



Oracleテクニカル・ホワイト・ペーパー
2013年9月

SPARC M6-32サーバーのアーキテクチャ

概要	2
SPARC M6-32サーバーの機能の概要.....	3
SPARC M6-32サーバーのコンポーネント	4
システム・アーキテクチャ.....	6
システム・コンポーネントの概要	6
システム・インターコネクットの信頼性機能.....	8
SPARC M6プロセッサ.....	9
Oracle Solarisによるマルチコアのスケラビリティ.....	11
I/Oサブシステム.....	12
完全な仮想化スペクトラム.....	14
ドメイン構成ユニット.....	15
ダイナミック・ドメイン.....	16
マルチスレッド・ハイパーバイザによってサポートされる論理ドメイン	17
Oracle VM Server for SPARC.....	18
Oracle Solaris Zones	19
信頼性、可用性、および保守性.....	21
高度な信頼性機能.....	21
エラー検出、診断、およびリカバリ	21
冗長/ホットスワップ対応コンポーネント	22
システム管理テクノロジー.....	23
Oracle ILOMとサービス・プロセッサ	23

電源管理.....	25
Oracle Enterprise Manager Ops Centerを使用したSPARC M6-32サーバー の管理.....	26
Oracle Solaris 11オペレーティング・システム.....	27
結論.....	29
追加情報.....	30
付録A:SPARC M6プロセッサのアーキテクチャ.....	31
SPARC M6プロセッサのキャッシュ・アーキテクチャ.....	32
SPARC M6コア・アーキテクチャ.....	33

概要

組織のテクノロジーへの依存度は、かつてないほど高くなっています。現在コンピューティング・システムは、製品の設計や顧客オーダーの実行など、あらゆる機能で重要な役割を果たしています。多くの場合、ビジネスの成功はITサービスの継続的な可用性によって決まります。メインフレーム・クラスの信頼性と保守性は、以前は一部のデータセンターでのみ必要でしたが、今や企業全体のシステムで不可欠となっています。また、データセンター・サーバーの強化と停電時のサービス継続も重要な関心事です。

可用性はもっとも重要ですが、コストを予算内に抑えたり、操作のしやすさを維持したりすることも必要です。組織はネットワーク化されたサービスをできるだけ効率的かつ経済的に提供するため、統合/仮想化戦略によってあらゆるIT資産を最大限に利用しようとしています。このため現在のITシステムの要件は、単なる処理能力という尺度を超えたものになっています。サーバーの利用率を最適化するための組み込みの仮想化機能および関連ツール、テクノロジー、プロセスとともに、柔軟性の高いサーバーが必要です。新しいコンピューティング・インフラストラクチャによって、テクノロジーやトレーニングへの現在の投資を保護することも必要です。

オラクルのSPARC M6-32サーバーは、従来のメインフレームの多くの利点を生かした、信頼性が高く管理しやすい垂直方向にスケーラブルなシステムです。また、関連コストや複雑さ、ベンダーの囲い込みなどの問題もありません。実際にこのサーバーによって、オープン・システムの価格でメインフレーム・クラスのシステム・アーキテクチャを提供できます。SPARC M6-32サーバーは、1~32基のプロセッサ、32TBのメモリ・サブシステム、高スループットのI/Oアーキテクチャによる対称型のマルチプロセッシング（SMP）スケーラビリティにより、統合されたワークロードで必要とされる高負荷作業を簡単に実行できます。また、最先端の仮想化テクノロジーを含む強力なOracle Solaris 10およびOracle Solaris 11オペレーティング・システムが実行されます。SPARC M6-32サーバーはダイナミック・ドメイン、Oracle VM Server for SPARC、動的再構成、およびOracle Solaris Zonesテクノロジーに対応しており、オープン・システム・コンピューティング・プラットフォームで高度なメインフレーム・クラスのリソース制御を実行できます。

SPARC M6-32サーバーの機能の概要

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。高速で待機時間の短いシステム・インターコネクトでメモリとI/Oに接続されている最新世代のSPARCプロセッサが使用されており、アプリケーションへのスループットが非常に優れています。また、計画的/計画外の停止時間を減らすように設計されており、停止を回避しリカバリ時間を短縮するための優れた信頼性/可用性/保守性（RAS）機能が搭載されています。このサーバーでは、高度なCPUの統合、データ・パスの整合性、Extended-ECC（エラー修正コード）メモリ、エンド・ツー・エンドのデータ保護、ホットスワップ対応のコンポーネント、耐障害性のある電源オプション、ハードウェアの冗長性などの設計機能により、信頼性が大幅に向上しています。

また、SPARC M6-32サーバーは構成の柔軟性が非常に優れています。オラクルの他のハイエンド・サーバーと同様、管理者はダイナミック・ドメイン（別名、物理ドメインまたはPDom）を使用して、1台のSPARC M6-32サーバーを電氣的に独立した複数のパーティションに物理的に分割できます（パーティションごとにOracle Solarisの独立したインスタンスが実行されます）。あるPDomでハードウェア/ソフトウェア障害が発生しても、他のPDomで実行されるアプリケーションには影響しません。PDomごとに、それ自体のOracle VM Server for SPARCのコピーを実行できます（ハイパーバイザベースの仮想化テクノロジー）。この環境では、論理ドメインが作成され、Oracle Solarisの複数のインスタンスを1つのPDomで実行できます。このように、Oracle Solaris Zonesによって物理ドメインと論理ドメインを混在させることができるため、さまざまな仮想化テクノロジーを複雑に組み合わせることができます。

動的再構成によって、重要なシステムを中断することなく、物理ドメイン内で、論理ドメイン間のハードウェア・リソースや仮想リソースの再割当てを行うことができます。SPARC M6-32サーバーの特性については、表1を参照してください。

