Oracle Elite
Engineering Exchange

Oracleホワイト・ペーパー
2013年10月

オラクルのSPARC M5-32サーバーと SPARC M6-32サーバー： ドメイン構成のベスト・プラクティス

ORACLE

はじめに

エンタープライズ統合の利点は十分に理解されています。ワークロード、アプリケーション、データベース、オペレーティング・システム・インスタンス、およびサーバーを統合することで、管理対象のリソース数を削減でき、その結果システム使用率の向上とコストの削減につながります。使用率が向上すると、ハードウェアを追加購入する必要性は低くなります。統合に、ITインフラストラクチャ全体の簡素化も組み合わせることができれば、データセンターの運用コストは大幅な削減が可能です。セキュリティの向上、予測可能性の高いサービス・レベルの確保、アプリケーションのデプロイの柔軟性の向上など、統合は戦略的な目標の達成にも役立ちます。オラクルのサーバー製品ラインにSPARC M6-32が加わり、価格と性能はBig Iron（高価な超高速マシン）またはその強化機能のコストが生じることなく、リニアに拡張されます。つまり、16 SPARC T5-2に合計32個のCPUを搭載した場合と、SPARC M6-32に32個のCPUを搭載した場合とは、同じように価格が設定されるということです。

そのため、従来このクラスのシステムに付随していた高い価格設定を外すことにより、ユーザーがサーバーを大型化する理由がまた1つ増えることになります。言い換えれば、SPARCプラットフォームの場合、小規模な多数のサーバーを調達しても、大規模なサーバーを1つ調達する場合と比べて安価になるわけではないということです。

統合のデプロイメントを成功させるためには、多数のアプリケーション・インスタンスに対応するスケーラビリティを持つサーバー・プラットフォームを選択することが求められます。さらに、このサーバー・プラットフォームには、ミッション・クリティカルなアプリケーションに必要となる高可用性や、多数のアプリケーションの管理を簡素化するためのリソース管理機能および仮想化機能、統合環境を管理するための各種ツールも必要です。

オラクルのSPARC M5-32サーバーとSPARC M6-32サーバーはこれらの要件のすべてに対応する、理想的なサーバー統合ソリューションです。SPARC M5-32サーバーとSPARC M6-32サーバーを利用すれば、ITマネージャーは迅速かつ動的に割当て可能なコンピューティング・リソースのプールを作成して、新しいワークロードや常に変化するワークロードに対応できます。

サーバーとアプリケーションを統合する理由

アプリケーションは従来、アプリケーション・インスタンスごとに1台のサーバーにデプロイされていました。しかし、複雑なエンタープライズ・アプリケーションの場合、このようなデプロイ方法では、Web層、アプリケーション層、データベース層のそれぞれに対して異なるサーバーが用意されることから、1つのアプリケーションを運用するためにデータセンターに多数のサーバーが必要になります。

また、多くのエンタープライズ・アプリケーションで、本番サーバーに加えてテスト用と開発用のサーバーも必要になります。一般的に、本番サーバーの初期導入の時点では、ワークロードの急増に対応するための余力（ヘッド・ルーム）が十分にありますが、アプリケーションの規模が拡大したときに、容量を追加する方法がサーバーの追加以外にないため、複雑性が高まります。サーバー数が増加するにつれ、管理を要するオペレーティング・システム（OS）インスタンス数も増加します。その結果、複雑性がさらに数段階高まり、ITの柔軟性が低下します。

通常、サーバーあたり1アプリケーションというデプロイメント・モデルの場合に、サーバー使用率は10～30%と非常に低くなるため、サーバー・リソースの使用効率が良くありません。各サーバーにはワークロードの急増に対応するための十分な規模が必要ですが、普段必要になるサーバーの容量はほんの一部です。

この点について、1つのアプリケーション・インスタンスを多数の小規模なサーバーで運用するような状況を図1に示します。これらのサーバーの1台1台に、ピーク時の容量要件に対応するための十分な余力が必要となり、容量の追加を必要としているサーバーや余分な容量があるサーバーと、余力を"共有"することはできません。

これらのサーバーで余力を共有して、必要に応じて貸し借りできるとすれば、サーバー使用率は高くなるでしょう。複数のアプリケーションを1台の大規模なサーバー上に統合し、このサーバー内でリソースをアプリケーション間で動的に移動させれば、ワークロードのピークと谷間が均一に近づき、全体的なコンピューティング要件が変動しにくくなります。統合対象のアプリケーションが増加するほど、サーバー使用率も均一化します。大規模なサーバー上に統合されたアプリケーションは共有の余力を活用できるため、アプリケーションの統合によって余分な容量が大幅に削減され、サーバー使用率が大幅に改善されます。

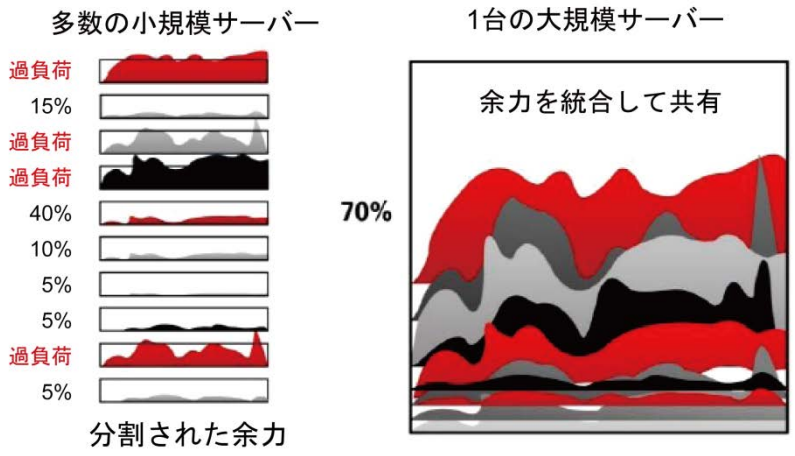

図1：大規模な対称型マルチプロセッシング・サーバーでの余力の統合と共有

サーバー使用率の向上は、サーバー・リソースの使用効率の向上を意味します。そのため、ROIが改善し、ワークロードの要件に対応するために必要となる合計のサーバー・ハードウェア数が減少します。

古い小規模なサーバーを少数の大規模な最新型サーバーに統合することには、使用率の向上以外にも多くの利点があります。最新型のサーバーは多くの容量を備え、パフォーマンスや電力効率、スペース効率に優れ、高度な可用性機能があり、管理も容易です。

統合の要件

統合に使用するサーバーでは、スケーラビリティ、大容量、および高可用性を備え、単純なアップグレード・パスを利用できる必要があります。また、既存のアプリケーションの再利用が可能であり、効率的な仮想化ツールおよびリソース管理ツールが付属していることも求められます。複数のアプリケーションが統合サーバー上に組み合わされるので、統合サーバーにはあらゆるタイプの大量のワークロードに対処できる処理能力が必要です。さらに、他のアプリケーションと統合された各アプリケーションのパフォーマンスは、専用サーバーに単独でデプロイされた場合に匹敵するか、それを超えるものである必要もあります。

統合は、当然ながら"かごの中に入れる卵を増やす"ことになるため、各アプリケーションを専用サーバーにデプロイした場合よりも、アプリケーション可用性に対するシステム障害の影響が大きくなります。

統合に使用するサーバーには、ハードウェアとソフトウェアの両方で、計画停止時間と計画外停止時間を短縮するための高可用性が求められます。統合サーバーでは、ほとんど停止することがないよう極度に高い信頼性が必要です。また、最小限の停止時間または停止時間なしでの再構成、アップグレード、および修復が可能となる高度な保守性も必要になります。

統合サーバーはおもに、古いアプリケーションを新しい環境で運用するために使用します。そのため、新しいアプリケーションだけではなく従来のアプリケーションが実行できる必要もあります。

統合環境には異なるタイプのワークロードが多く存在し、これらの多様なワークロードのそれぞれに固有のパッチ、リソース、セキュリティ、パフォーマンスの要件があります。多くの場合、複数のアプリケーションを管理するための十分なツールがオペレーティング・システムに用意されていますが、そうでない場合は、アプリケーションを効率的に実行するための個別の環境が必要です。統合サーバー内のリソース・プールをパーティション化し、必要に応じて複数のアプリケーション向けにデプロイするための仮想化ツールおよびリソース管理ツールが必要になります。仮想化によってアプリケーションの分離が強制され、また、リソース管理によって各アプリケーションのパフォーマンス要件に確実に対応できるようになります。

大規模で垂直方向にスケーラブルなSMPサーバー上での統合

オラクルのSPARC M6-32のような大規模な対称型マルチプロセッシング（SMP）サーバーには多数のプロセッサとI/Oスロット、および数テラバイトのRAMが、すべて1台のキャビネット内に搭載されており、単一の大規模なOSインスタンスにデプロイすることも、リソース管理されるドメインに分けることも柔軟に可能です。

基本的に、垂直方向にスケーラブルなサーバーは、さまざまなサイズおよびタイプの大量のワークロードに対応可能な大規模なリソース・プールであり、統合やアプリケーションのデプロイを簡素化します。新しいアプリケーションを1台の大規模なSMPサーバー上にデプロイできるため、新しいアプリケーションを導入するたびにサーバーをインストールする必要がありません。また、既存のアプリケーションは、利用可能な余力を利用して拡張できます。

