



Oracleテクニカル・ホワイト・ペーパー
2014年7月

10,000台以上のVMware仮想マシンの Oracle ZFS Storage Applianceによるデプロイ

ORACLE

はじめに	2
システム・コンポーネントの概要	3
VMwareのブート・ストームおよび動的キャッシュの重要性.....	5
10,000台以上のVMware仮想マシンの同時ブート.....	6
仮想マシンの大規模デプロイメントを実行する場合のストレージに関する考慮事項	8
Oracle ZFS Storage Applianceのディスク・レイアウト.....	8
ネットワーク設定	11
NFS、プロジェクト、およびシェア	13
VMwareのNFSプロトコルに関する推奨事項.....	15
仮想マシンの設定に関する考慮事項	15
結論	16
付録A：ベンチマーク結果	17
SPC-2の結果	17
付録B：参考資料	17

はじめに

仮想マシンが複数あるインフラストラクチャをホスティングするストレージは、通常、次に示す4つの主要なロードに対応できる必要があります。

- プロビジョニング、クローン作成、スナップショットおよび仮想マシンのライブ・マイグレーションの各操作 – プロビジョニング（OSイメージを供給する処理）は一般的に順次処理であるため、I/Oレイテンシの影響をほかの処理よりも強く受けます。イメージ・コピーを使用するプロビジョニングは、データ重複排除などのテクノロジーと組み合わせたとしても、処理が重くなります。対照的に、スナップショット/クローン作成を使用するプロビジョニングは軽量であるため、この方法を実行できる場合は大いに推奨される処理方法です。
- ダンプ・デバイス – ダンプ・デバイスは通常クラッシュ時にのみ使用されるため、ここでのアクティビティはあまりありません。多数の仮想マシンが同時にクラッシュすることはまれですが、ありえないことではありません。たとえば、Linuxゲストの構成でkernel.hung_task_panicが設定されていると、共通する1つのイベントが原因で多数のLinuxゲストが同時にパニックを起こす可能性があります。言うまでもなく、通常はこのパラメータを設定すべきではありません。なぜなら、クラッシュ・ダンプが発生すると、大半の仮想マシンの操作を上回る集中的なワークロードが発生するからです。
- 仮想マシンのワークロードまたはフットプリント – ワークロードは、各仮想マシン内で実行しているアプリケーションによって生成されます。アプリケーションのI/Oパターンは、オンライン・トランザクション処理（OLTP）、メール・サーバー・オペレーション、Webサーバー・ワークロードなどさまざまで、多岐にわたる可能性があります。
- ブート・プロセス – これが、本書で説明する中心的な項目です。ブート・プロセスでは、プロトコルがiSCSIなのか、ファイバ・チャネル（FC）なのか、ネットワーク・ファイル・システム（NFS）なのかに関係なく、同じ量のワークロードが発生するものです。理由についてはこれから説明しますが、検証の結果、VMwareのワークロードにはNFSが最適な選択肢であることが分かりました。

こうしたさまざまなストレージI/Oワークロードからは、毎秒何千ものI/O処理が発生する可能性があります。これに対応するには、応答時間およびレイテンシが短く、パフォーマンスがよいインテリジェントなストレージ・アーキテクチャが必要です。

このホワイト・ペーパーでは、Oracle ZFS Storage Applianceを使用した大規模なVMware仮想マシンのデプロイメントで最適なI/Oパフォーマンスとスループットを得るためのストレージに関する考慮事項について説明します。

ストレージに関する考慮事項と推奨事項で説明するおもな内容は、VMwareのNFSプロトコルの構成オプションおよびチューニング・オプション、推奨されるディスク・レイアウト、Oracle ZFS Storage Applianceと連携するVMware vSphere 5.x環境の正しいIPネットワーク・インフラストラクチャ設計です。

このホワイト・ペーパーの要点は次のとおりです。

- VMwareのブート・ストームの概要と考慮事項
- NFSプロトコルとOracle ZFS Storage Applianceを使用して10,000台を超える大量のVMware仮想マシンをブートする場合のストレージに関する考慮事項

- NFSプロトコルに対応するVMware 10GbEネットワーク・インフラストラクチャとチューニング・オプション
- 10,000台を超えるVMware仮想マシンに対応するOracle ZFS Storage Applianceのディスク・プール・レイアウト
- Oracle ZFS Storage Applianceを使用したVMware仮想マシンの同時ブートとパフォーマンス結果
- 何千台ものVMware仮想マシンをOracle ZFS Storage Applianceにデプロイする場合のNFSに関する考慮事項

システム・コンポーネントの概要

以下の表は、本書で使用されているハードウェア構成、オペレーティング・システム、およびソフトウェア・リリースを示しています。

表1に、使用したハードウェアを示します。

表1: リファレンス・アーキテクチャで使用したハードウェア

機器	数量	構成
ストレージ	1クラスター (2コントローラ)	Oracle ZFS Storage ZS3-2クラスター コントローラあたり256ギガバイト (GB) のダイナミック・ランダム・アクセス・メモリ (DRAM) 20.3テラバイト (TB) のディスク・ドライブ×4 - Oracle Storage Drive Enclosure DE2-24C コントローラあたり4枚の10ギガビット・イーサネット (GbE) ネットワーク・インタフェース・カード (NIC) 73GBログ・デバイス×4
IPネットワーク・スイッチ	2	10GbEネットワーク・スイッチ