表1：SPARC M6-32サーバーの特性

SPARC M6-32サーバー					
エンクロージャ	キャビネット×1				
SPARCプロセッサ	<table border="0"> <thead> <tr> <th>M6</th> <th>M5</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> 3.6GHz (48MBのレベル3 (L3) キャッシュ) 最大32基のオラクルの12コア SPARC M6プロセッサ スレッド×8 (コアあたり) SPARC M6 プロセッサ×2 (CPUメモリ・ユニット (CMU) あたり) </td> <td> <ul style="list-style-type: none"> 3.6GHz (48MB L3キャッシュ) 最大32基のオラクルの6コア SPARC M5プロセッサ スレッド×8 (コアあたり) SPARC M5プロセッサ×2 (CMUあたり) </td> </tr> </tbody> </table>	M6	M5	<ul style="list-style-type: none"> 3.6GHz (48MBのレベル3 (L3) キャッシュ) 最大32基のオラクルの12コア SPARC M6プロセッサ スレッド×8 (コアあたり) SPARC M6 プロセッサ×2 (CPUメモリ・ユニット (CMU) あたり) 	<ul style="list-style-type: none"> 3.6GHz (48MB L3キャッシュ) 最大32基のオラクルの6コア SPARC M5プロセッサ スレッド×8 (コアあたり) SPARC M5プロセッサ×2 (CMUあたり)
M6	M5				
<ul style="list-style-type: none"> 3.6GHz (48MBのレベル3 (L3) キャッシュ) 最大32基のオラクルの12コア SPARC M6プロセッサ スレッド×8 (コアあたり) SPARC M6 プロセッサ×2 (CPUメモリ・ユニット (CMU) あたり) 	<ul style="list-style-type: none"> 3.6GHz (48MB L3キャッシュ) 最大32基のオラクルの6コア SPARC M5プロセッサ スレッド×8 (コアあたり) SPARC M5プロセッサ×2 (CMUあたり) 				
メモリ	<ul style="list-style-type: none"> 最大32TB DIMMスロット×64 (CMUあたり) 16GBと32GBのDIMMをサポート 				
内部I/Oスロット	<ul style="list-style-type: none"> 最大64枚のx8 PCIe Generation 3 (Gen3) カード ホットプラグ可能なキャリア上のロープロファイル・カード 				
内部ストレージ	<ul style="list-style-type: none"> 最大32台のSerial Attached SCSI (SAS) ドライブ 				
ダイナミック・ドメイン	<ul style="list-style-type: none"> 最大4つの物理ドメイン 				
論理ドメイン	<ul style="list-style-type: none"> 最大512のゲスト 				

SPARC M6-32サーバーのコンポーネント

SPARC M6-32サーバーはエンタープライズ・システム・キャビネットにマウントされ、最大16台のCPUメモリ・ユニット (CMU) と4台のI/Oユニット (IOU) をサポートします。完全構成の場合、SPARC M6-32サーバーには32基のプロセッサ・チップ、32TBのメモリ、64個のPCI Express (PCIe) 内部ショート・スロットが搭載され、4つの物理ドメインに分割できます。各物理ドメインは、最大128の論理ドメインをサポートできます。また、SPARC M6-32サーバーは最大32台のディスク・ドライブをサポートします。SPARC M6-32サーバーでは、12個の電源と18台のファン・ユニットにより電源供給と冷却が行われます。SPARC M6-32サーバーの前面図と背面図については、図1と図2を参照してください。