垂直方向にスケーラブルなハイエンドSMPサーバー

あらゆるサーバーは同じような中核的コンポーネントで構成されていますが、サーバー・アーキテクチャが異なる場合、これらのコンポーネントの組合せ、接続方法、および使用方法も異なります。

垂直方向にスケーラブルなサーバー（通常は8基以上のプロセッサをホストする大規模なSMPサーバー）は、1つのOSインスタンスで複数のプロセッサ、メモリ・サブシステム、およびI/Oコンポーネントを管理します。これらのコンポーネントは1台のシャーシ内に格納されます。オラクルのSPARC M6-32サーバーのような垂直方向のスケーラビリティにもっとも優れたサーバーでは、仮想化ツールによりパーティション化して、サーバー・リソースのサブセットを使用して複数のOSインスタンスを作成することもできます。仮想化ツールは、ワークロードやセキュリティ要件、および可用性要件に応じてリソースの共有や分離を行うために使用します。

垂直方向にスケーラブルな設計において、システム・インターコネクトは一般的に、緊密に結合されたセンタープレーンまたはバックプレーンとして実装されます。これらは、低レイテンシと高帯域幅の両方を実現します。垂直型システム（またはSMPシステム）では、メモリは共有され、ユーザーに対しては1つの実体として示されます。すべてのプロセッサとすべてのI/O接続がすべてのメモリに同等にアクセスできるため、データのデプロイ状況について考慮する必要はありません。オラクルのハイエンドSPARC SMPサーバーは1993年よりリニアなスケーラビリティを達成してきており、緊密に結合された高速で低レイテンシインターコネクトの価値を実証してきました。

キャッシュ・コヒーレンシ・インターコネクトにより、キャッシュまたはメモリ上のデータの場所にかかわらず、すべてのデータの場所に関する情報が維持されます。SMPサーバーでは、内蔵インターコネクトによりすべてのデータ移動が自動的かつ透過的に処理されるため、クラスタ・マネージャやネットワーク・インターコネクトはありません。リソースをシャーシに追加するには、追加のプロセッサ、メモリ、およびI/Oサブアセンブリが搭載されたシステム・ボードを取り付けます。垂直的なアーキテクチャに大規模なSMPサーバーのクラスタを追加して、1つの大規模なアプリケーションに対して使用することもできます。

ハイエンドSMPサーバーは、アプリケーションのデプロイと統合を大幅に簡素化します。大規模なSMPサーバーには、容易にパーティション化できるプロセッサ、メモリ、およびI/Oリソースで構成される大きなプールが備わっています。Oracle Solaris Resource Managerを使用して、このリソース・プールをアプリケーションに動的に割り当て、Oracle Enterprise Manager Ops Centerなどの標準的なシステム管理ツールを使用して操作することができます。

SPARC M5-32サーバーとSPARC M6-32サーバーの統合テクノロジー

これ以降の項では、多数のアプリケーションをまとめてデプロイしてシステムの使用率を向上させ、コンピューティング・リソースの使用を最適化して、IT投資からより多くのROIを引き出すための統合テクノロジーについて見ていきます。次ページの図2に、オラクルの現行のSPARC Enterprise Mシリーズにおいて無償で利用できるさまざまなレベルの仮想化テクノロジーを示します。

仮想化スタックでもっとも低いレイヤーはSPARCプラットフォームです。SPARCプラットフォームは、仮想化の第1レベルとしてSPARC Enterprise Mシリーズのダイナミック・ドメイン機能（物理ドメイン、またはPDomとも呼ばれる）を提供します。これは電気的に分離したハードウェア・パーティションであり、他のPDomにはまったく影響せずに電源のオン/オフと操作を行うことができます。

仮想化の第2レベルでは、各PDomがさらにハイパーバイザ・ベースのOracle VM Server for SPARCパーティション（LDomとも呼ばれる）に分割されます。このパーティションがそれぞれ独自のOracle Solarisカーネルを実行し、独自にI/Oリソースを管理します。Oracle VMの元で、バージョンの異なるOracle Solarisが異なるパッチ・レベルを実行することも珍しくありません。Oracle VMは、ソフトウェア・ライセンシングの目的上Oracleハード・パーティションとしても認識されます。

仮想化の第3レベルはOracle Solaris Zonesです。これは、もっとも粒度の高い仮想化レベルであり、Oracle Solarisの機能の1つです。Oracle Solaris Zonesの各ゾーンは、共通のOracle Solarisカーネルとパッチ・レベルを共有します。作成と再起動の点で柔軟性がきわめて高いという利点があり、非常に高速で軽量です。Oracle Solarisの各インスタンスは、Oracle Solaris Resource Managerを利用して、アプリケーションで消費できるCPUまたはメモリを制限し、Oracle Enterprise Manager Ops Centerによって管理されます。

こうした仮想化技術のすべてが、多数のアプリケーションを1台のサーバーに統合する上で高い威力を発揮します。これ以降では、これらの仮想化とリソース管理の各種テクノロジーについて詳細に説明します。

パフォーマンスの重要性が高いワークロードを実行するために、Oracle VM Server for SPARCのドメインは、各ドメインを独自のPCIeスロットに直接接続するような構成が可能です。この構成では、PDomあたりのOracle VM Server for SPARCの総数が制限されますが、パフォーマンスと分離性に関してはかなり有利になります。Oracle SuperCluster M6-32はこのタイプのOracle VM Server for SPARCドメインを使用しますが、この構成はSPARC M5-32とSPARC M6-32の両プラットフォームにも同様に適用できます。

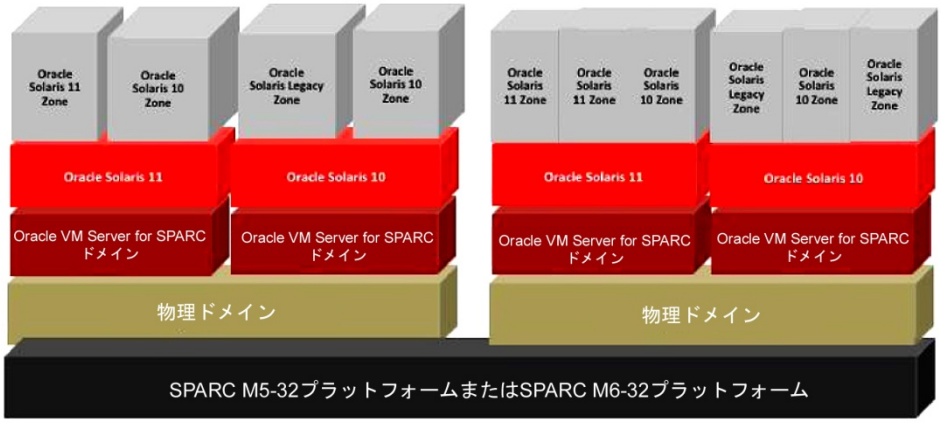

図2：SPARC M5-32サーバーまたはSPARC M6-32サーバーの仮想化テクノロジー・スタック

ダイナミック・ドメイン

前述したように、ダイナミック・ドメイン（物理ドメイン、またはPDomとも呼ばれる）は、電子的に分離されたパーティションを可能にする機能です。PDomを利用すれば、複数のアプリケーションや複数のOracle Solaris OSコピーを1台のサーバー上で分離できます。管理者はダイナミック・ドメインの機能によってハードウェア障害やセキュリティ障害を分離して、各ドメインに対する影響を制限できます。その結果、優れたレベルのシステム可用性とセキュリティを確保できます。ダイナミック・ドメインの機能は現在6世代目であり（以前はオラクルのSPARC Enterprise Mシリーズ・サーバーで利用できました）、UNIXサーバー市場においてもっとも成熟し定着しているパーティション化オプションとなっています。前述したように、Oracle VM Server for SPARCを実行することによってPDomをさらに仮想化できるため、独立した複数のOracle Solarisインスタンスが同じ物理ドメイン内で共存することも可能です。

ダイナミック・ドメインを使用すれば、ソフトウェアおよびハードウェアのエラーや障害が、障害の発生したドメインを超えて伝播することはありません。ダイナミック・ドメイン間で障害を完全に分離することで、あらゆるハードウェアまたはソフトウェアのエラーによるアプリケーションへの影響が限定的なものになります。そのため、これらのサーバーでは高いレベルの可用性が維持されます。このような可用性は、多数のアプリケーションを統合する場合に不可欠です。ダイナミック・ドメインの機能は、各ドメインの管理を分離するため、あるドメインでのセキュリティ違反によって他のドメインが影響を受けることはありません。