表2に、使用した仮想マシン・コンポーネントを示します。

表2: リファレンス・アーキテクチャで使用した仮想マシン・コンポーネント

オペレーティング・システム	数量	構成
Oracle Linux 6.2	10,000以上	Linux仮想マシン

表3に、使用したソフトウェアを示します。

表3：リファレンス・アーキテクチャで使用するソフトウェア

ソフトウェア	バージョン
Oracle ZFS Storage Appliance OS	2013.1.2.0
Vdbench	50401
Sun StorageTek Workload Analysis Tool (SWAT)	3.0.2 – MOS Patch 10350687
VMware vCenter Server	5.1u2
VMware ESXハイパーバイザ・ソフトウェア	5.1u2

VMwareのブート・ストームおよび動的キャッシュの重要性

大量の仮想マシンまたは仮想デスクトップが同じ時間枠の中でブートされると、ブート・ストームが発生してパフォーマンスとサービスが低下します。仮想マシン環境および仮想デスクトップ・インフラストラクチャ環境では、最初の起動操作でストレージI/Oワークロードが集中的に発生すると同時に、アプリケーションおよびオペレーティング・システムでも大量のディスク読み込み操作が実行されます。こうした状況では、仮想マシンが生成する異例の予測不能な大量の読み込みI/Oワークロードをストレージ・デバイスで処理できなくなる、という深刻な事態が起きかねません。ストレージI/Oが低下したり仮想デスクトップが遅くなったり応答しなくなったりするだけでなく、ストレージ・デバイスが原因でネットワーク・パフォーマンスが大幅に低下する可能性もあります。

Oracle ZFS Storage Applianceは、極端なI/Oワークロード、特に仮想デスクトップ・インフラストラクチャや仮想マシンが生成する予測不能なI/Oワークロードを、低レイテンシで処理できるように設計されています。この効率性を実現する鍵は、動的キャッシュの管理機能です。ハイブリッド・ストレージ・プールというOracle ZFS Storage Applianceの機能が、メモリ、フラッシュおよび物理ディスクを統合し、データを効率的に配置してパフォーマンスを向上させます。Oracle ZFS Storage Applianceでは、適応型のインテリジェントな一連のI/O管理アルゴリズムを使用し、ダイナミック・ランダム・アクセス・メモリ (DRAM)、読み込み/書き込みに最適化されたフラッシュベースのSSD、SASディスクといったハードウェア・リソースの使用効率を最大化できるようにして、パフォーマンス効率を最適化します。読み込み/書き込みパスは、個々のパフォーマンス要件およびデータ整合性要件に対処できるように、それぞれ異なる方法で処理されます。

DRAMは、読み込みを高速化するためのプライマリ・キャッシュとして使用されます。DRAMはディスクやフラッシュのいずれよりもはるかに高速にトランザクション・ワークロードを処理できるメディア・タイプであるため、DRAMで処理される読み込み操作の割合を高くすることで、システム全体のパフォーマンスが劇的に向上します。読み込みキャッシュとして使用されるDRAMの部分は適応型置換キャッシュ (ARC) と呼ばれます。ARCに割り当てるDRAMの量は、システム全体のパフォーマンスが最大化されるように、オペレーティング・システムによって動的に管理されます。ARCは、もっとも最近使用された (MRU) ブロック、使用頻度のもっとも高い (MFU) ブロック、もっとも過去に使用された (LRU) ブロック、使用頻度のもっとも低い (LFU) ブロックに分けられています。この適応性が高い動的キャッシュのアーキテクチャは次の図で確認できます。

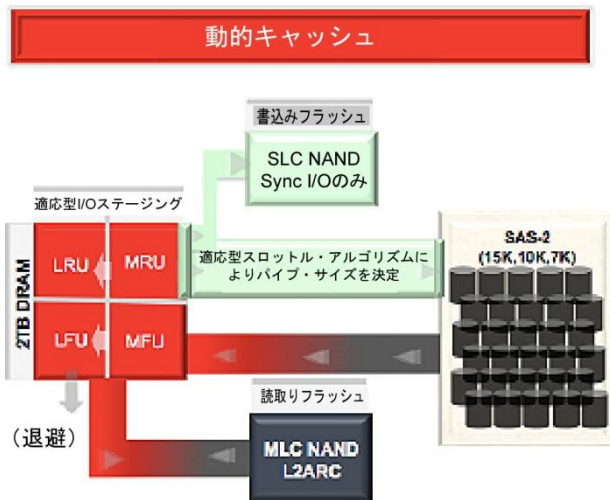


図1: Oracle ZFS Storage Applianceの動的キャッシュ機能

図1で使用されている略語とその意味は次のとおりです。SLC（シングル・レベル・セル）、MLC（マルチ・レベル・セル）、NAND（消去、書き込み、読み込みの各機能を高速化するフラッシュ・テクノロジー）。

データセット全体で“もっともホットな”部分をDRAMに保持する、というのがこの考え方です。ARCが飽和状態になり、ARC内のよりクールなデータとよりホットなデータの交換が必要になると、ハイブリッド・ストレージ・プールはDRAM内のもっともクールなデータを読み込みフラッシュ・キャッシュ・デバイスへ退避させます。これはレベル2 ARC（L2ARC）と呼ばれるもので、Oracle ZFS Storage ApplianceではSSDが使用されています。ARCまたはL2ARCに保持しておくほどホットではないと判断されたデータに対して読み込みリクエストがあった場合は、回転ディスクからデータを取得する必要があるため、こうした読み込み操作ではレイテンシが長くなります。ただし、実際には、インストールしたベース・システムで幅広いサンプリングを実施し、ARCヒット率が80%を超えるようにすることが一般的です。

ブート・ストームの発生時には、ブート・イメージをキャッシュするのに十分なL1 ARCが用意されると同時に、仮想マシン環境または仮想デスクトップ環境から発生するランダムI/Oワークロードが高いパフォーマンスで処理されます。