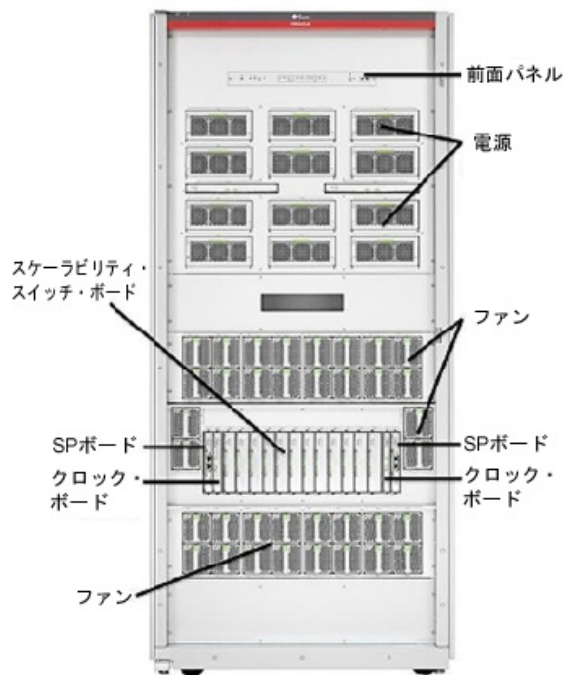


図1：SPARC M6-32サーバー・エンクロージャ（前面）

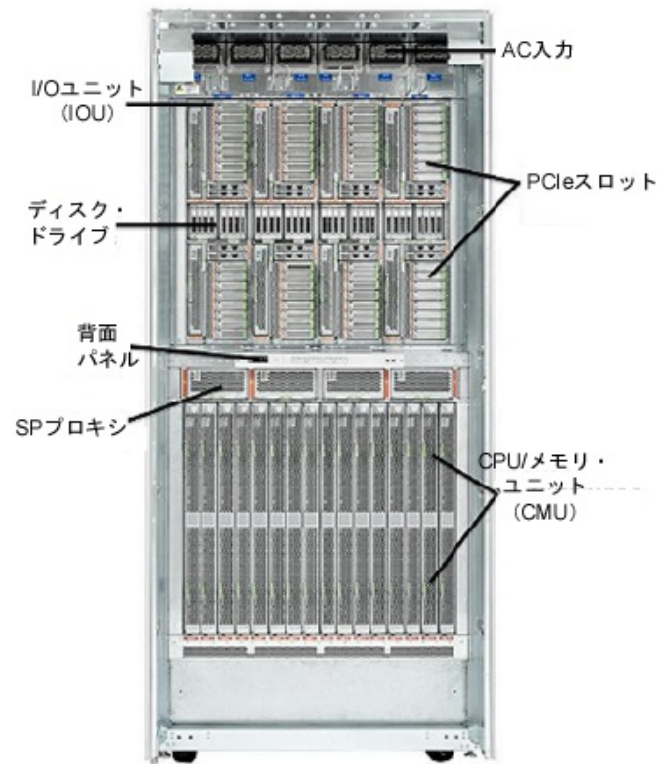


図2：SPARC M6-32サーバー・エンクロージャ（背面）

システム・アーキテクチャ

IT組織は、ワークロードの増加と効率性の実現に苦慮してきた結果、少数の強力なシステムで処理要件に対応した方が経済的なメリットが大きいということを理解するようになりました。SPARC M6-32サーバーでは、インターコネクト、プロセッサ、メモリ・サブシステム、I/Oサブシステムと一緒に機能することで、サーバー統合のニーズに対応したスケーラブルな高性能プラットフォームを実現しています。組織はこのサーバーを利用することで複数のプロジェクトを1つのプラットフォームにロードし、低コストでアプリケーションの実行を高速化できます。

システム・コンポーネントの概要

SPARC M6-32サーバーは、特に高い信頼性、優れたパフォーマンス、および真のSMPスケーラビリティの実現を重視して設計されています。このサーバー内のあらゆるサブシステムの特性と機能は、これを目指としています。このサーバー内では高帯域幅のシステム・バス、強力なSPARC M6プロセッサ、高密度なメモリ・オプション、および高速なPCIe拡張スロットの組合せにより、エンタープライズ・アプリケーションの高レベルなアップタイムとスループット、および信頼性の高いスケーリングが実現されています。

システム・インターコネクト

システム・インターコネクトにより、SPARC M6-32サーバーの高レベルなパフォーマンス、スケーラビリティ、信頼性がさらに向上します。SPARC M6-32サーバー内のスケーラビリティ・スイッチにより、CPU、メモリ、およびI/Oサブシステム間のポイント・ツー・ポイントの接続が可能です。コンポーネント間に複数のバス・ルートがあることでパフォーマンスが向上し、スイッチに障害が発生してもシステムの運用を継続できます。実際、これらのサーバーで使用されるシステム・インターコネクトによって、ピーク帯域幅が3,072GB/秒にもなり、オラクルの前世代のハイエンド・サーバーよりシステム・スループットが実質的に向上しています。

メモリ

SPARC M6-32サーバーのメモリ・サブシステムでは、スケーラビリティとスループットが向上しています。SPARC M6-32サーバーでは、DDR3 DIMMテクノロジーが使用されています。単一のシステム・ボード内で複数のDIMMサイズはサポートされませんが、DIMM容量はシステム・ボードによってさまざまです。使用可能なDIMMサイズは16GBと32GBです。メモリ・サブシステムの詳細については、表2を参照してください。

表2：メモリ・サブシステムの仕様

SPARC M6-32サーバー	
最大メモリ容量	32TB
DIMMスロット	SPARC M6プロセッサあたり32個
CMUあたりの増加	CMUあたり16、32、または64枚のDIMM

SPARC M6-32サーバーのメモリ・サブシステムは、パフォーマンスはもちろん、信頼性も重視して構築されています。メイン・メモリに保存されるデータ用にECC保護が実装されており、次の高度な機能により早期診断と障害分離が可能です。このためシステムの整合性を維持し、アプリケーションの可用性を上げることができます。

- **メモリ・パトロール** - メモリ・パトロールにより、メモリのエラーが定期的にスキャンされます。この事前予防的な機能により、障害のあるメモリ領域によってシステム・エラーやアプリケーション・エラーが発生する前にその使用を防ぎ、システムの信頼性を上げることができます。
- **メモリのExtended-ECC** - これらのサーバーのメモリのExtended-ECC機能によって1ビット・エラーの修正が可能となり、（メモリ・デバイスの障害により発生しうるバースト読取りエラーなどの）イベントに関係なく継続的な処理がサポートされます。この機能はIBMのChipkillテクノロジーと似ています。
- **メモリ・レーン・スペアリング** - SPARC M6プロセッサのメモリ・コントローラとDIMMの間のメモリ・レーンは、データ・パス内で1つのレーンが喪失しても引き続き機能します。

システム・クロック

SPARC M6-32サーバーは、信頼性を重視して設計されています。特に、クロック・ボードは内部の冗長コンポーネントを使用して構築されています。また、すべてのSPARC M6-32サーバーには冗長クロック・ボードが搭載されています。可用性が大幅に向上し、メンテナンスもずっと簡単になっており、あるボードで障害が発生しても他のボードを使用してシステムを再起動できます。

PCIeテクノロジー

SPARC M6-32サーバーでは、PCIeバスを使用してI/Oサブシステム内での高速なデータ転送が可能です。PCIeテクノロジーでは、データ転送の最大速度が元のPCIテクノロジーの2倍で、スループットが8GT/秒にもなります。実際に、PCIeは高速インターコネクト（ファイバ・チャンネル、InfiniBand、ギガビット・イーサネットなど）に対応できるように開発されました。SPARC M6-32サーバーにはロープロファイル（LP）キャリアが搭載されており、IOUに対してPCIeカードを個別に（ホットプラグ対応で）抜き差しできます。SPARC M6-32サーバーには、64個のLP PCIe Gen3スロットがあります。

