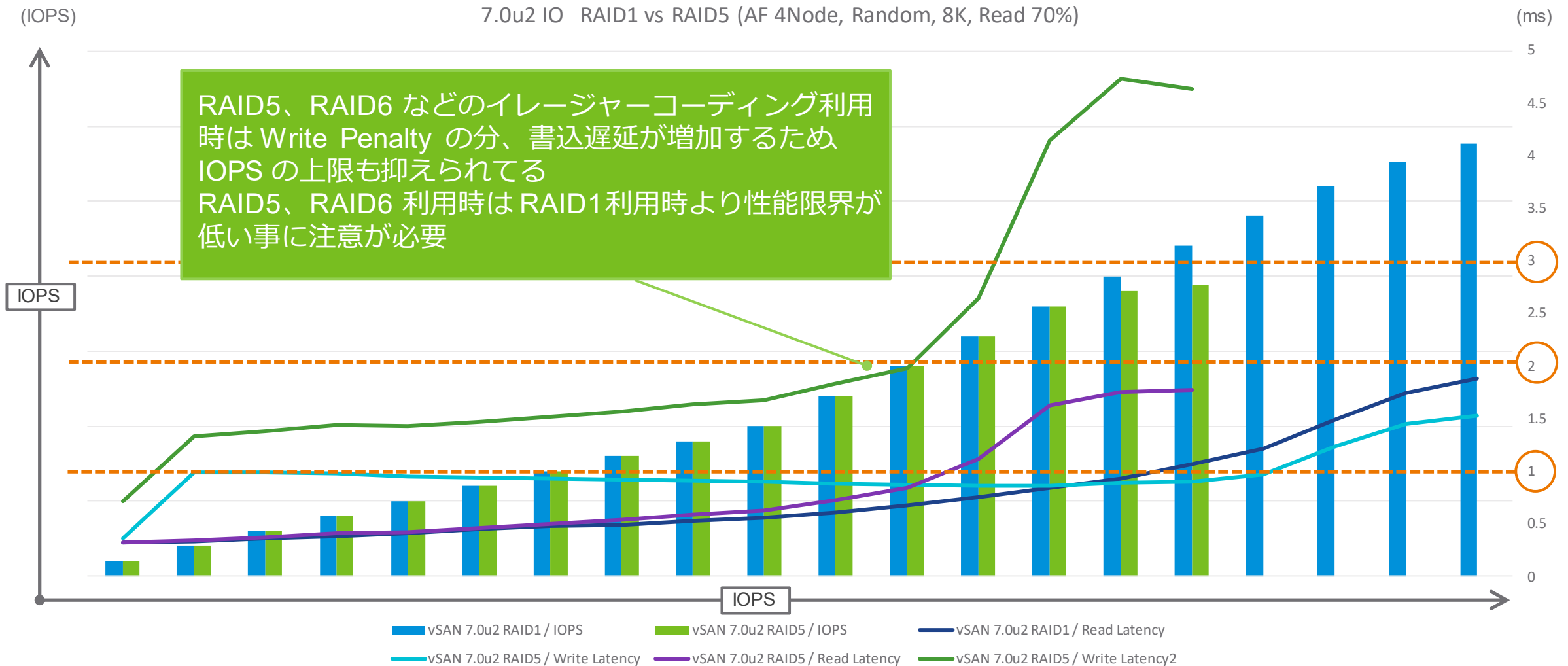
RAID5 / RAID6 構成時の Write Penalty について



※通常のストレージと同様に Erasure Coding 利用時はパリティ計算分の IO が増加
※ IO 性能要件、容量要件に合わせて VM 毎にストレージポリシーを選択して適材適所のデータ配置で運用することが vSAN のベストプラクティス