Oracle ZFS Storage Applianceのほかの重要なコンポーネントは、マルチコアCPUとそのマルチスレッド機能を最大限に活用する対称型マルチプロセッシング（SMP）アーキテクチャに搭載されたマルチコアCPUとオペレーティング・システムです。重要なのは、VMwareやその他の仮想化テクノロジーからは、SMPアーキテクチャに最適なワークロードが発生するということです。

10,000台以上のVMware仮想マシンの同時ブート

インテリジェントなキャッシュ・アーキテクチャを備え、クラスタ構成であることから、Oracle ZFS Storage Appliance ZS3-2で膨大な数の仮想マシンを動作させることができます。Oracle ZFS Storage ZS3-2クラスタ・システムでは、何台の仮想マシンを最適なパフォーマンス・レベルでブートおよびサポートできるのでしょうか。答えは次のグラフに示されています。

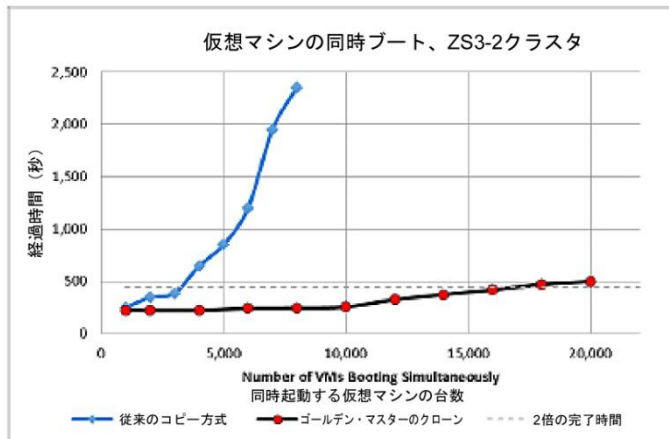


図2：Oracle ZFS Storage Appliance ZS3-2のヘッドあたりのVMware仮想マシンの同時ブート・ディスク数

基本的に、従来のコピー方式を使用する仮想マシンの場合は、1つのOracle ZFS Storage ZS3-2システムでブートできる仮想マシンの最大数は、ヘッドあたり8,000台であり、Oracle ZFS Storage ZS3-2クラスタあたりでは16,000台です。仮想マシンのクローンを使用し、ゴールデン・マスター方式またはVMwareリンク・クローン・テクノロジーを利用する場合は、結果がさらによくなり、Oracle ZFS Storage ZS3-2システム1つでクラスタ内の仮想マシンを20,000台までブートできる可能性があります。

グラフ中の点線は、応答時間が短縮されるおおよその境界（カットオフ）を表しています。従来のコピー方式またはゴールデン・マスター・クローン方式のいずれを使用する場合も、500台の仮想マシンをブートするのにかかる時間はほぼ同じです。カットオフが発生するのは、最小限の移入にかかった時間の2倍の時間がかかり始めたときです。これは、従来の方式の場合はおおよそ3,500台の仮想マシンをブートしたときであり、クローンを使用した場合は16,000台を超える仮想マシンをブートしたときです。20,000台の仮想マシンを同時にブートする場合でさえ、ブート・プロセスの完了までおおよそ8.5分しかかかりません。

この数字は楽観的に思えるかもしれませんが、というのも、この数字はシステムのブートのみを対象にしたもので、ブート・ストームが発生している間はほかのワークロードを処理するための機能がコントローラに残されていないためです。しかしながら、ブートが本当に同時であるなら、すべての仮想マシンが同時にブートされるので、これはきわめて筋が通っていますし、ブートが終了した後はコントローラ上のワークロードは基本的にゼロまで低下します。

重要な点は、300,000 NFS ops/秒の処理能力を持つシステムに対して、20,000台の仮想マシンのブート・イメージを処理しながら、2つのコントローラ・ヘッドのそれぞれで1秒あたり150,000を超えるNFS操作を処理していたことです。これは、次のOracle ZFS StorageのAnalyticsのスクリーンショットを見れば分かります。このパフォーマンスは、報告されているベンチマーク・スコア（210k ops/秒）を大幅に上回っていますが、その理由は、基本的にすべての読み込み操作がL1ARCで対処されたことにあります。

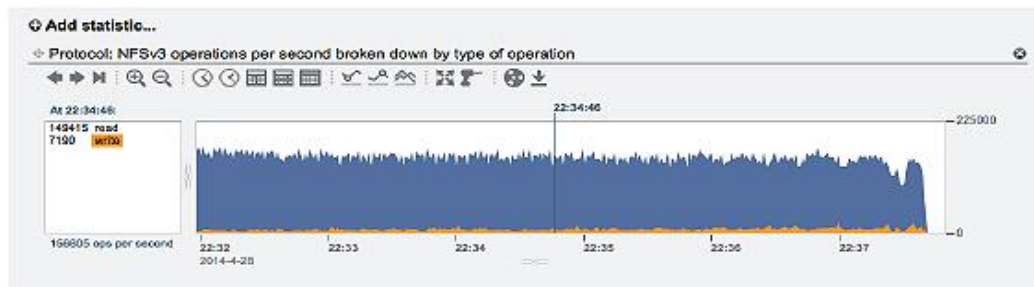


図3：Oracle ZFS Storage Appliance ZS3-2のAnalyticsに操作のタイプ別に分類され表示されたNFSプロトコルのパフォーマンス結果

Oracle ZFS Storage Appliance ZS3-2システムを使用したVMware仮想マシンの大規模なデプロイメントを計画している場合は、前のグラフの青い線に注意してください。こうしたデプロイメントはVMware仮想マシン・サーバーによくあることですが、その場合はサーバーごとに従来の（または完全な）VMware vmdkコピーを使用することになります。