電力と冷却

SPARC M6-32サーバーでは、電力と冷却に別のモジュールを使用しています。システム全体に設置されたセンサーによってプロセッサと主要なASICの温度、および複数個所の周辺温度が測定されます。電源/冷却サブシステムのハードウェア冗長性と環境監視を組み合わせることで、電源やファンに障害が発生してもサーバーの運用を継続できます。

ファン・ユニット

SPARC M6-32サーバーでは、完全に冗長化されたホットスワップ対応のファンがプライマリ冷却システムとして機能します。1つのファンに障害が発生すると、SPによって障害が検出され、残りのファンを高速動作に切り替えることで、エアフローの減少分を補完します。SPARC M6-32サーバーはこのような状況でも通常どおり運用できるため、障害の発生したユニットの保守を十分に時間をかけて行うことができます。アプリケーションの処理を中断せずにファン・ユニットを交換できます。

電源

表3のとおり、冗長電源/電源コードを使用することでSPARC M6-32サーバーのフォルト・リジリエンスが向上します。ホットスワップ対応の冗長電源によって電力が供給されるため、電源に障害が発生してもサーバーは引き続き運用できます。電源ユニットはホットスワップ対応のため、システムを実行しながら取り外しや交換を行うことができます。

SPARC M6-32サーバーでは三相電源が使用されます。三相電源によって電力のデュアル・フィードが可能となります。すべてのSPARC M6-32サーバーに、6本の電源コードがすべて付属します。

表3：SPARC M6-32サーバーの電源および冷却仕様

SPARC M6-32サーバー	
ファン・ユニット	<ul style="list-style-type: none"> ファン・ユニット×36 N+1冗長
電源	<ul style="list-style-type: none"> 7,000ワット ユニット×12 N+1冗長 三相
電源コード	<ul style="list-style-type: none"> 電源ケーブル×3 (シングル・フィード) 電源ケーブル×6 (デュアル・フィード)

オペレータ・パネル

SPARC M6-32サーバーにはオペレータ・パネルが搭載されています。このパネルを使用してサーバー・ステータスの表示、サーバーの識別およびユーザー設定情報の保存、運用/保守モードの切替え、ドメインの電源投入を行います（図3）。サーバーの起動時には、前面パネルのLEDステータス・インジケータによって、SPとサーバーの稼働状況が分かります。



図3：SPARC M6-32サーバーのオペレータ・パネル

システム・インターコネクットの信頼性機能

システム・インターコネクットには冗長性/信頼性機能が組み込まれているため、SPARC M6-32サーバーの安定性が向上しています。インターコネクットによってシステム・バス上のデータが完全にECC保護されるため、損失や破損を防ぐことができます。CPUやI/Oコントローラで1ビット・データ・エラーが検出されると、ハードウェアによってデータが修正され転送が実行されます。SPARC M6-32サーバーには、デグレード可能なスケラビリティ・スイッチとバス・ルートがあります。インターコネクット内でまれにハードウェア障害が発生すると、再起動時に残りのスケラビリティ・スイッチが使用され、障害のあるスイッチが分離されて簡単に動作を再開できます。コヒーレンシ・リンクごとにレーンのスペアリングがサポートされているため、特定のリンクでのレーンの喪失を防ぐことができます。

SPARC M6プロセッサ

SPARC M6プロセッサは、SPARC M5プロセッサの後継製品です。SPARC M5プロセッサは6コアとして設計されましたが、SPARC M6プロセッサは12のコアを提供します。SPARC M6プロセッサはコンピューティング、セキュリティ、I/Oが1つのチップに統合された統合性の高いチップであるため、高い費用をかけてハードウェアとソフトウェアをカスタム開発する必要がなくなります。以前のSPARCプロセッサとのバイナリ互換性に加えて、他のプロセッサと比べて少ないスペースと電力要件で高いパフォーマンスを実現できます。このため組織は、最大限の効率性と予測可能性で新しいネットワーク・サービスを迅速に展開できます。SPARC M6プロセッサについては、図4を参照してください。



図4：組織はSPARC M6プロセッサのおかげで、最大限の効率性と予測可能性で新しいネットワーク・サービスと計算集中型のワークロードを迅速に展開できます。

オラクルの次世代のマルチコア/マルチスレッド・プロセッサの設計時に、社内の設計チームがまず念頭に置いていたのは次の目標です。

- オラクルのSPARC64 VII+プロセッサより、スループット演算機能を大幅に上げる（同レベルのパフォーマンスを必要とするワークロード向け）。
- ネットワーク負荷の大きいワークロード用に、ネットワーク・パフォーマンスを上げる。
- エンド・ツー・エンドのデータセンター暗号化のパフォーマンスを大幅に上げ、ハードウェア内に新しい暗号実装を追加する。
- サービス・レベルを上げ、計画的/計画外停止時間を減らす。
- コストを削減しながらデータセンターの機能を上げる。

オラクルのマルチコア/マルチスレッド・アーキテクチャは非常に柔軟性が高いため、プロセッサ、コア、統合コンポーネントのさまざまなモジュールを組み合わせることができます。SPARC M6プロセッサは、SPARC T4、T5、M5プロセッサで使用されているものと同じオラクルのS3コア・アーキテクチャを利用しています。

SPARC M6プロセッサの設計では、メモリ待機時間がパフォーマンス向上の真のボトルネックであることが認識されています。このプロセッサでは、各プロセッサ内のコアを再設計し、新しい浮動小数点パイプラインを設計し、ネットワーク帯域幅をさらに増やすことで、SPARC64 VII+プロセッサの約6倍のスループットを実現しています。