Oracle VM Server for SPARC

Oracle VM Server for SPARCは、独立したオペレーティング・システム・インスタンスを実行する完全な仮想マシンを実現し、オラクルのSPARC Tシリーズと新しいSPARC Mシリーズをベースとするすべてのプラットフォームで使用できます。LDom、または論理ドメインとも呼ばれ、各オペレーティング・システム・インスタンスには、専用のCPU、メモリ、ストレージ、コンソール、および暗号化デバイスが含まれます。LDomには、仮想化機能の多くが基盤となるハードウェアによってネイティブで提供され、またCPUもメモリも直接ドメインに割り当てられるため仮想化のオーバーヘッドが生じないという特長があります。I/Oを直接ドメインに割り当てればパフォーマンスが高くなるというメリットがある一方、I/Oを仮想化すれば、ハードウェア・リソースの利用率が向上し、ライブ・マイグレーションも可能になるというメリットがあります。システムで使用できるドメインの数は物理スレッドの数によって制限されますが、SPARC M5-32またはSPARC M6-32の場合、サーバーまたはPDomあたり128ドメインという上限があります。ルート・ドメイン・モデルがOracle SuperCluster M6-32の場合と同様にデプロイされており、ドメインが完全なPCIeスロットの所有権を排他的に付与されている場合、このタイプのドメインの数はPDomで使用可能なI/Oカードの数量によって制限されます。システムで使用できるドメインの数は物理スレッドの数によって制限されますが、SPARC M5-32またはSPARC M6-32の場合、サーバーまたはPDomあたり128ドメインという上限があります。ルート・ドメイン・モデルがOracle SuperCluster M6-32の場合と同様にデプロイされており、ドメインが完全なPCIeスロットの所有権を排他的に付与されている場合、このタイプのドメインの数はPDomで使用可能なI/Oカードの数量によって制限されます。LDomドメイン内でOracle Solaris Zonesを使用すると、仮想化の第3のレイヤーが成立し、このようにOracle VM Server for SPARCのドメイン数が減るという影響が緩和されます。

Oracle VM Server for SPARCは、システム間でドメイン間のライブ・マイグレーションを実行する機能を備えています。ライブ・マイグレーションという名前が示すように、ソース・ドメインとアプリケーションを中断もしくは停止させる必要はありません。そのため、同じサーバー上または別のサーバー上の別のPDomに論理ドメインを移行できます。Oracle VM Server for SPARCの実装環境でライブ・マイグレーションを使用するのは一般的ですが、SPARC M5-32またはSPARC M6-32のプラットフォームで予想される主たるワークロードは、もっともパフォーマンスの高いI/Oを必要とする可能性があるため、そうなるとライブ・マイグレーションは使用できなくなります。一方、SPARC M5-32またはSPARC M6-32のプラットフォームにデプロイでき、ライブ・マイグレーションが理想的な二次的なワークロードも数多くあります。

ダイナミック・ドメインの上層に論理ドメインを重ねれば、電気的に分離された複数のドメインに複数のオペレーティング・システムを同時にデプロイできる柔軟性が得られます。これらのドメインのすべてでOracle Solarisが稼働し、そこでさらにOracle Solaris Zonesをホストすれば、新たな仮想化レイヤーが追加されます。

Oracle Solaris

Oracle Solaris OSは、ある特定のサーバーまたはドメイン内のすべてのプロセッサに対して多数のアプリケーション・プロセスのスケジューリングを非常に効率よく実行し、ワークロードに基づいてプロセッサ間でプロセスを動的に移行します。たとえば、多くの企業は仮想化ツールを使用せずに1台のSPARCサーバー上で100個を超えるOracle Databaseインスタンスを運用しています。Oracle Solarisでは、SPARCのすべてのコアおよびスレッドですべてのデータベース・プロセスを効率的に管理してスケジューリングできます。

この方法により、垂直方向にスケーラブルな大規模サーバーで、サーバー上に存在する多くのユーザーやアプリケーション・インスタンスに、必要に応じてリソースを割り当てることができます。Oracle Solaris OSを利用してワークロードのバランスを取ることで、必要な処理リソースを削減できるため、より少ないプロセッサやメモリ容量で運用して、調達コストを削減できます。Oracle Solarisによって、柔軟性が向上し、ワークロード処理が分離され、サーバー使用率が最大になる可能性が高まります。

Oracle Solaris Zones

統合環境では時に、各アプリケーションを個別に管理できる状態を維持する必要があります。セキュリティ要件の厳しいアプリケーションや他のアプリケーションとの共存が難しいアプリケーションもあるため、ITリソース使用率を制御し、それぞれのアプリケーションを分離し、同じサーバー上で複数のアプリケーションを効率的に管理できる必要があります。

Oracle Solaris Zonesテクノロジー（旧称Oracle Solaris Containers）は、Oracle Solarisを実行するすべてのサーバーで利用できる、コンピューティング・リソースの仮想化を行うためのソフトウェア・ベースの手法です。障害が分離された複数のセキュアなパーティション（ゾーン）を1つのOracle Solaris OSインスタンス内に作成できます。複数のゾーンを稼働させて、1つのOSインスタンス内に異なる多数のアプリケーションを共存させることが可能です。

ゾーン環境には、リソース使用量の高度なアカウンティング機能も含まれます。この非常に粒度の細かい大規模なリソース追跡機能によって、一部の統合環境で必要とされる高度なクライアント課金モデルに対応できます。

Oracle Solaris Resource Manager

Oracle Solaris Resource Managerとは、Oracle Solaris Zonesを含むOracle Solarisインスタンス内で、CPU、メモリ、およびI/Oの各リソース消費をアプリケーション間で割り当て、共有できる一連の技術です。Oracle Solaris Resource Managerは、リソース・プールを利用してシステム・リソースを制御します。各リソース・プールにはリソース・セットと呼ばれるリソースの集合を格納できます。このリソース・セットには、プロセッサ、物理メモリ、またはスワップ領域を含めることができます。さらに、必要に応じて、リソース・プール間でリソースを動的に移行できます。また、Oracle Solaris 11では、ネットワーク・サービス仮想化の機能も大幅に強化されました。

フェアシェア・スケジューラ

Oracle Solaris Resource Managerには、リソース・プール内で使用できる高度なフェアシェア・スケジューラが組み込まれています。管理者はフェアシェア・スケジューラを使用して、1つ以上のプロセスで構成されるワークロードにプロセッサ・シェアを割り当てます。

このシェアによって、他のワークロードと比較した相対的な重要度をワークロードに指定できます。フェアシェア・スケジューラはこの重要度をプロセッサ・リソースの割合に変換して、リソースをワークロードに対して予約します。ワークロードでプロセッサ・リソースが要求されない場合は、それらのリソースが他のワークロードによって使用されることもあります。ワークロードに対するシェアの割当てによって、最小限のプロセッサ・リソースを効率的に予約して、重要なアプリケーションに必要となるサーバー・リソースを確実に割り当てることができます。

統合テクノロジーの管理

Oracle Enterprise Manager Ops Center

サーバー統合のおもな目的に、管理を要するサーバーおよびOSインスタンスの個数を削減してサーバー管理を簡素化することがあります。Oracle Enterprise Manager Ops Center 12cは、システム・インフラストラクチャ資産の管理を統一された1つの管理コンソールに統合することで、この目的を達成します。

Oracle Enterprise Manager Ops Center 12cでは、高度なサーバー・ライフ・サイクル管理機能を通じて、ファームウェア、オペレーティング・システム、および仮想マシンを含む、サーバー、ストレージ、およびネットワーク・ファブリックの管理を統合した集約型のハードウェア管理手法を実現します。Oracle Enterprise Manager Ops Center 12cでは、資産の検出、資産のプロビジョニング、監視、パッチ適用、およびワークフローの自動化を利用できます。また、物理サーバーに加えて仮想サーバーの検出と管理も可能であり、SPARC M5-32、SPARC M6-32、およびOracle SuperCluster M6-32のようなハイエンド・サーバーやデータセンター内の他のOracleサーバーすべての管理を簡素化します。Oracle Enterprise Manager Ops Center 12cは、Oracle Premier Support契約を結ぶすべてのOracleサーバーの顧客に無償で提供されます。

SPARC M5-32サーバーとSPARC M6-32サーバーによる統合のレイヤー化

垂直方向に拡張可能なシステムのもっとも重要な側面は、デプロイメント・モデルの柔軟性です。水平方向に拡張される環境の場合、選択できる仮想化技術は通常、VMだけです。垂直方向に拡張可能なシステムの場合、多くの層で仮想化が可能であり、それが利用率と簡便性の向上につながっています。

SPARC M6-32とSPARC M5-32には、インフラストラクチャのレベルでレイヤー化される仮想化がおもに3種類あります。

1. Oracle Solaris Zones：1つのOSインスタンス内で、複数アプリケーションの共存とリソース管理が可能です。
2. Oracle VM Server for SPARC：同じ物理インフラストラクチャで複数のOSインスタンスが共存でき、ハードウェア・リソースは動的に再割当てされます。
3. ダイナミック・ドメイン：1台のサーバーを独立分離した複数のサーバーにパーティショニングします。

それぞれの仮想化技術に独自の利点があります。一般的にはゾーンが、動的なリソース利用率も柔軟性ももっとも高くなりますが、分離性はもっとも低く、サービスの粒度も低くなります。ダイナミック・ドメイン機能は、分離性がもっとも高くなりますが柔軟性が劣ります。もっとも適切なデプロイメント・モデルはおそらく、これらのテクノロジーすべてを複合したアプローチでしょう。オラクルのソフトウェア・ライセンシングの目的上、ダイナミック・ドメイン、Oracle VM Server for SPARC、およびOracle Solaris Zonesはすべてライセンス・ソフトウェアのハード・パーティションとみなされます。[1]