VMware仮想デスクトップの大規模なデプロイメントの場合は、黒線で示した結果のほうが重要です。このデプロイメントの場合は、ゴールデン・イメージからのストレージ・クローン作成やVMwareリンク・クローンなど、異なるクローン作成方式を使用できます。

仮想マシンの大規模デプロイメントを実行する場合のストレージに関する考慮事項

この項では、NFSプロトコルとOracle ZFS Storage Applianceを使用して大量のVMware仮想マシンをブートする場合のストレージに関する考慮事項について説明します。作業環境が、次に示す構成以上になっていることを確認してください。

- Oracle ZFS Storage Appliance、アクティブ/アクティブ・モード
- Oracle ZFS Storage Applianceソフトウェア・リリース2013.1.2.x以上
- コントローラあたり256GB以上のDRAM (L1ARC) キャッシュを搭載したOracle ZFS Storage Applianceコントローラ
- Oracle ZFS Storage Applianceコントローラあたり2基以上のIntel® Xeon® E5-2658 0 2.10 GHz CPU
- コントローラあたり4枚以上のデュアル10GbEネットワーク・インターフェース

Oracle ZFS Storage Applianceのディスク・レイアウト

使用例1は、Oracle ZFS Storage Appliance ZS3-2システムに接続したVMware vSphere5.x環境で大量の仮想マシンをブートする場合に推奨される最小ディスク・ストレージ・レイアウトです。使用するのは小規模なLinux仮想マシン、または（約10GBのシン・プロビジョニングOSブート・ディスクを保持する）小規模な仮想デスクトップで、仮想マシン1台あたりのフットプリントはおよそ3010PSであると想定すると、推奨される最小ディスク構成は次のようになります。

- 44個以上の4TB SAS-2（7200RPMの容量ディスク・ドライブ）で構成された2つのミラー化ディスク・プール
- ストライプ化されたログ・プロファイルを使用するLogZilla用に2つ以上の73GB SSDデバイス
- 4枚の10GbEネットワーク・インタフェース・カード

注：VMwareでサポートされる仮想マシンは、ESXi 5.xホスト1台あたり512台までであるため、サイジングを検討する場合は、使用可能なESXi 5.xサーバーが十分にあり、大量の仮想マシンをホスティングできることを確認してください。このホワイト・ペーパーで紹介する使用例では、メモリと仮想CPUのサイズが小さい構成の仮想マシンを使用しました。VMware仮想マシン1台あたりのメモリは1GBで、仮想CPUは1つです。この構成に準じると、VMware仮想マシンが10,000台の場合は20台以上のVMware ESXi 5.xサーバーが必要になり、仮想マシンが16,000台の場合は32台のVMware ESXiサーバー、仮想マシンが20,000台の場合は40台以上のVMware ESXiサーバーが必要になります。

VMwareのサイジングを検討する場合は、次のURLにある『構成の上限VMware vSphere5.5』ドキュメントを参照してください。

<http://www.vmware.com/files/jp/pdf/vsphere-55-configuration-maximums.pdf>

下の図は、ヘッドあたり44台の4TB SAS-2 7200RPMディスクで構成した例です。DE2-24CはOracle Storage Drive Enclosure DE2-24Cのことで、

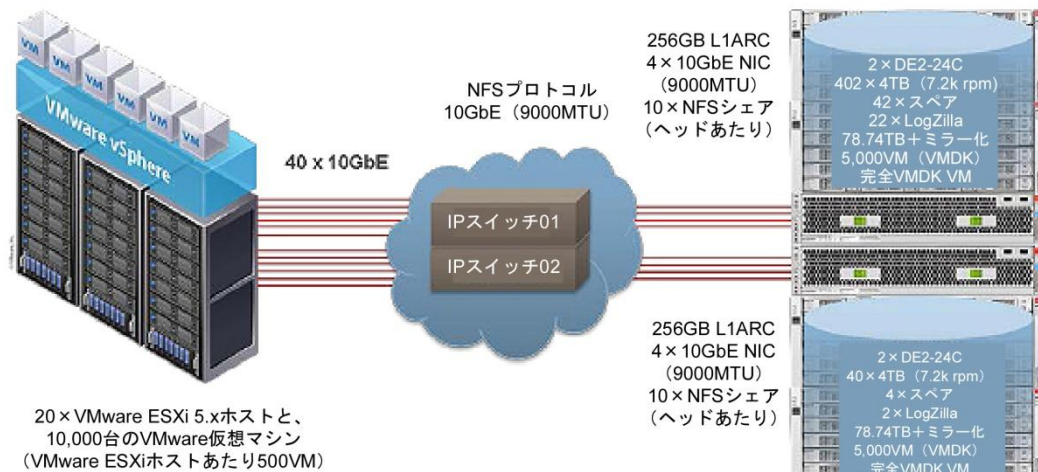


図4：Oracle ZFS Storage Appliance - 10,000台のVMware仮想マシンに容量ディスクを使用した、サイズの小さい仮想マシン/仮想デスクトップ用のディスク・レイアウト

使用例2は、中～大量のVMware仮想マシンまたは仮想デスクトップをブートする場合に推奨される最小ディスク・ストレージ・レイアウトです。このブートでは、“ゴールデン・イメージ”のクローン作成・テクノロジーまたはVMwareリンク・クローン・テクノロジーも使用します。VMware vSphere5.x環境はOracle ZFS Storage ZS3-2クラス構成に接続されています。また、このオプションはパフォーマンス、容量、クローン作成・テクノロジーが組み合わさった別のストレージ・レイアウトでもあります。この使用例も、“ゴールデン・イメージからクローン作成される”約100～300GB（シン・プ