各SPARC M6プロセッサには12個のコアが搭載されており、改善されたLRU（Least Recently Used）アルゴリズムによって、コアごとに最大8つのスレッド（プロセッサあたり96スレッド）を切り替えてスレッドを選択できます。さらに、各コアには2つの整数実行パイプラインがあるため、1個のSPARCコアで同時に2つのスレッドを実行できます。SPARC M6プロセッサはSPARC64 VII+プロセッサとは異なり、8つのスレッドのうちの1つをフェッチして、命令をパイプラインの各段階に伝達していき、フェッチの3段階で選択段階に渡します。

スレッド命令は、2つのデコード・グループに分類され、デコード、リネーム、ピックの段階を経て発行段階に進んだ後、実行する命令の種類に応じて、後続の4つの実行パイプラインの1つに送られます。

ここまでは、スレッドからのいずれの命令も、命令の種類に関係なくパイプラインを経由します。サイクルあたりの発行段階ごとに、サイクルあたり2つの実行命令が発行されます。

12コアのSPARC M6プロセッサでサポートされるスレッド・モデルの簡単な概要説明については、図5を参照してください。

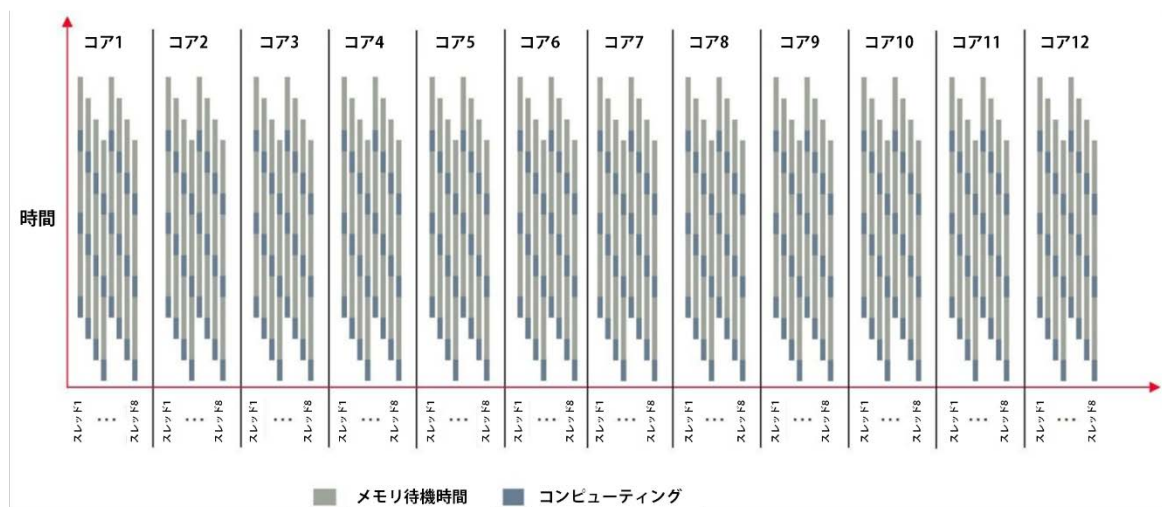


図5：12コアのSPARC M6プロセッサ1基で最大96のスレッドをサポートします。各コアで同時に実行できるスレッド数は最大2つです。

SPARC M6プロセッサは、提供するコア数が2倍である点を除けば、事実上SPARC M5プロセッサと同じです。SPARC M5が提供するコア数は6ですが、SPARC M6は12のコアを提供します。SPARC M6-32サーバーはSPARC M6プロセッサとSPARC M5プロセッサの両方をサポートします。これらのプロセッサを組み合わせる方法については、ドメインの項で説明します。

Oracle Solarisによるマルチコアのスケラビリティ

Oracle Solaris 10および11リリースは、特にSPARC M6-32システムの多くのリソースを完全に利用できるように設計されています。つまりOracle Solarisには、垂直/水平方向にスケラリングされる環境で、仮想化、最適な使用、高可用性、卓越したセキュリティ、および優れたパフォーマンスを実現するための主要な機能が搭載されています。Oracle Solarisは、広範なSPARCおよびx86ベースのシステムで動作し、既存のアプリケーションとの互換性が保証されています。SPARC M6プロセッサに基づくシステムのもっとも魅力的な機能の1つは、Oracle Solarisおよびそのサポート対象のアプリケーションに対し、このシステムが使い慣れたSMPシステムとして示されることです。さらに、Oracle Solarisリリース10と11には、オラクルのマルチコア/マルチスレッド・アーキテクチャ上のアプリケーションのパフォーマンスを改善する数多くの機能が搭載されています。