統合の考え方

統合のオプションが複数ある場合には、まず統合を考えた当初の理由を念頭に置き、その初期要件に基づいてもっとも適切なソリューションを推し進めると良いでしょう。

- 運用効率の最大化
    - 統合の利点は、ハードウェア・コストの削減だけに限ったことではありません。統合の最大の利点は、運用モデルの標準化と管理対象オブジェクト数の削減によって得られる簡便性です。
    - 可能な限りスタックの上層までを統合することによって、管理対象オブジェクトの数が無理なく減り、可能な限りの標準化も進みます。
- ワークロード効率の最大化
    - 分離を進めた場合に生じるトレードオフの1つとして、仮想化のオーバーヘッドが増えるという可能性があります。この点を踏まえて、分離の追加は必要な場合にだけ行うようにしてください。
    - 現在のOSインスタンスのフットプリントに比べれば、ごく小さいワークロードもあります。可能な場合には、OSインスタンスごとに複数のワークロードの共存を試みてください。

両極端の2つと、その間にいくつかのオプションを考えることができます。

[1] これらの技術をハード・パーティション境界として使用する際の最新のルールについては、Oracle Partitioning Policyを参照してください：http://www.oracle.com/us/corporate/pricing/partitioning-070609.pdf

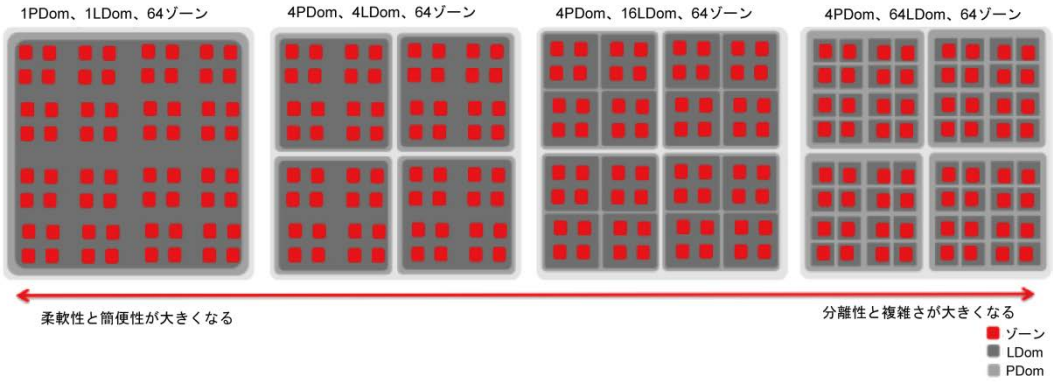

図3：分離性と柔軟性のレベルが異なるワークロードをデプロイするオプション

独立した4つのダイナミック・ドメイン（PDom）を用意し、各PDomで16個のOracle VM Server for SPARC論理ドメイン（LDom）を実行して、各ワークロードで1つのOSインスタンスを使用すれば、PDomあたり最高で64個のワークロードをデプロイできます。

このレイアウトでは分離のレベルが最高になりますが、その代わりとしてかなり複雑になります。64個のLDomをサポートするには、128個の仮想ブート・ディスクを構成しなければならないからです。各ドメインと64個の固有OSインスタンスにサービスを提供するには、さらに追加でサービス・ドメインが必要になります。

これとは正反対に、システム全体で1つのPDomを用意し、そこで1つだけのOracle Solarisインスタンスを実行して、64個のゾーンそれぞれでワークロードを実行することもできます[2]。

このオプションは、リソース利用率と運用の簡素化という点ではもっとも効率的です。しかし、単一の障害ドメインが生じるため、保守性と管理性に多くの問題が生じ、計画停止に対しても無計画停止に対しても大きな影響を及ぼします。この図で、できるだけ左寄りになることを目指し、分離と保守性の要件が必須の場合だけ右方向に進むようにしてください。

実際には、当該のワークロードの特性に基づいた最適なソリューションは、この両極端の間のどこかに収まります。このホワイト・ペーパーは、企業が情報に基づいて選択できるように、仮想化テクノロジーの3つのレイヤーについて十分に論じることを目的としています。

[2] 64個のゾーンが上限ということではありません。理論上の制限は、Oracle Solarisインスタンスあたり8,000ゾーン以上です。

ダイナミック・システム・ドメイン（PDom）

SPARC M5-32サーバーとSPARC M6-32サーバーの特徴はバランスの取れたスケーラビリティの高いSMP設計です。高速で低レイテンシのシステム・インターコネクトを利用してメモリとI/Oに接続されているSPARCプロセッサが使用されます。

ドメイン構成ユニット（DCU）は、システムにおけるPDomのハードウェア構成要素であり、1台、2台、または4台のCPU/メモリ・ユニット（CMU）ボード、1台のI/Oユニット、および1台のサービス・プロセッサ・プロキシ（SPP）ボードによって構成されます。

システムは、PDomと呼ばれる最大4つの障害分離パーティションに物理的に分割され、Oracle SolarisまたはOracle VM Server for SPARCの独立したインスタンスがそれぞれで実行されます。

PDomは、シャーシの他のPDomからハードウェアとして完全に分離した独立のサーバーのように動作します。1つのPDomでハードウェアまたはソフトウェアの障害が発生しても、シャーシの他のPDomには影響しません。オラクルのソフトウェア・ライセンシングの目的上、PDomはハード・パーティションとみなされます。つまり、本番データベースに8ソケットのPDomを使用する場合、8ソケットにコア/ソケット数を掛けた分のライセンス料金が必要になるということです。

PDomには、1、2、3、または4個のDCUを含めることができます。DCUが複数の場合は、スケーラビリティ・スイッチ・ボード（SSB）を使用してメモリ・アクセスと、システム全体を通じたキャッシュ・コヒーレンシを実現できます。PDomに1つのDCUしか含まれない場合、SSBを使用する必要はなく、"境界"PDomとして構成することによってPDomをさらに分離することもオプションとして可能です。これは、SSBへのアクセスを無効にする機能を果たします。

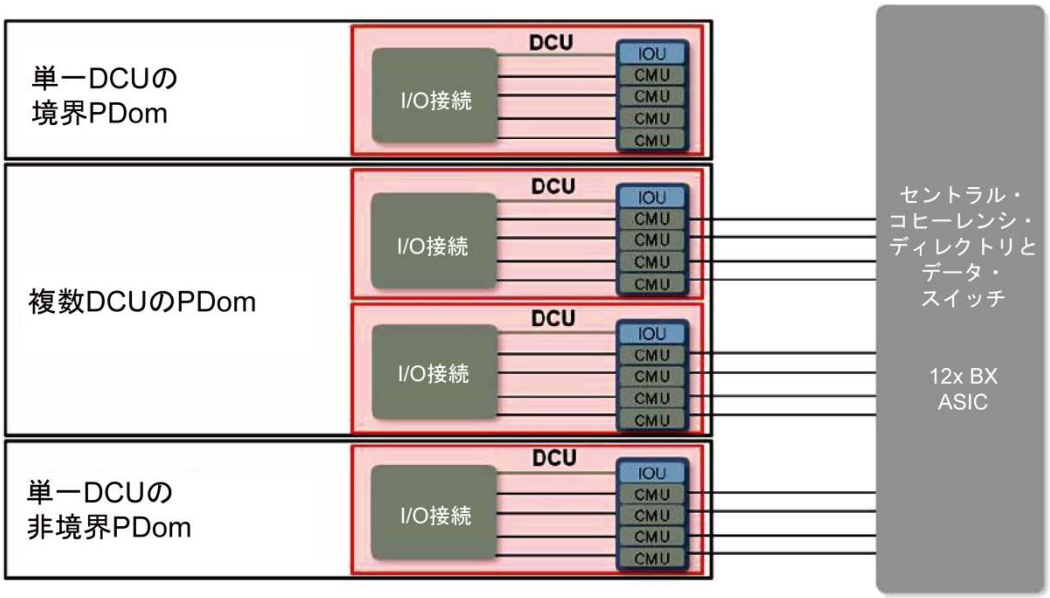

図4：単一DCUと複数DCUの物理ドメイン

PDomのサイジング：パフォーマンス、可用性、および柔軟性に対する影響

単一ドメイン構成ユニットのPDom

このPDomは、すべてのトラフィックが8ソケットのローカル・インターコネクトに分離されるため、最高のパフォーマンスと最高の可用性が得られます。

1つのプロセッサごとに7x153.6Gb/秒の専用ローカル・コヒーレンシ・リンクを使用して、すべての通信をローカル・コヒーレンシ・プレーン内で配信できるため、最大帯域幅と低レイテンシが実現します。

8ソケット、8TBのドメインであれば、今日のシステムで実行されるワークロードの大部分に対応できる十分な規模であり、このドメインのタイプ/サイズがSPARC M5-32ドメインまたはSPARC M6-32ドメインのデプロイの一般的な構成になります。

このタイプは、境界PDomの場合と非境界PDomの場合があります。境界PDomはスケーラビリティ・スイッチ・ボードから分離され、これがデフォルトのデプロイ・オプションです。

複数ドメイン構成ユニットのPDom

この構成では、1つのシステム・イメージで処理能力、メモリ、およびI/Oを大きくすることができます。PDomのこのモードで構成する場合は、すべてのメモリ・キャッシュ情報をスケーラビリティ・スイッチ・ボード（SSB）に渡して、DCU間のメモリ・トラフィックを有効にする必要があります。