ロビジョニング)のWindows仮想デスクトップを使用することを想定していますが、データ用のVMware仮想ディスクと、ISO、バックアップおよびイメージ・データ・ストア用のディスク領域を追加で備えています。

- ・ ディスク・プールあたり、およびヘッドあたり44台(以上)の900GB SAS-2(10k RPMのパフォーマンス・ディスク・ドライブ)と、ストライプ化されたログ・プロファイルを使用するLogZilla用の73GB SSDデバイス1つ以上で構成された1つのミラー化ディスク・プール
- ・ ディスク・プールあたり、およびヘッドあたり5台(以上)の900GB SAS-2(10k RPMのパフォーマンス・ディスク・ドライブ)と、ストライプ化されたログ・プロファイルを使用するLogZilla用の73GB SSDデバイス2つ以上で構成された1つのミラー化ディスク・プール
- ・ ディスク・プールあたり、およびヘッドあたり44台(以上)の4TB SAS-2(7200RPMのパフォーマンス・ディスク・ドライブ)と、ストライプ化されたログ・プロファイルを使用するLogZilla用の73GB SSDデバイス2つ以上で構成された1つのミラー化ディスク・プール

図5に使用例2を示します。

注：パフォーマンスを最大化するには、900GB 10k RPMのパフォーマンス・ディスクを仮想化環境に使用することが強く推奨されます。また、この点についてはディスク・レイアウトやサイジングのフェーズで検討することをお勧めします。

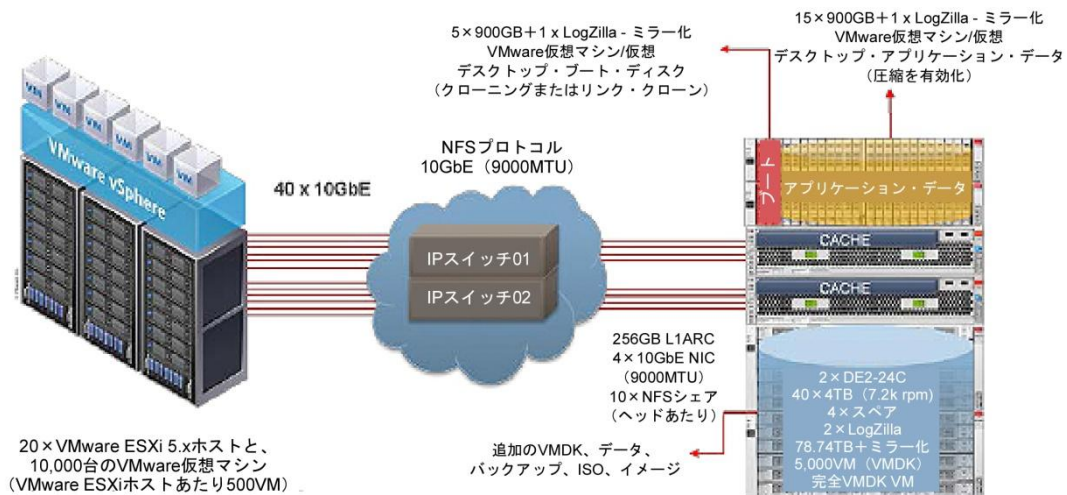


図5：Oracle ZFS Storage Appliance - 10,000台のVMware仮想マシンに容量ディスクを使用した、サイズの小さい仮想マシン/仮想デスクトップ用のディスク・レイアウト

使用例3は、Oracle ZFS Storage Appliance ZS3-2システムに接続したVMware vSphere5.x環境で中～大量の仮想マシンをブートする場合に推奨されるディスク・ストレージ・レイアウトです。中～大量のLinuxまたはWindows仮想マシン/仮想デスクトップ(100～300GB以上のシン・プロビジョニングOSブート・ディスクを保持)を使用し、仮想マシン1台あたりのフットプリントを約50IOPSと想定した場合に推奨される最小構成を次の図に示します。

- ディスク・プールあたり、およびヘッドあたり130台（以上）の4TB SAS-2（7200RPMの容量ディスク・ドライブ）と、ストライプ化されたログ・プロファイルを使用するLogZilla用の73GB SSDデバイス4つ以上で構成された2つのミラー化ディスク・プール
- ディスク・プールあたり、およびヘッドあたり40台（以上）の900GB SAS-2（10k RPMのパフォーマンス・ディスク・ドライブ）と、ストライプ化されたログ・プロファイルを使用するLogZilla用の73GB SSDデバイス4つ以上で構成された2つのミラー化ディスク・プール
- 4枚の10GbEネットワーク・インターフェース・カード

注：100～300GBのシン・プロビジョニングOSブート・ディスクを保持するような比較的大規模なVMware仮想マシンの場合は、ディスク領域を節約するためにLZJB圧縮（LZJBアルゴリズム）を使用することを検討してください。LZJBアルゴリズムは高速で、CPUリソースをあまり消費しません。LZJBを実装したユーザーのほとんどがディスク使用率の向上を確認しており、パフォーマンスの向上が見られることもあります。