- **高速暗号化** - Oracle Solaris (初期リリース) の暗号化フレームワークとSPARC M6プロセッサにより、高速暗号化がサポートされます。SPARC M6プロセッサでは、暗号化ハードウェアの実装にアクセスできません。今回初めて、ユーザー・レベルの命令によって、コプロセッサではなく適切なパイプライン自体に暗号が実装されます。つまり、ハードウェアベースの暗号をより効率的に実装でき、特権レベルでの変更が不要になるため、暗号化アルゴリズムの計算が大幅に効率化されます。また、命令パイプライン自体に実装された各種暗号を、データベース運用ではるかに効率的に利用できます。
- **クリティカル・スレッドの最適化** - Oracle Solarisリリース10および11では、ユーザーまたはプログラマーがOracle Solarisスケジューラにクリティカル・スレッドを認識させることができます (このためには、CLIまたは関数のシステム・コールによって優先順位を60以上に上げます)。この操作を実行すると、スレッドが自動的にシングル・コアで実行され、このコアのすべてのリソースがコア自体に集められます。利用可能なCPUより多くの実行可能スレッドがある場合は、このシングル・スレッドはシングル・コアでは実行されません。他のスレッドのリソースが不足しないように、この制限が設けられています。Oracle Solarisでは、クリティカル・スレッドの最適化のさらなる強化が予定されています。
- **マルチコア/マルチスレッドの認識** - Oracle Solarisリリース10および11ではSPARC M6プロセッサの階層が認識されるため、スケジューラによって、使用可能なすべてのパイプラインで効果的に負荷分散できます。これらの各プロセッサを96個の論理プロセッサとして使用する場合も、Oracle Solarisではサポートするコアおよびスレッド間の相関関係を把握し、スレッドを迅速かつ効率的に実装できるようにします。
- **きめ細かな管理** - Oracle Solarisリリース10および11には、SPARC M6プロセッサの個別のコアとスレッド (論理プロセッサ) を有効または無効にする機能が搭載されています。また、Oracle Solarisの標準機能 (プロセッサ・セットなど) によって、論理プロセッサのグループを定義し、このグループに対してプロセスやスレッドをスケジューリングできます。
- **インタフェースのバインディング** - Oracle Solarisでは、必要に応じて、プロセスや個々のスレッドをプロセッサやプロセッサ・セットに非常に柔軟にバインドできます。
- **仮想化ネットワークおよびI/Oのサポート** - Oracle Solarisには、SPARC M6プロセッサ上のコンポーネントとサブシステムをサポート/仮想化するためのテクノロジーが含まれます (オンチップのPCIeインタフェースのサポートなど)。高パフォーマンス・ネットワーク・アーキテクチャの一部として、オラクルのマルチコア/マルチスレッド対応のデバイス・ドライバが提供されているため、仮想化フレームワーク内で実行されるアプリケーションがI/Oデバイスやネットワーク・デバイスを効率的に共有できます。

- **Oracle SolarisにおけるNon-Uniform Memory Accessの最適化** - メモリは各SPARC M6プロセッサによって管理されるため、これらの実装はNon-Uniform Memory Access (NUMA) アーキテクチャに相当します。NUMAアーキテクチャの場合、プロセッサが自身のメモリへのアクセスに要する時間は、別のプロセッサによって管理されるメモリへのアクセス時間よりやや短くなります。Oracle Solarisには、特にアプリケーションへのNUMAの影響を減らし、NUMAアーキテクチャのパフォーマンスを改善できる次のテクノロジーが搭載されています。
 - **メモリ配置の最適化 (MPO)** - Oracle Solarisでは、MPOによってサーバーの物理メモリ全体のメモリ配置を改善することで、パフォーマンスが向上します。MPOにより、システム内の適切なバランスを維持しながら、メモリが、そのメモリにアクセスするプロセッサにできる限り近い位置に配置されます。結果として、MPOにより、多数のデータベース・アプリケーションの実行速度が大幅に向上します。
 - **Hierarchical Lgroup Support (HLS)** - HLSによって、複雑なメモリ待機時間の階層を持つシステムのパフォーマンスが最適化され、Oracle SolarisのMPO機能が改善されます。Oracle Solarisでは、HLSによってメモリとの距離を識別され、アプリケーションの待機時間が最短になるようにリソースが割り当てられます。特定のアプリケーションがデフォルトでローカル・リソースを使用できない場合、Oracle Solarisでは、HLSによって、もっとも近くのリモート・リソースが割り当てられます。
- **Oracle Solaris ZFS** - Oracle Solaris ZFSでは、世界唯一の128ビットのファイル・システムによってデータ管理を劇的に進歩させることで、複雑なストレージ管理の概念を自動化および統合し、無限のスケラビリティを実現します。Oracle Solaris ZFSは、I/Oの発行順序に関する従来の制約をほぼすべて解消するトランザクション・オブジェクト・モデルに基づいており、パフォーマンスが大幅に向上します。また、発見されにくいデータ破損の検出と修正を行う64ビットのチェックサムですべてのデータが保護され、データの整合性が維持されます。
- **セキュアで堅牢なエンタープライズ・クラス的环境** - Oracle Solarisには恣意的な犠牲は必要ありません。既存のSPARCアプリケーションを変更せずに、引き続きSPARC M6プラットフォーム上で実行できるため、ソフトウェアへの投資が保護されます。また、マルチレベルの認証セキュリティによって、Oracle Solaris環境が侵入から保護されます。Oracle Solarisの障害管理アーキテクチャによって、Oracle Solarisの予測的自己修復などの要素がハードウェアと直接通信できるため、計画停止時間と計画外停止時間の両方が短縮されます。Oracle Solaris DTraceなどの効果的なツールにより、システムのリソースを最大限に利用できるようにアプリケーションを調整できます。

I/Oサブシステム

強力なI/Oサブシステムは、今日の大規模なデータセットの効率的な移動/操作に不可欠です。SPARC M6-32サーバーはI/Oの拡張とパフォーマンスに非常に優れているため、組織はシステムを迅速に拡張して進化するデータ・ストレージのニーズに対応できます。

PCIeテクノロジーの使用は、SPARC M6-32サーバー内のI/Oサブシステムのパフォーマンスにとって非常に重要です。PCIeブリッジによって、I/Oユニットのメイン・システムとコンポーネント間（PCIeスロットと内部ドライブなど）を接続できます。

サーバーでは、PCIeアダプタ・カードを簡単にホットプラグするためにPCIeカセットを使用します。PCIeカードはPCIeホットプラグをサポートしており、管理者が実行中のサーバーのPCIeカセットにマウントしたり、内部PCIeスロットに挿入したりすることができます。

SPARC M6-32サーバーには、4台のI/Oユニットがあります。各IOUには16個のPCIe Gen3 x8スロットと最大8台のドライブ、および最大4枚のI/Oカードが搭載されています。各IOUは特定のドメイン構成ユニット（DCU）専用です。これについては、ドメインの項で説明します。シングルIOUのコンポーネントについては、図6を参照してください。