ただし、複数のDCU間でのメモリ待機時間は1つのDCU内より大きくなるため、パフォーマンスに影響するという点を考慮してください。たとえば、8ソケットDCUの2つの独立したPDomと、16ソケットDCUの1つのPDomで多数のワークロードを実行する場合を考えてみます。16ソケットのPDomのほうが柔軟性は高く、PDom内の動的なリソース再割当ても向上しますが、8ソケットの2つのPDom間でワークロードを分散する場合と同じパフォーマンスを達成できるよう適切に動作するためには、Oracle Solarisでメモリのデプロイメントとスレッドのデプロイメントの最適化が必要になります。8ソケットの2つのPDomでは、すべてのメモリ・アクセスについて待機時間は最大222ナノ秒と保証されるのに対し、16ソケットのPDomではOracle Solarisがメモリのデプロイメントとスレッドのデプロイメントを確保すると想定されるものの、一部のメモリ・アクセスでは最悪の場合329ナノ秒の待機時間が発生する場合もあるからです。実際のパフォーマンスの差は、DCU間のメモリ呼出しが必要になる頻度によって異なります。

待機時間（ナノ秒）	オラクルのSPARC Enterprise M8000	オラクルのSPARC Enterprise M9000-32	SPARC M5-32 SPARC M6-32
CPUローカル	342	387	160
XBグループ/DCU内	402	447	222
キャビネット内	402	464	329

LDomを使用する場合は、DCUの境界を越えないようにLDomリソースを定義することによって、待機時間による影響を最小限に抑える（ほぼゼロにする）ことが可能です。

信頼性と可用性の観点から見ると、単一DCUドメインと複数DCUドメインとのハードウェア上の差は、SSBに依存します。SSBの平均故障間隔（MTBF）は、サーバーの想定ライフ・サイクルより長いので、リカバリ不能な障害から再起動につながる確率はほとんどありません。

保守性の観点から見ると、ドメインが大きくなるほど、そのドメインで計画保守が必要になったとき、再デプロイメントと中断が必要なワークロードも大きくなります。

Oracle SuperCluster M6-32のPDom

Oracle SuperCluster M6-32は、エンジニアド・システムとして使用できます。Oracle SuperCluster M6-32には固定のハードウェア構成セットが事前定義されており、特定のPDom構成を多数サポートすることが可能です。2個または4個のPDomを、シングルまたはデュアルのDCUから作成できます。この点については、付録：Oracle SuperCluster M6-32の構成ルールを参照してください。

Oracle VM Server for SPARC

Oracle VM Server for SPARCドメイン（論理ドメイン、またはLDomとも呼ばれる）は、個別の論理リソース・グループから構成される仮想マシンです。論理ドメインは、1つのコンピュータ・システム内で独自のオペレーティング・システムとIDを持ちます。論理ドメインはそれぞれ独立して作成、破棄、再構成、および再起動できるため、サーバーで電源の入れ直しが必要ありません。各論理ドメインで各種のアプリケーション・ソフトウェアを実行でき、パフォーマンスとセキュリティの目的でそのソフトウェアの独立を保つことができます。オラクルのソフトウェア・ライセンスの目的上、LDomはハード・パーティションとみなされます。

各論理ドメインで扱うことができるのは、ハイパーバイザによって利用可能になっているサーバー・リソースのみです。Logical Domain Managerを使用すると、制御ドメインを介してハイパーバイザを操作できます。そのため、ハイパーバイザはサーバーのリソースを強制的にパーティション化し、複数のオペレーティング・システム環境に一部のサブセットのみを提供します。このパーティショニングとプロビジョニングが、論理ドメインを作成する際の根本となるメカニズムです。次の図に、4つの論理ドメインをサポートするハイパーバイザを示します。この図では、論理ドメインの機能を構成する以下のレイヤーも示されています。

- ユーザー/サービス、またはアプリケーション
- カーネル、またはオペレーティング・システム
- ファームウェア、またはハイパーバイザ
- ハードウェア（CPU、メモリ、I/Oなど）

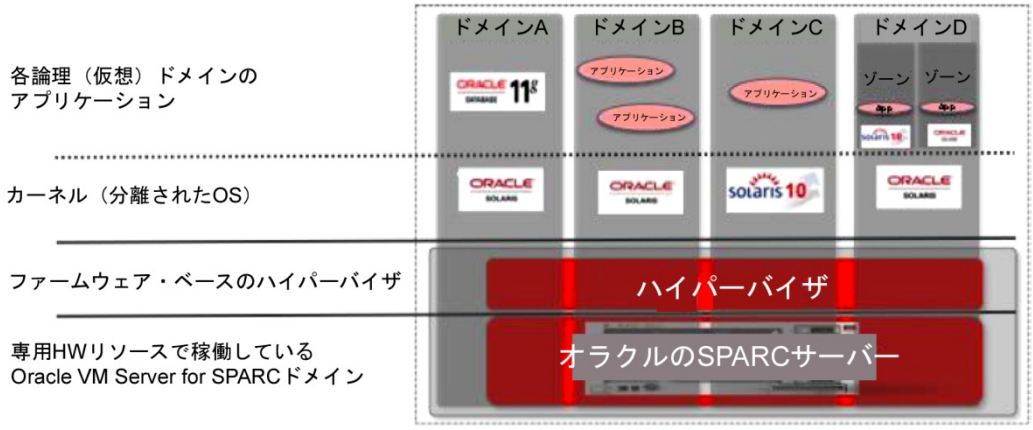

図5：Oracle VM Server for SPARCの仮想化

特定のSPARCハイパーバイザがサポートする各論理ドメインの数と機能は、サーバーによって異なります。ハイパーバイザは、CPU、メモリ、I/Oというサーバー・リソース全体のサブセットを特定の論理ドメインに割り当てることができます。そのため、複数のオペレーティング・システムがそれぞれ独自の論理ドメイン内で独自にサポートされます。リソースを個々の論理ドメイン間で再デプロイするときの粒度も任意です。たとえば、CPUは、CPUの1スレッドの粒度で論理ドメインに割り当てることができます。

各論理ドメインは、次の各リソースを独自に持つ完全に独立したマシンとして管理できます。

- カーネル、パッチ、およびチューニング・パラメータ
- ユーザー・アカウントと管理者
- ディスク
- ネットワーク・インタフェース、メディア・アクセス制御（MAC）アドレス、およびIPアドレス

論理ドメインはそれぞれ独立して停止、開始、および再起動されるため、サーバーで電源の入れ直しは必要ありません。

ダイナミック・ドメイン内のLDom

Oracle VM Server for SPARCは、物理的に分離されたドメインをさらに多くのドメインに分割する柔軟性を備えているため、ダイナミック・ドメインのみを使用する場合より多くのOSを独立して実行できます。

SPARC M5-32サーバーとSPARC M6-32サーバーは最大4つのダイナミック・ドメインをサポートするので、多くのデプロイメントでは必要に応じてOracle VM Server for SPARCテクノロジーを利用し、論理ドメイン・レベルでワークロードの分離をさらに進められると考えられます。

制御ドメイン

PDomを最初に作成するとき、プライマリ・ドメインまたは制御ドメインと呼ばれるドメインが作成されます。SPARC M5-32システムとSPARC M6-32システムでは、この初期ドメインでOracle Solaris 11.1以降を実行する必要があります。こうして作成される最初の制御ドメインは、全CPU、全メモリ、および全I/Oのリソースを含めてPDomで使用可能なすべてのハードウェアを所有します。

Oracle Solaris 11.1を実行するドメインが1つしか必要ない場合、その構成ではOracle VM Server for SPARCの機能が不要なため、これ以上の作業は必要ありません。このタイプの使用方法は、垂直方向に拡張されるきわめて大規模なワークロードで、CPUもメモリ・リソースも大量に必要な構成に適しています。それ以外の場合は、追加のドメインを作成し、必要に応じてI/O所有権をドメインに割り当てるようにOracle VM Server for SPARCを構成する必要があります。

I/O、ルート、サービス、およびゲストの各ドメイン

Oracle VM Server for SPARCのデプロイメントに存在する各種のドメインを表すために、いくつもの名前が使用されます。しかも、ドメインは同時に複数のタイプでありうるため、よけいに複雑です。たとえば、制御ドメインは常にI/Oドメインでもあり、通常はサービス・ドメインでもあります。このホワイト・ペーパーでは、各種のOracle VM Server for SPARCドメインを表すために次の用語を使用します。

制御ドメイン：サーバー仮想化の管理制御ポイント。ドメインの構成とリソースの管理に使用されます。電源を入れたときブートされる最初のドメインであり、通常はサービス・ドメインでもあります。制御ドメインは1つだけです。

I/Oドメイン：物理I/Oデバイス、すなわちPCIeルート・コンプレックス、PCIeデバイス、またはシングル・ルートI/O仮想化（SR-IOV）の機能が割り当てられています。所有するデバイスのパフォーマンスと機能をネイティブで持ち、仮想化レイヤーによって仲介されません。I/Oドメインは複数存在する場合もあります。