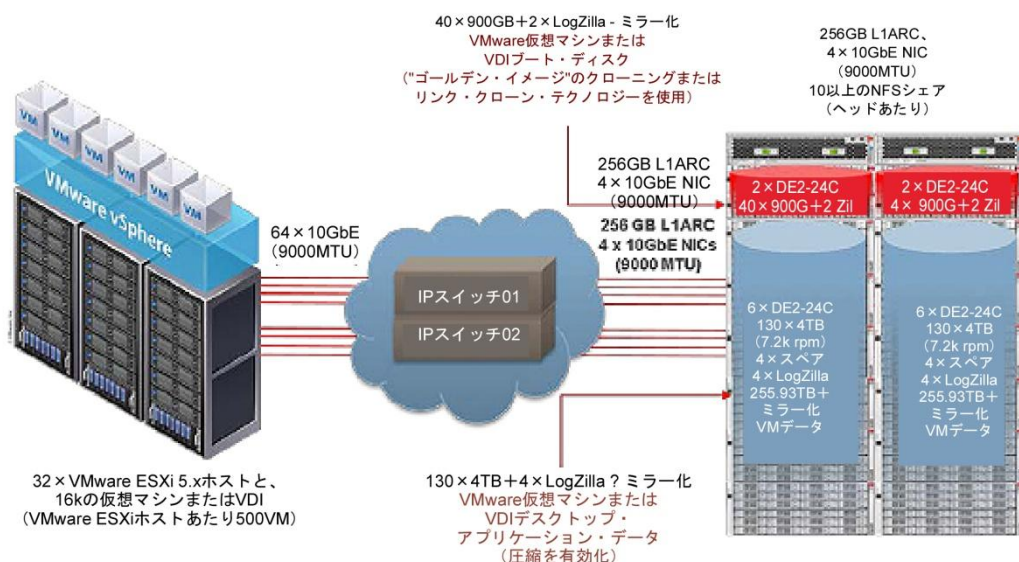


図6: Oracle ZFS Storage Appliance - 容量ディスクを使用したストレージ・レイアウト

ネットワーク設定

『Best Practices for Oracle ZFS Storage Appliance and VMware vSphere 5.x』というホワイト・ペーパー（本書の最後の参考資料を参照）でも説明していますが、NFS向けネットワークの推奨設定は次のとおりです。

- ストレージ・トラフィックをその他のネットワーク・トラフィックから分離します。そのためには、VLANタギング、ネットワーク・セグメンテーション、またはNFSトラフィックのみを処理する専用のIPスイッチを使用します。

- 2つ以上のデュアル10GbE SFP+ PCIe 2.0 Low Profileアダプタを使用します。Oracle ZFS Storage Applianceコントローラあたり合計4つの10GbE接続（10GbEオンボード・インタフェース4つ）は、標準のOracle ZFS Storage ZS3-2コントローラ構成で提供されます。
- すべての10GbE NICおよびIPネットワーク・スイッチで9000バイトのMTUジャンボ・フレームを有効にします。これは非常に重要です。

注：10GbE NICは、IEEE 802.3adリンク・アグリゲーション制御プロトコル（LACP）を使用して（コントローラあたり）1つのチャンネルにバンドルできます。IPMP構成を使用するとネットワークの高可用性が実現され、リンク・アグリゲーションを使用するとネットワークのパフォーマンスが向上します。これらの2つのテクノロジーは互いに補完する関係にあり、仮想デスクトップ環境でのネットワーク・パフォーマンスと可用性の両方の利点を実現するためにまとめて展開できます。

- LACP用、およびソース・アドレスとIPアドレスに基づいたアウトバウンド・ポートの選択用には、LACPポリシーL3を利用します。スイッチの通信モードには、接続のネゴシエーションやリンク・ステータスの監視のためにLACPメッセージを送受信するLACPアクティブ・モードを使用します。
- 次の図の構成に示すように、LACPメッセージ間のタイマー間隔はShortを使用します。

注：一部のネットワーク・スイッチ・ベンダーは、LACPプロトコルをサポートしていません。この場合は、LACPモードを“Off”に設定してください。詳しくは、スイッチ・ベンダーのドキュメントを参照してください。