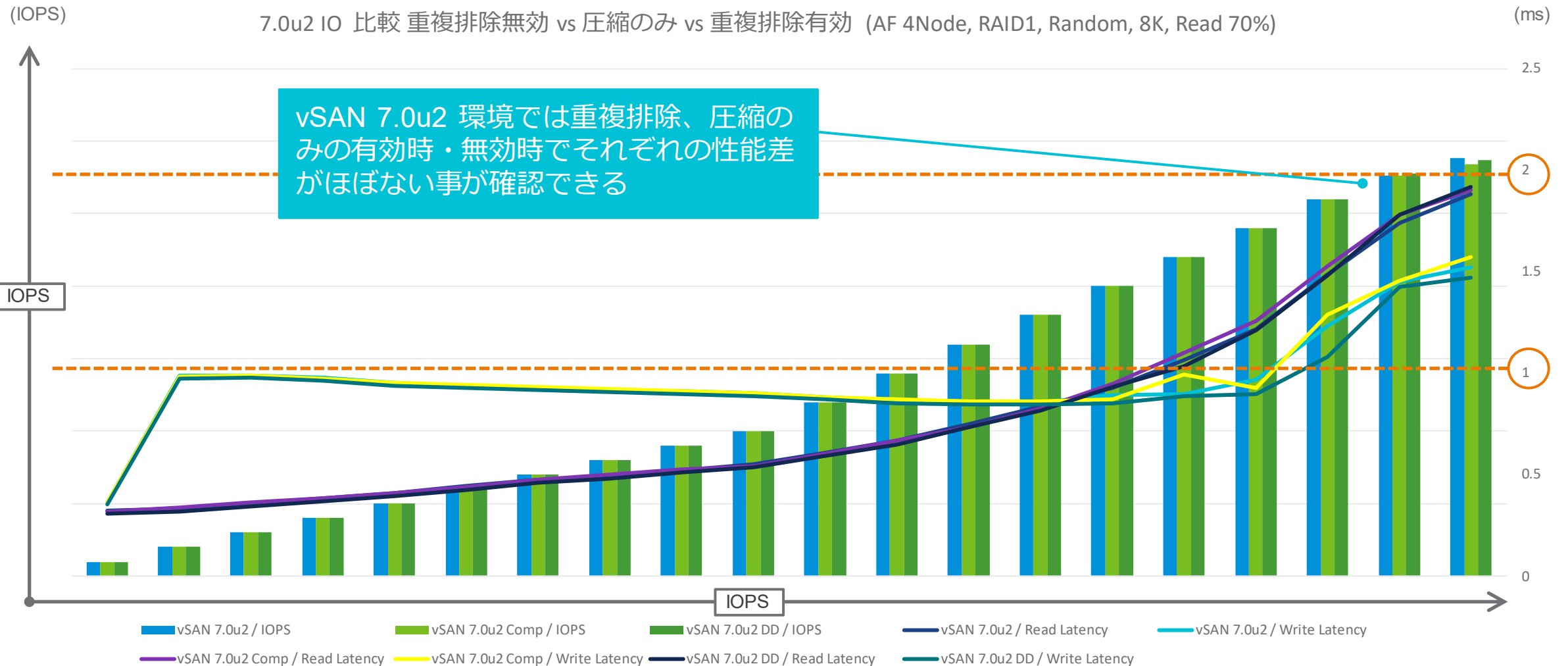
RAID1 (ミラー) vs RAID (イレージャーコーディング) の性能比較