サービス・ドメイン：仮想ネットワークと仮想ディスク・デバイスをゲスト・ドメインに提供します。サービス・ドメインは複数存在する場合もあります。ゲスト・ドメイン用に仮想化するために物理I/Oリソースを所有する必要があるため、サービス・ドメインは常にI/Oドメインです。ほとんどの場合、サービス・ドメインにはPCIeルート・コンプレックスが割り当てられており、その場合はルート・ドメインと呼ぶことができます。

ゲスト・ドメイン：すべてが仮想デバイスで、物理デバイスを持たないドメイン。仮想ネットワークと仮想ディスク・デバイスは1つ以上のサービス・ドメインによって提供されます。一般的には、このドメインでアプリケーションが実行されます。1つのシステムには通常、複数のゲスト・ドメインがあります。

ゲスト・ルート・ドメイン：1つ以上のPCIeルート・コンプレックスが割り当てられているが、前述したサービス・ドメインのようにサービスを提供するのではなく、ドメイン内でアプリケーションを実行するために使用されるドメイン。

物理的には、サービス・ドメインとゲスト・ルート・ドメインとの間には使用方法しか差がないため、単にルート・ドメインと呼ばれることも少なくありません。SPARC M5-32サーバーおよびSPARC M6-32サーバー上のダイナミック・ドメインにおけるOracle VM Server for SPARCの構成は、従来のサーバー上での構成方法と何ら変わりません。追加のドメインを作成し、CPU、メモリ、およびI/Oをそのドメインに割り当てる際に制御ドメインが使用されます。CPUとメモリの割当ては比較的簡単

ですが、ドメインの使用目的と、I/Oをドメインに割り当てる方法は、ユースケースによって大きく異なります。

おおまかに言うと、Oracle VM Server for SPARCを実行する際には通常3つのモデルが使用されます。

モデル	説明	特徴	一般的なユースケース
単一の制御ドメイン/サービス・ドメイン	このモデルでは、制御ドメインがすべてのルート・コンプレックスを所有し、すべてのゲスト・ドメインに対して仮想デバイスを作成します。	もっとも柔軟性の高いモデルであり、比較的小さいドメインが多数あって障害の影響が少ないケースに適しています。制御ドメインが停止した場合には、すべてのゲスト・ドメインが影響を受けます。ゲスト・ドメインに対してはライブ・マイグレーションが可能です。	テスト環境や開発環境に理想的です。可用性が水平方向の拡張によって確保される軽量な本番環境にも適しています。
複数のサービス・ドメイン	1つ以上のサービス・ドメインが作成され、ルート・コンプレックスがそのサービス・ドメインに割り当てられます。ゲスト・ドメインのための冗長I/Oも構成可能です。	上のモデルと似ていますが、制御ドメインまたはサービス・ドメインで障害が発生してもゲスト・ドメインがそれほどは影響されない点が異なります。	より高い可用性が求められる本番環境に適しています。
ゲスト・ルート・ドメイン	ゲスト・ルート・ドメインの場合、ルート・コンプレックスがゲスト・ドメインに直接割り当てられ、そのI/Oを直接所有します。	複数の仮想ディスクと仮想ネットワーク・サービスを作成する必要がないため、もっとも単純なモデルですが、ルート・コンプレックスと同じ数のドメインしか持てないため、柔軟性ももっとも低くなります。ただし、これらのゲストはベアメタルのパフォーマンスで稼働し、相互に独立しています。	パフォーマンスの高い独立したドメインが必要とされる少数のドメインを持つ環境に理想的です。

これらの各種デプロイメント・モデルについては、各種のホワイト・ペーパーやWebキャストで詳細に説明されています。http://www.oracle.com/jp/technologies/virtualization/oracle-vm-server-for-sparc/resources/index.htmlを参照してください。

ゲスト・ルート・ドメイン

ここでは、ゲスト・ルート・ドメインについて詳細を説明します。SPARC M5-32ベースとSPARC M6-32ベースのシステムで想定されるワークロードは、この運用モデルに適しており、特にOracle SuperCluster M6-32はこの前提で構成されるからです。

ゲスト・ルート・ドメインとは、1つ以上のアプリケーションを直接ホストしており、サービス・ドメインに依存しないドメインを表す概念です。具体的には、ドメインI/O境界が1つ以上のPCIルートコンプレックスによって正確に定義されます。

そのため、Oracle VM Server for SPARCテクノロジーで使用可能な他のどのモデルと比べても大きい相違が多々あります。特にすべてのサービスをソフトウェア・ベースの仮想化を通じてゲストVMに提供する従来型の"Thick"モデルを利用する他のハイパーバイザに比べると、その利点は明らかです。

- パフォーマンス：すべてのI/Oがネイティブ（ベアメタル）であり、仮想化のオーバーヘッドがありません。
- 単純さ：ゲスト・ドメインとそれに関連するゲスト・オペレーティング・システムが、PCIeルート・コンプレックス全体を所有します。I/Oの仮想化が不要で、このタイプのドメインは構成がサービス・ドメイン・モデルより若干単純です。
- I/O障害の分離：ゲスト・ルート・ドメインはI/Oを他のドメインと共有しません。そのため、PCIeカード（NICまたはHBA）で障害が発生してもそのドメインしか影響を受けません。サービス・ドメイン・モデル、ダイレクトI/Oモデル、またはSR-IOVモデルでは、これらのコンポーネントを共有するすべてのドメインが障害の影響を受けるため、その点で対照的です。
- セキュリティの向上：コンポーネントや管理ポイントの共有が少なくなります。

1つの物理ドメインは、1個から4個のDCU（DCU0、DCU1、DCU2、DCU3）で構成されます。DCUごとに1つのI/Oユニット（IOU）が関連付けられ、各IOUは最大16基のPCIeスロット、8つの10 GbEポート、8台のHDD/SSD、4個のEMSモジュールをサポートします。

DCUあたりのルート・コンプレックスの数は16、DCUは最大4個なので、SPARC M5-32サーバーでもSPARC M6-32サーバーでも合計では64のルート・コンプレックスを使用できます。これらのルート・コンプレックスには、pci_0～pci_63という名前が付けられています。各ルート・コンプレックスには1基のPCIeスロットが関連付けられ、DCUあたり4つのルート・コンプレックスが、ローカル・ブートと10 GbEポートへのアクセスによって、EMSモジュールにアクセスします。

典型的な構成はゲスト・ルート・ドメインが16、ドメインあたりのPCIeスロットが4基であり、それぞれがブート時にはローカル・ディスクにアクセスします。Oracle SuperCluster M6-32には特定のドメイン構成ルールがあり、PDom内のドメインについて最適なサイズとレイアウトが定められています。その詳細は付録に記されています。これらのルールは、このプラットフォームにおけるゲスト・ルート・ドメインのレイアウトに関してオラクルが推奨するものなので、SPARC M5-32とSPARC M6-32のドメイン・レイアウトにも適用が可能です。

Oracle SuperCluster M6-32を使用しない場合、オプションはゲスト・ルート・ドメインだけではないということに注意してください。SPARC M5-32サーバーとSPARC M6-32サーバーの場合、サービス・ドメインを使用するシステムも、ゲスト・ルート・ドメインを使用するシステムも有効な場合があります。実際、同じダイナミック・ドメインを、アプリケーションが稼働する2つのルート・ドメインで構成し、2つのサービス・ドメインが完全に仮想化された多数のゲスト・ドメインにサービスを提供することも、異なるダイナミック・ドメインをまったく別個に構成することもできます。

ゾーン

Oracle Solarisには、Oracle Solaris Zonesと呼ばれる仮想化機能が組み込まれており、ソフトウェアによって定義される柔軟な境界に基づいてソフトウェア・アプリケーションとサービスを分離することができます。ハイパーバイザ・ベースの仮想化とは異なり、Oracle Solaris Zonesが提供するのはOSレベルの仮想化です。複数の物理マシンではなく複数のOSインスタンスのように見えます。Oracle Solaris Zonesを使用すると、オペレーティング・システムの単一インスタンスから多数のプライベート実行環境を作成でき、環境全体と個別のゾーンの完全なリソース管理が可能です。オラクルのソフトウェア・ライセンシングの目的上、制限された、または専用のCPUはハード・パーティションとみなされます。

OS仮想化の性質から、Oracle Solaris Zonesはオーバーヘッドと待機時間の非常に小さい環境を実現します。これによって、1つのシステムで数百または数千のゾーンを作成することが可能になります。Oracle Solaris ZFSとネットワーク仮想化が完全に統合されているため、他の仮想マシン実装環境では問題になりうる領域についても、実行と格納のオーバーヘッドが小さくなります。Oracle Solaris Zonesでは、ベアメタルに近いI/Oのパフォーマンスが得られるため、これらのソフトウェア・コンポーネントがきわめて高いI/Oパフォーマンスを発揮します。

Oracle Solaris 11には、完全に仮想化されたネットワーク・レイヤーがあります。データセンターのネットワーク・トポロジ全体を、仮想化されたネット、ルーター、ファイアウォール、およびNICを使用する単一のOSインスタンス内で作成できます。このように仮想化されたネットワーク・コンポーネントは、可観測性やセキュリティ、柔軟性、リソース管理性が高くなります。物理的なネットワーク・ハードウェアが一部不要になるため、コストを削減しながら柔軟性の向上を図ることができます。ネットワーク仮想化ソフトウェアはサービス品質にも対応し、おもなアプリケーションのために適切な帯域幅を確保できます。