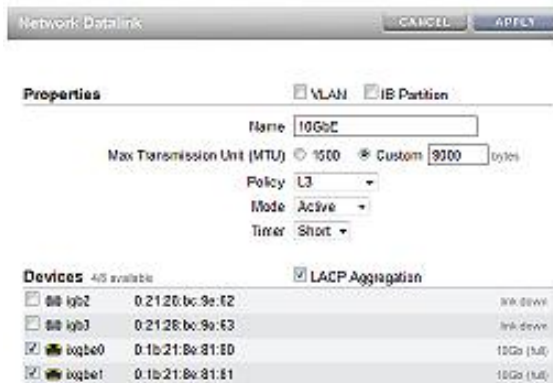


図7：Oracle ZFS Storage ApplianceでのLACP、ジャンボ・フレーム、およびMTUの構成

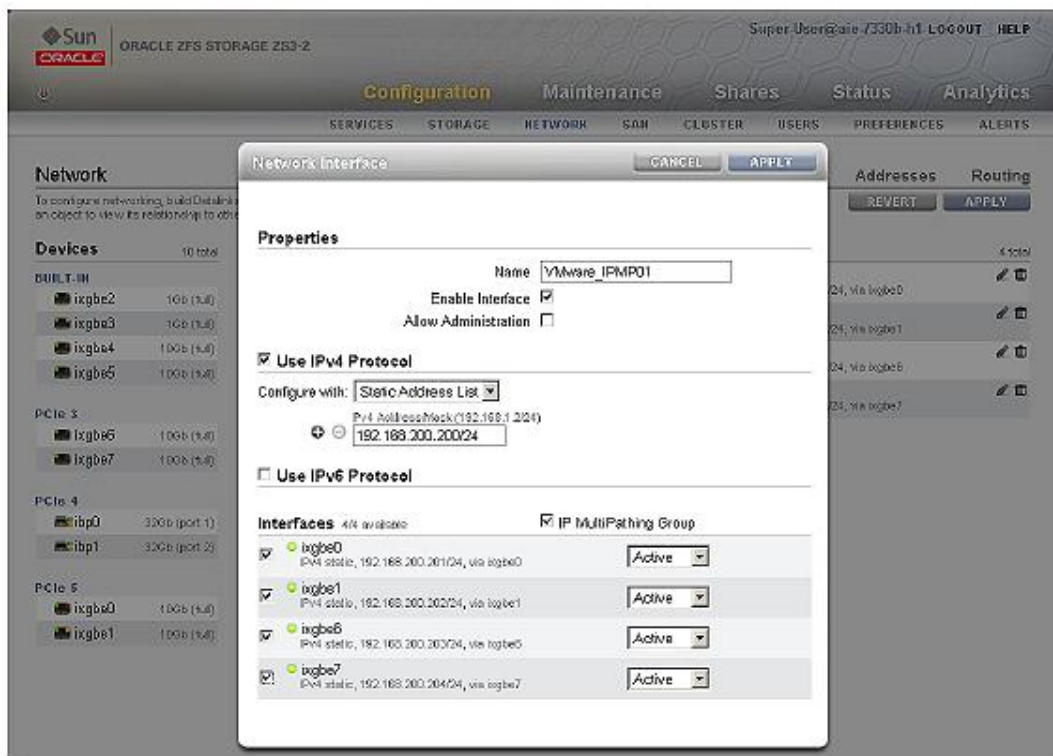


図8: Oracle ZFS Storage ApplianceでのIPMP構成

NFS、プロジェクト、およびシェア

この例は、Oracle ZFS Storage Applianceヘッドあたり1つのプロジェクトと10個の異なるNFSシェアを持つ、アクティブ/アクティブ・モードの1つのOracle ZFS Storage Applianceクラスタを表しています。NFSシェアにはVMware_01~VMware_20と名前が付けられています。VMware仮想マシンが10,000台以上あるような大規模なデプロイメントに推奨される最小構成は、Oracle ZFS Storageヘッドあたり10個のOracle ZFS Storage Appliance NFSシェアを1セット持つ構成です。VMware仮想マシンが10,000台ある使用例では、Oracle ZFS Storage Appliance NFSシェアのそれぞれで500台の仮想マシンがホスティングされ、Oracle ZFS Storage ApplianceヘッドあたりのVMware仮想マシンは合計5,000台です。

次の図は、ヘッドあたり5,000台のVMware仮想マシンをホスティングする場合にOracle ZFS Storage Applianceに推奨されるNFSの（最小）構成です。図10は、Oracle ZFS Storage Applianceのブラウザ・ユーザー・インターフェイス（BUI）に表示されたファイル・システム構成およびマウントポイント構成です。

The screenshot shows a table with the following columns: NAME, SIZE, and MOUNTPOINT. There are 10 rows, each representing a VMware LUN.

NAME	SIZE	MOUNTPOINT
VMware_01	31K	/export/VMware_01
VMware_02	31K	/export/VMware_02
VMware_03	31K	/export/VMware_03
VMware_04	31K	/export/VMware_04
VMware_05	31K	/export/VMware_05
VMware_06	31K	/export/VMware_06
VMware_07	31K	/export/VMware_07
VMware_08	31K	/export/VMware_08
VMware_09	31K	/export/VMware_09
VMware_10	31K	/export/VMware_10

図9 : 10,000台のVMware仮想マシンをホスティングする場合にOracle ZFS Storage Applianceに推奨されるNFS構成

The screenshot shows the Properties window for a VMware LUN. The 'Mountpoint' is set to /export/VMware_01. The 'Inherit from project' checkbox is unchecked. The following properties are configured:

- Mountpoint: /export/VMware_01
- Read only:
- Update access time on read:
- Non-blocking mandatory locking:
- Data deduplication (warning):
- Data compression: Off
- Checksum: Fletcher4 (Standard)
- Cache device usage: All data and metadata
- Synchronous write bias: Latency
- Database record size: 128K
- Additional replication: Normal (Single Copy)
- Virus scan:
- Prevent destruction:
- Restrict ownership change:

図10 : Oracle ZFS Storage ApplianceのBUIに表示されたシェア構成

注 : VMware、およびVMware IPネットワークのNFSプロパティについて詳しくは、本書の最後の参考資料の項に記載したホワイト・ペーパー『Best Practices for Oracle ZFS Storage Appliance and VMware vSphere5』を参照してください。

VMwareのNFSプロトコルに関する推奨事項

重要なのは、VMwareのNFSとTCP/IPの詳細設定を変更することです。これらの構成を変更すると、Oracle ZFS Storage Applianceのフェイルバック操作時またはテイクオーバー操作時のVMware NFSデータ・ストアの高可用性が保証されます。これらを設定すると、VMware ESXiホストあたりに許可されるNFSデータ・ストアの最大数、およびVMware TCP/IPスタックの重要な変更も設定されます。

次の表に示されているパラメータを変更するには、VMware vCenter 5.xサーバーに移動し、VMwareサーバーを選択します。「Software」タブを選択し、「Advanced Settings」をクリックします。「NFS and TCP/IP」を選択し、次のオプションを変更します。

表4 : Oracle ZFS Storage Appliance上のVMware vSphere 5.1データ・ストアに推奨されるNFSとTCP/IPの詳細設定	
オプション	値
NFS.HeartbeatTimeout	5
Nfs.Sendbuffersize	264
Nfs.Receivebuffersize	256
Nfs.MaxVolumes	256
Net.TcpipHeapMax	128
Net.TcpipHeapsize	32
Nfs.heartbeatfrequency	20
Nfs.heartbeatdelta	12
Nfs.heartbeatmaxfailures	10

この構成変更は、クラスタのすべてのVMwareホスト・メンバーで実行する必要があります。新しい設定を有効にするには、各VMwareホストを再起動する必要があります。

仮想マシンの設定に関する考慮事項

仮想マシンのデータ・レイアウトの推奨事項、およびVMware仮想マシンがOracle ZFS Storage Applianceで動作するためのベスト・プラクティスを次に示します。