容量効率と性能はトレードオフだが、限界性能を求めない環境ではRAID5/6は有効な選択肢



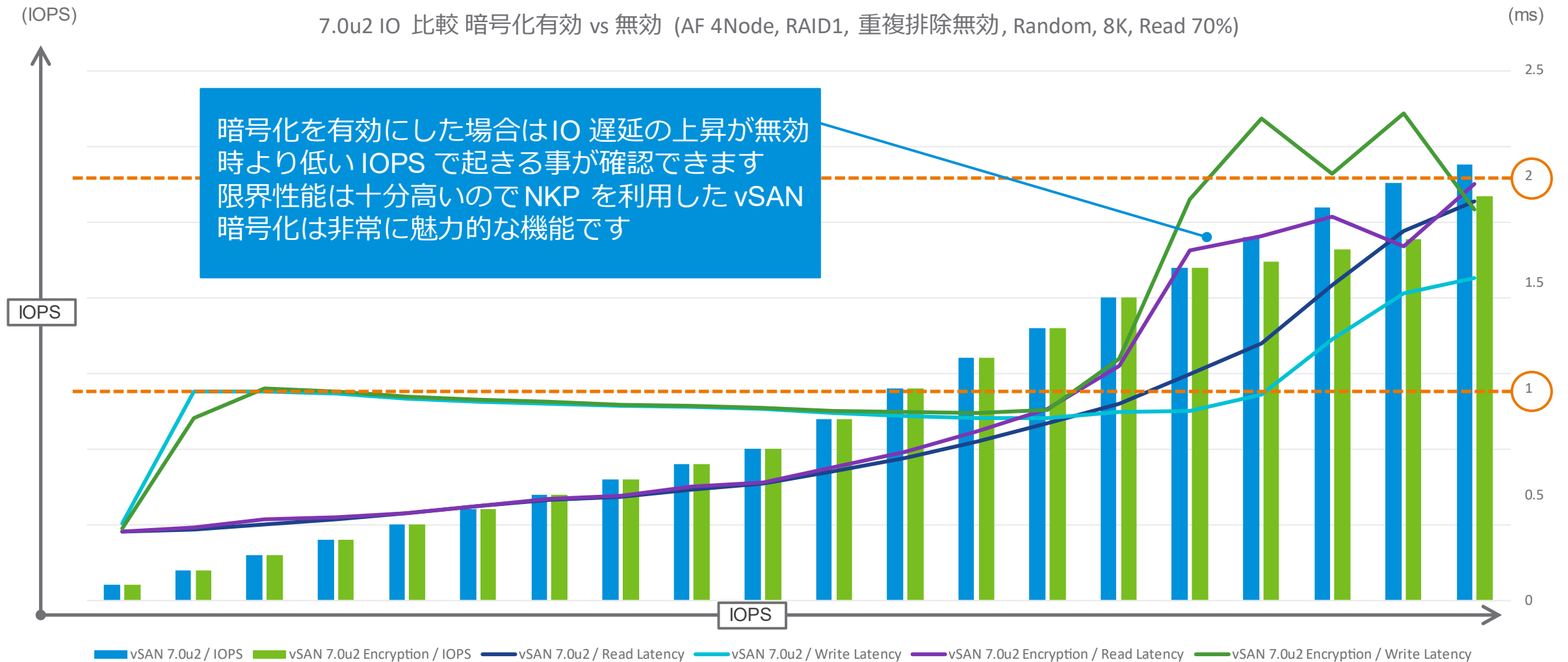
性能劣化のトレードオフ無しで利用可能な容量削減機能

お勧めは vSAN 7.0u1 から実装された「圧縮のみ」機能の利用



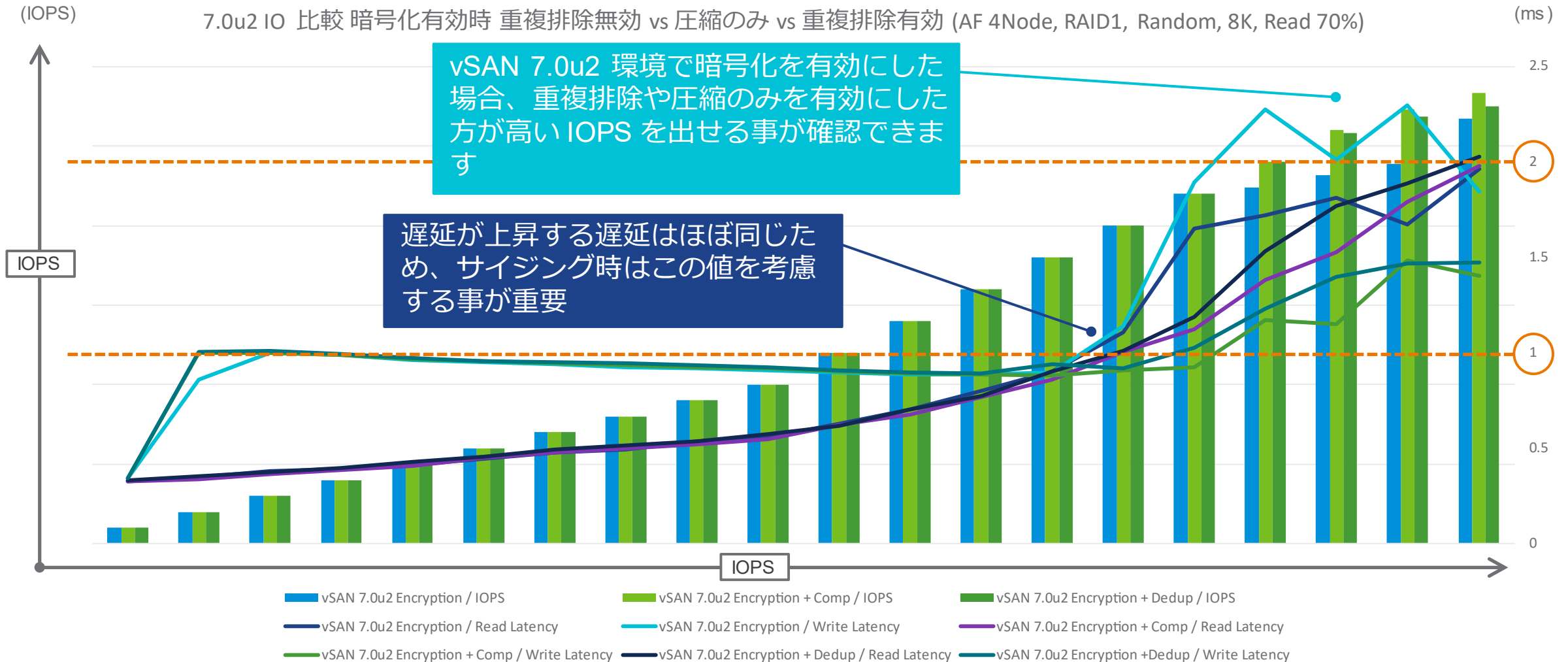
vSAN7.0u2 での vSphere NKP を利用した暗号化と性能の関係

外部 KMS を利用しない vSphere NKP で十分な性能を確認









暗号化有効時は「重複排除」や「圧縮のみ」を利用したほうが高性能

暗号化処理を実行するデータが削減される事が影響していると考えられる



実機性能検証から見えてきたポイント

1. 新しいvSANバージョンでは機能だけでなく性能も大幅にUP   
2. 重複排除、暗号化などの機能も利用しやすく、大幅に安定化 
3. vSAN は物理ドライブの性能をかなり限界まで引き出せるので、汎用的なドライブの組み合わせで外部ストレージ構成よりはるかに低遅延なIOを提供可能  
4. 2CPU 構成、複数ディスクグループ構成で Max 性能を引き出す場合は 25GbE + JumboFrame 利用が推奨
※ 特に高IO 高スループット環境では高い帯域のネットワークを強く推奨
→ ネットワークの推奨デザインは次回詳細解説します
5. 次回セッションではさらに高性能、All NVMe vSAN を利用したらどのくらいのIO性能が得られるのか実機検証を元に詳細解説します

【徹底攻略塾】 VMware Wednesday eXtra : vSAN Deep Dive Series

https://japancatalog.dell.com/c/isg_seminar_vmware_webinar/

2022年 3月 30日 (水) 16:00 - 17:00

ソフトウェアの進化でパフォーマンスも大幅進化！？
VMware vSAN Performance Deep Dive

2022年 5月 25日 (水) 16:00 - 17:00 (予定)  次回ご期待ください

「支える」アーキテクチャを理解してデザインを考える！
VMware vSAN Architecture Deep Dive

2022年 7月 27日 (水) 16:00 - 17:00 (予定)

vSAN なら運用もこんなにシンプルに！ GUI & CLI vSAN 運用詳細解説
VMware vSAN Management Best Practice & What's Next ?



Thank You