Oracle Solaris Zonesでは、バージョンの古いOracle Solarisをゾーン内で実行することも可能で、これをブランド・ゾーンと呼びます。Oracle Solaris 10のグローバル・ゾーンを実行すると、その中でOracle Solaris 8やOracle Solaris 9を実行できます。これを利用すれば、レガシー・アプリケーションを新しいプラットフォームに簡単に統合できることになります。また、Oracle Solaris 11のグローバル・ゾーン上でOracle Solaris 10ゾーンを実行することによって、Oracle Solaris 10のワークロードがOracle Solaris 11のネットワーク仮想化機能を利用できます。

Oracle Solaris ZonesはOracle Solaris DTraceとも統合されます。DTraceは、動的な実装を可能にして、アプリケーションおよびカーネルのアクティビティをトレースできるOracle Solarisの機能です。管理者はこのDTraceを使用して、ソフトウェア・スタックを通じたJavaアプリケーションのパフォーマンスを調べることができます。Oracle Solaris Zones内部とグローバル・ゾーンにおける可視化が実現するため、管理者はボトルネックを特定、排除して簡単にパフォーマンスを最適化できます。

ユースケース

ここまでの各項で説明してきたように、仮想化の3つのレイヤーにはそれぞれ独自の機能があります。その組合せを変えると、アプリケーション・ワークロードの特定の要件に基づいて最適な組合せで柔軟性と分離性を得ることができます。

これのテクノロジーを組み合わせた何通りかの典型的なサンプルを、以下に示します。

構成の例

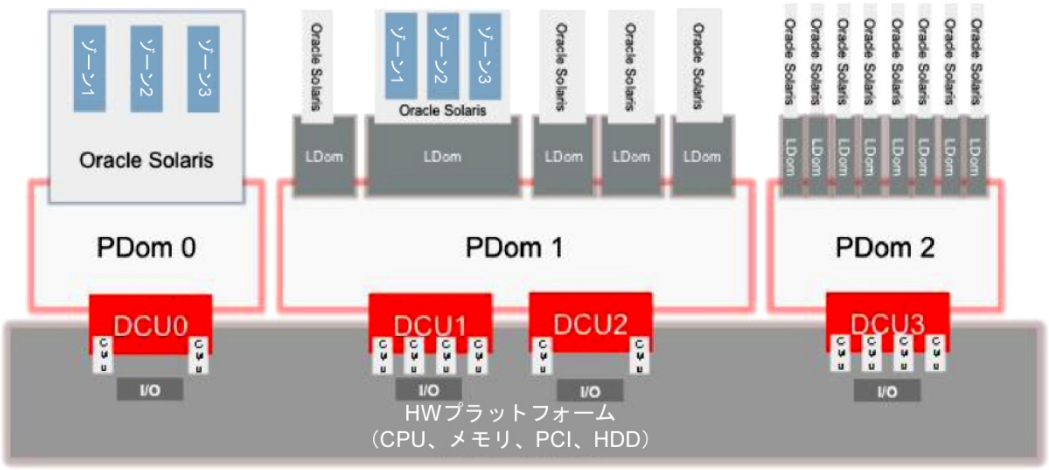

図6：複数の物理ドメインの構成オプション

上に示した例の場合、PDom 0ではOracle Solaris 11をベアメタルで直接実行しており、仮想化のオーバーヘッドが生じないため最大のパフォーマンスが得られます。ゾーンを使用すると、パフォーマンスを損ねることなくワークロードを分離できます。Oracle Solaris OSイメージは1つだけです。この実装では、スタックのもっとも高いレベルで統合が可能になります。

PDom 1では、サーバー仮想化のレイヤーとしてLDomが導入されています。仮想化とパフォーマンスの最適なバランスを維持するために、LDomはルート・ドメインとして構成され、各LDomがそれぞれ独自のPCIeスロットに直接アクセスします。LDomでは、ワークロードをさらにOSレベルにまで分離できます。各LDomには独自のOSインスタンスが必要なため、上のシナリオと比べると管理はやや複雑になります。各LDom内部では、Oracle Solaris Zonesを使用してさらにワークロードが分離されます。こうした実装によってハードウェア分離されたドメインが作成され、パフォーマンスをまったく損ねることなく同じハードウェア上でワークロードを分離できます。

PDom 2では、仮想I/OでLDomを使用し、最高の柔軟性を達成しようとしています。複数のI/Oサービス・ドメインがあり、これが仮想I/Oを他のLDomすべてにエクスポートします。このような編成の場合、ワークロードまたは環境ごとにLDomを作成できます。この実装は完全に仮想化された環境であり、独立分離した多数のドメインが独自のOSインスタンスを持つため、仮想化のオーバーヘッドが生じてパフォーマンスが低下します。LDomの数を減らしても同レベルの仮想化粒度を達成でき、Oracle Solaris Zonesによって作成される複数のゾーンを使用してスタックの統合度が高くなります。

もちろん、ここに挙げた3つの例は相互に排他的なものではなく、この3つをどのように組み合わせることも可能です。

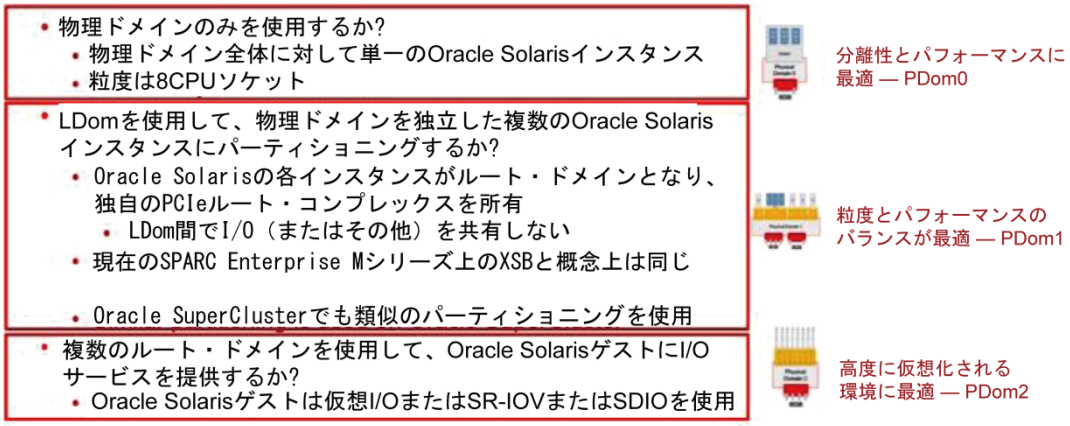

図7：推奨されるPDom構成のユースケース

I/Oの仮想化によって、柔軟性、動的な割当て、およびI/Oリソースの使用率は最大になりますが、場合によってはネイティブI/Oより待機時間が長くスループットが低くなります。ネイティブI/O特性が重要な場合には、ルート・ドメインの使用が適切です。使用可能なルート・ドメインの数は、ルート・コンプレックスの数によって、またI/Oカードで使用可能なPCIeスロットの数によって決まります。このような場合、SPARC M5-32サーバーとSPARC M6-32サーバーにOracle Virtual Networkingを使用して、ネットワークとFCトラフィックを1つのInfiniBandカードに集約すれば、SANとイーサネットのネットワーク接続両方に必要なPCIeスロットの数が半分になり、ソリューションとして効果的です。

結論

ここまででも、SPARC M5-32サーバーとSPARC M6-32サーバーが、旧世代のSPARC Enterprise Mシリーズにおける動的なシステム・ドメイン構成機能と、現行のSPARC Tシリーズ・システムにおけるOracle VM Server for SPARC（LDom）機能とを併せ持っていることは明らかです。これによって、Oracle Solaris Zonesでレイヤー化された仮想化ソリューションが実現し、多種多様な要件に対応できます。32TBのメモリ・フットプリントを持つSPARC M5-32サーバーとSPARC M6-32サーバーのサイズによって、数千のアクティブ・スレッドに対して最新の機能を提供されます。SPARC M5-32サーバーとSPARC M6-32サーバーは、可用性と保守性の両方を重視してプロセッサのコアから新たに設計されており、エンタープライズ・クラスのアプリケーションで新しいレベルの

パフォーマンス、可用性、そして使い安さが発揮されます。また、このサーバーは、ダイナミック・ドメイン、Oracle VM Server for SPARC、およびOracle Solaris Zonesによる高度なリソース制御により、企業がハードウェア資産の使用を最適化できるという点でも優れています。組織は、オラクルの高速でスケーラブルなSPARC M6-32サーバーを導入することで、スレッド単位の大幅なパフォーマンスと柔軟性の向上を実現できます。これは、ビジネスにおける競争戦略上、大きな利点になります。

SPARC M6-32の導入は、可用性の理由からも垂直方向のパフォーマンス関係の理由からもSPARCベースの小規模なシステムに適さない高負荷のワークロードを実行する際のニーズに基づいて決定されると考えられます。多くの場合、大規模なPDomを必要とする単一または複数のワークロードがあると考えられますが、小規模なPDomと複数のPDomへのデプロイメントが最適と考えられるワークロードが多数追加される可能性も高いでしょう。こうした柔軟性から、SPARC M6-32は理想的な統合プラットフォームとなっています。