- VMware仮想マシンのバージョン8を使用します。
- ストレージを効率化するには、シン・プロビジョニング仮想ディスク・ドライブを使用して仮想マシンを構成し、パフォーマンスを向上させるには、VMware準仮想化SCSIコントローラ・タイプを使用してVMware仮想マシンを構成します。
- パフォーマンスを向上させるには、VMXNET3ネットワーク・アダプタを使用します。
- VMwareクライアント・ツールをインストールします。これらのツールとそのインストール方法について詳しくは、次のリンクを使用してください。

http://www.vmware.com/files/jp/pdf/vmware-tools-installation-configuration_JA.pdf

- パーティション・アライメントは重要です。仮想マシンが正しいパーティション・アライメントで動作していることを確認してください。

結論

Oracle ZFS Storage Applianceは、仮想化環境のための優れたパフォーマンスを提供します。このアーキテクチャを活用することで、ブート・ストーム・トラフィックの影響を受けずに大量の仮想マシンを最適なパフォーマンスと低レイテンシでデプロイすることができます。Oracle ZFS Storage Applianceの各種機能とインテリジェント・キャッシュ・テクノロジーは何十万ものIOPSに対応できるように設計されており、ハイブリッド・ストレージ・プール機能は、仮想マシンおよび仮想デスクトップの数が上限に達している環境のサポートに必要な馬力を提供します。

Oracle ZFS Storage ApplianceのパフォーマンスとVMware vSphereハイパーバイザの組合せは、仮想化環境に最適な選択肢です。

付録A：ベンチマーク結果

Oracle ZFS Storage Applianceのテスト結果について詳しくは、次のWebサイトを参照してください。

SPC-2の結果

<http://www.spec.org/sfs2008/results/res2013q3/sfs2008-20130819-00227.html>

http://www.storageperformance.org/benchmark_results_files/SPC-2/Oracle_SPC-2/B00067_Oracle_ZFS-ZS3-4/b00067_Oracle_ZFS_Storage_ZS3-4_SPC-2_full-disclosure-report.pdf

http://www.storageperformance.org/results/benchmark_results_spc2/#sun_spc2

付録B：参考資料

このドキュメントで説明した製品に関連する追加情報については、次のリソースを参照してください。

Sun ZFS Storage Appliance、Sun ZFS Storage 7000、ZFS Storage Applianceはすべて、同じOracle ZFS Storage Appliance製品ファミリのことを指しています。引用されているドキュメントの中には、現在でもこれらの過去のネーミング規則に従っているものがあります。

- Oracle ZFS Storage Appliance ドキュメント・ライブラリ（インストール、分析、顧客サービス、管理ガイドなど）：
<http://www.oracle.com/technetwork/documentation/oracle-unified-ss-193371.html>
- 『Oracle ZFS Storage Appliance Administration Guide』はOracle ZFS Storage Applianceのヘルプ・コンテキストでも参照できます。
Oracle ZFS Storage Applianceのヘルプ機能には、ブラウザ・ユーザー・インタフェースからアクセスできます。
- Oracle ZFS Storage Applianceソフトウェア・リリースは、次のURLからダウンロードできます。
<https://wikis.oracle.com/display/FishWorks/Software+Updates>
<http://www.oracle.com/technetwork/jp/server-storage/sun-unified-storage/downloads/index.html>
- Oracle Support Center
<http://www.oracle.com/jp/support/index.html>
- パッチと更新プログラムはMy Oracle Support (MOS) からダウンロードできます (Oracle ZFS Storage Software Patchesの下を検索してください)

- Oracle ZFS Storage Applianceの各種プラグイン
<http://www.oracle.com/technetwork/server-storage/sun-unified-storage/downloads/zfssa-plugins-1489830.html>
- オラクルのストレージ製品に関する情報
<http://www.oracle.com/jp/products/storage/overview/index.html>
- 『Best Practices for Oracle ZFS Storage Appliance and VMware vSphere5.x』をはじめとするOracle ZFS Storage Applianceのテクニカル・ホワイト・ペーパーおよびソリューション概要
<http://www.oracle.com/technetwork/jp/server-storage/sun-unified-storage/documentation/index.html>
- VMware
<http://www.vmware.com>

ORACLE®

10,000台以上のVMware仮想マシンの

Oracle ZFS Storage Applianceによるデプロイ

2014年7月、バージョン1.0

著者：Application Integration Engineering、
Anderson Souza

Oracle Corporation
World Headquarters
500 Oracle Parkway

Redwood Shores, CA 94065

U.S.A.

お問い合わせ窓口

Oracle Direct

TEL 0120-155-096

URL oracle.com/jp/direct



Oracle is committed to developing practices and products that help protect the environment

Copyright © 2014, Oracle and/or its affiliates. All rights reserved. 本文書は情報提供のみを目的として提供されており、ここに記載される内容は予告なく変更されることがあります。本文書は一切間違いがないことを保証するものではなく、さらに、口述による明示または法律による黙示を問わず、特定の目的に対する商品性もしくは適合性についての黙示的な保証を含み、いかなる他の保証や条件も提供するものではありません。オラクル社は本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクル社の書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。

OracleおよびJavaはOracleおよびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。

IntelおよびIntel XeonはIntel Corporationの商標または登録商標です。すべてのSPARC商標はライセンスに基づいて使用されるSPARC International, Inc.の商標または登録商標です。AMD、Opteron、AMDロゴおよびAMD Opteronロゴは、Advanced Micro Devicesの商標または登録商標です。UNIXはX/Open Company, Ltd.によってライセンス提供された登録商標です。0611

Hardware and Software, Engineered to Work Together