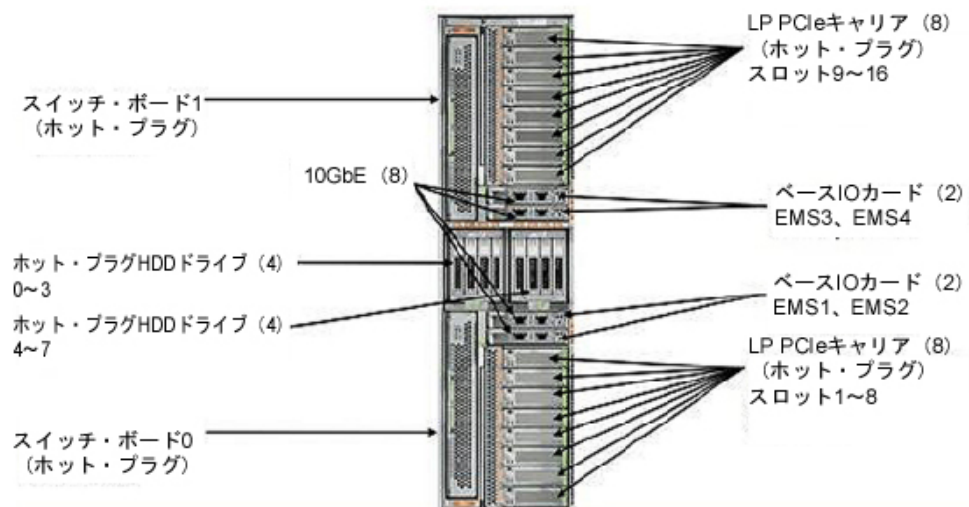


図6：IOUのコンポーネント

各PCIeスロットにはキャリアが付いており、PCIeカードを個別にホットプラグできます。スイッチ・ボード0によって、PCIeスロット1~8とベースI/Oカード1および2が制御されます。スイッチ・ボード1によって、PCIeスロット9~16とベースI/Oカード3および4が制御されます。ディスク・ドライブはデュアル・ポートのため、ディスクは、ペア1と2のベースI/Oカード、およびペア3と4の2番目のベースI/Oカードによって制御されます。

物理ドメインのブート時には、SPARC M6プロセッサによってPCIeスロットが制御されます。フル・ロードのDCUのSPARC M6プロセッサとPCIeスロットのマッピングについては、図7を参照してください。

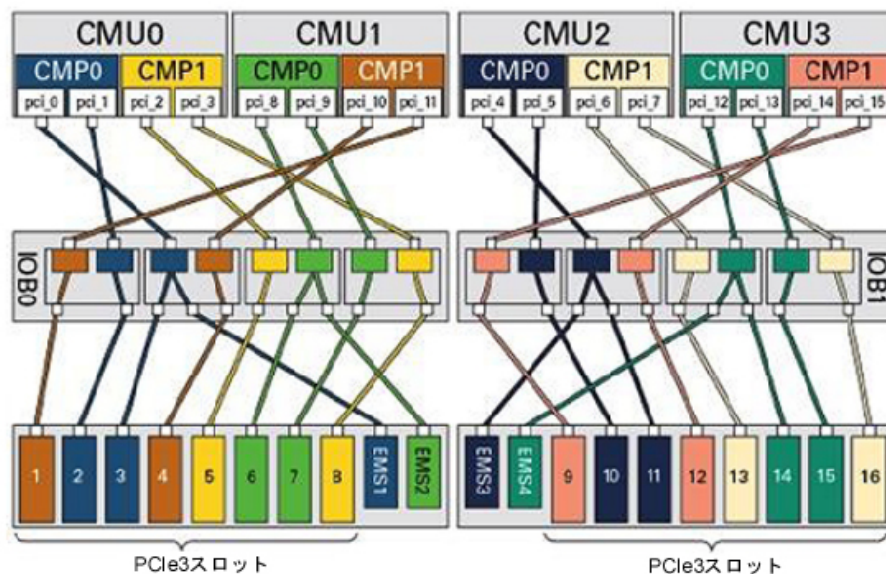


図7：SPARC M6-32サーバーのI/Oサブシステム（DCUあたり）

完全な仮想化スペクトラム

組織では、使用率が増加する中で各種ワークロードを少数の強力なシステムに統合しようとしているため、仮想化テクノロジーへの注目が高まっています。SPARC M6-32サーバーは特に仮想化向けに設計されており、3つのレベルの仮想化テクノロジー（ダイナミック・ドメイン（物理ドメインなど）、論理ドメイン、およびOracle Solaris Zones（OSの仮想化））がすべて搭載されています。

- ダイナミック・ドメインは、1つの大規模システムを複数の障害分離サーバーに分割するために使用されます。
- Oracle VM Server for SPARCを使用して作成された論理ドメインは、ダイナミック・ドメインの仮想化に使用されます。これにより、複数の仮想マシン（VM）をホストし、その仮想マシンごとにOracle Solarisの自身のインスタンスを実行できます。
- Oracle Solaris ZonesによってOSを仮想化できます。これにより、Oracle Solarisのシングル・インスタンスによって、アプリケーションを互いに安全に分離し、特定のシステム・リソースを各ゾーンに割り当てることもできます。

もっとも重要なのは、オラクルの仮想化テクノロジーが、高価なアドオンとしてではなく、システムの一部として提供されている点です。SPARC M6-32サーバーのすべての仮想化テクノロジーの例については、図8を参照してください。