SPARC M6-32クラスのマシンにワークロードをデプロイする際には、多くのユースケースを検討できます。一般的に、考慮するすべてのワークロードが8ソケットの構成要素に簡単に適合する場合には、8ソケットの境界PDomを作成するのがもっとも簡単な方法です。最高のパフォーマンスと分離性が得られる一方、その中で柔軟に割り当てられるドメイン数も十分であり、それでいて保守性が損なわれるほどには多すぎないからです。

SPARC M6-32を購入する理由が、8ソケットおよび8TB RAM以上を必要とする大規模な単一のOSインスタンス・イメージを実行することにある場合には、ワークロードに適したサイズで複数DCUのPDomを作成する必要があります。

粒度の高いワークロードが必要になるだけでなく、ダイナミック・ドメインの上に直接ゾーンを作成する方法でも、あるいはOracle VM Server for SPARCを追加して分離性を高める方法でも実現できます。

Oracle VM Server for SPARCドメインを使用する際には、ゲスト・ルート・ドメイン・モデルを使用してパフォーマンスのきわめて高い少数のドメインをデプロイする方法も、標準ゲスト・モデルを使用して小規模な多数のドメインをデプロイする方法もあります。どちらの場合でも、Oracle Solaris Zonesで作成したゾーンは同じようにOracle VM Server for SPARCドメインの上に重ねることができます。

どの場合でも、必要なレベルの分離性と保守性をもっとも簡単に達成できるモデルを選択するようにしてください。

可用性のベスト・プラクティス

このホワイト・ペーパーでは、高可用性（HA）については詳しく説明しませんでした。しかし、HAは、SPARC M5-32デプロイメントまたはSPARC M6-32のデプロイメントのアーキテクチャを決定する上で非常に重要な要素です。

コンピュート・ノード間やリモート・レプリケーション間でディザスタ・リカバリを目的としてクラスタ化するなど、高可用性のベスト・プラクティスを適用する場合は常にビジネス要件を考慮する必要があります。たとえば、同じPDom内にデプロイしたクラスタの両方のノードにHAを実装してはなりません。一方、複数の層（Web、アプリケーション、データベース）を異なるドメインにデプロイし、共通のクラスタ化テクノロジーまたは水平方向の冗長性を利用して他のノードに複製することはできます。

個々のワークロードごとにこれらの概念の詳細を論じたホワイト・ペーパーを、オラクルは多数発行しています。Oracle Technology Network（OTN）で、Maximum Availability ArchitectureとOptimized Solutionsのセクションを参照してください。

まとめ

ごく単純に言えば、SPARC M5-32ベースとSPARC M6-32ベースのデプロイメントには、以下のような概要のガイドラインが当てはまります。

- 8ソケット以上のドメインを必要とする具体的な要件がない限り、デフォルトで8ソケットのダイナミック・ドメインを使用する。8ソケット・ドメインを最大4つまで構成できる。
- さらに分離が必要な場合には、ダイナミック・ドメイン内でOracle VM Server for SPARCを使用する。少数の大規模なドメインに対してはルート・ドメイン・モデルを使用し、多数の小規模なドメインに対しては仮想化したI/Oゲスト・モデルを使用する。
- どんな場合でも、Oracle Solaris Zonesを使用してアプリケーションをドメイン内にカプセル化する。柔軟で動的なリソース制御とセキュリティ分離のためにはゾーンを使用する。
- 高可用性を構成するには、アプリケーションレベルで水平方向の拡張性を、またはアプリケーションレベルのクラスタ化を利用するか、クラスタ化製品を使用してゾーンまたはドメインのレベルでワークロードをクラスタ化する。

Elite Engineering Exchangeについて

Elite Engineering Exchange（EEE）は、部門を越えたグローバル組織で、オラクルのエリート・セールス・コンサルタント（SC）とシステム・エンジニア（Product Engineering）とで構成されています。EEEは、共同コラボレーション、顧客との双方向コミュニケーション、市場トレンド、そして新世代製品のテクノロジーの方向性に対する深い洞察を通じて、Product Engineeringと各分野の専門家トップとを直接つなぎます。EEEは、実世界の顧客体験をエンジニアリングに、エンジニアリングにおける技術的な情報と洞察をSCに直接結び付け、その双方によって、オラクルのお客様のご要望の変化に対応できるソリューションをお届けします。

付録：Oracle SuperCluster M6-32の構成ルール

Oracle SuperCluster M6-32はエンジニアド・システムであり、9台のExadata Storage Serverと3台のInfiniBandスイッチで構成されるOracle Exadata Storage Expansion Half Rackと、SPARC M6-32サーバーが組み合わされています。これを補足するのが、ラックに設置されたOracle ZFS Storage Applianceです。

SuperCluster構成の特徴をすべて詳細に説明することは、このホワイト・ペーパーの対象範囲を超えていますが、ここではOracle SuperCluster M6-32のドメイン（PDomとLDom）構成オプションについておもな特徴を説明しておきます。

エンジニアド・システムであるOracle SuperCluster M6-32は、限られた数で固定の構成を提供するように設計されており、それによってパフォーマンス、スケーラビリティ、および可用性の観点でベスト・プラクティスを提示しようとしています。言い換えれば、SPARC M6-32サーバーは多様な形の構成が可能であるものの、SuperClusterを構成できる種類は限定されているということです。そのため、インストール・ベースを通じて厳格にテストされた構成の一貫性が保証され、万一問題が発生した場合に迅速な対応と分析が可能になります。

Oracle SuperCluster M6-32ドメインの構成要素

SuperClusterは2つまたは4つのPDomで構成され、PDomは事前定義された多くの組合せで固定構成DCUを組み合わせて作成されます。次に、Oracle VM Server for SPARCテクノロジーを使用してPDomあたり最大4つのルート・ドメインLDomを作成します。

- 各DCUは、多数の固定コンポーネントで構成されます。
- 10GbEポート×2 を具備した ベースI/Oカード×4
- HDD×8（各900GB）
- デュアル・ポートInfiniBand HCA×4
- クアッド・ポート1GbE NIC×1
- 空きPCIeスロット×11

DCUの可変要素はDCU内のCMUの数で、次のうちのいずれかです。

- 2CMU（SPARC M6プロセッサ×4）、または
- 4CMU（SPARC M6プロセッサ×8）

最後にDCUには、SPARC M6プロセッサごとに、容量16GBまたは32GB32枚のデュアル・インライン・メモリ・モジュール（DIMM）をすべて装着する必要があります。これで、プロセッサあたりのRAMは512GBまたは1TBになります。DIMM容量の選択肢は、Oracle SuperCluster M6-32内のすべてのDCUに適用されることに注意してください。

PDom構成：ベースと拡張

DCUが2つのみの場合、システム構成のオプションは単一DCUの2PDomだけです。単一DCUのPDomは通常の構成アプローチとみなされ、"ベース構成"のPDomと呼ばれます。

一方、DCUが4つある場合、2DCUずつの2PDomか、単一DCUの4PDomかいずれかの構成オプションがあります。これを"拡張構成"のPDomと呼びますが、その原理は単純です。8ソケット以上のCPUを割り当てた1つのLDomが必要か、単一DCUで得られる以上のI/O容量を複数のLDomで必要とするかということです。

ほとんどの場合は、単一DCUのPDomを使用するものと想定されます。

最後に、同じラック内で共存させるのではなく、SPARC M6-32のラック間でDCUを分割することができます。このオプションは、単一ラックのソリューションよりRASが若干改善されるため、きわめて高い可用性または追加の分離が必要な場合のためだけに用意されています。

LDom構成：1個から4個のLDom

ドメイン構成ルールは、PDomに対するLDomの割当て方法を決定し、通常または拡張の各PDomレイアウトについて4つの構成を定義します。LDomは常にルート・ドメインとして構成され、各ドメインが1つ以上のルート・コンプレックスの所有権を排他的に付与されます。これによって、LDomのそれぞれに対してベアメタル・パフォーマンスが保証されます。

通常のPDomの場合、次の4つの構成があります。

1つのLDomで、それにすべてのリソースを割り当てる（大規模な1LDom）

2つのLDomで、I/Oリソースをその2者間で均等に分割する（中規模な2LDom）

3つのLDomで、1つが大規模、2つが小規模（中規模な1LDom、小規模な2LDom）

4つのLDomで、I/Oリソースを均等に分割する（小規模な4LDom）

拡張PDomの場合、1つのLDomが巨大であるという前提になります。構成は上の場合と同じですが、最初のLDomにも最初のDCUのI/Oがすべて割り当てられる点が異なります。

Oracle SuperCluster M6-32についての結論

以上で見てきたように、Oracle SuperCluster M6-32の構成ルールは、ルート・ドメイン・モデルを利用するドメイン構成の指針となるベスト・プラクティスとして最適な例です。と同時に、Oracle SuperCluster M6-32またはM5-32の構成を伴わない構成のブループリントとしても利用できます。

ORACLE

オラクルのSPARC M5-32サーバーと
SPARC M6-32サーバー：ドメイン構成の
ベスト・プラクティス
2013年10月
著者：Michael Ramchand、Tom Atwood、
Michele Lombardi、Henning Henningsen、
Martien Ouwens、Ray Urciuoli、Roman Zajcew

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

Hardware and Software, Engineered to Work Together