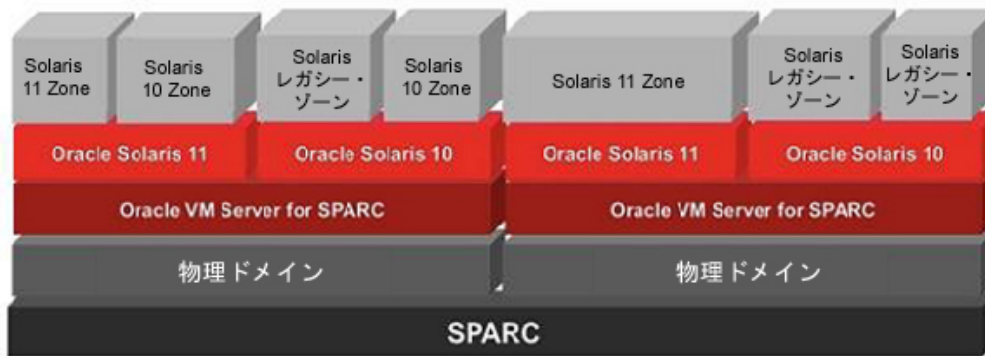


図8：SPARC M6-32サーバーの仮想化テクノロジー・スタック

ドメイン構成ユニット

各ドメイン構成ユニットは、4台のCPUメモリ・ユニット、1台のI/Oユニット、およびサービス・プロセッサ・プロキシ（SPP）ボードをグループ化したものです。DCUは、物理ドメインの構成に使用される構成要素です。DCU内のSPARC M6プロセッサは、すべて相互に直接的に通信します。異なるDCU内のSPARC M6プロセッサは、スケーラビリティ・スイッチ・ボード（SSB）を使用して相互に通信する必要があります。DCU0のレイアウトについては、図9を参照してください。

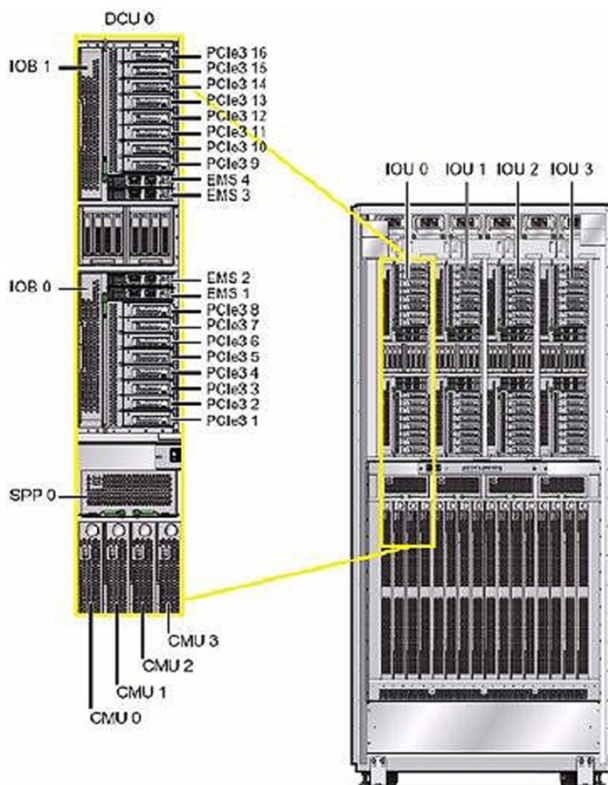


図9：DCU0のレイアウト

SPARC M6-32サーバー内で使用可能な4台のDCUのレイアウトについては、図10を参照してください。

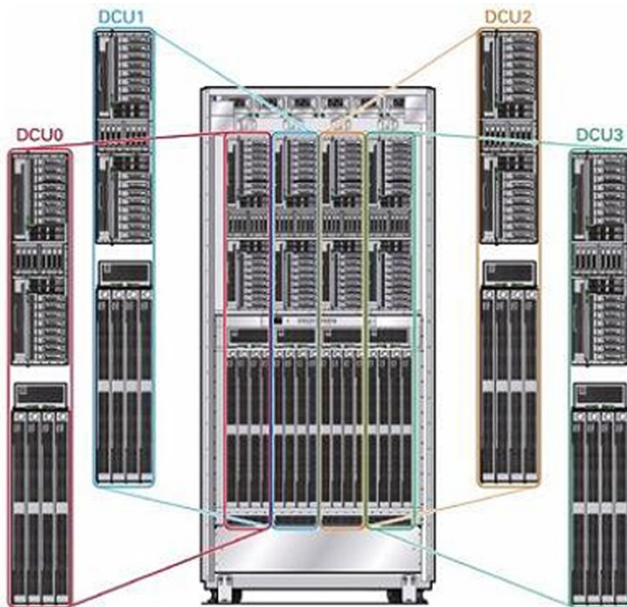


図10：SPARC M6-32サーバー内の4台のすべてのDCU

SPARC M6プロセッサとSPARC M5プロセッサを同じSPARC M6-32サーバーで組み合わせて使用する場合、CMUボードがDCU内にグループ化されている点に留意することが重要です。DCUに実装できるのは同じタイプのプロセッサのみです。つまり、SPARC M5プロセッサを使用したCMUボードがDCU内に2つのみ含まれる場合、DCU内の空きスロットに実装できるのはSPARC M5ベースのCMUボードのみになります。SPARC M6ベースのCMUボードは別のDCUに実装する必要があります。ただし、ダイナミック・ドメイン（物理ドメインまたはPDom）を定義する場合は、SPARC M5ベースのDCUとSPARC M6ベースのDCUを同じPDom内に組み合わせることが可能です。

ダイナミック・ドメイン

コストと管理上の負担を減らすため、多くの組織がサーバーを統合しようとしています。ただし、1台のサーバーで複数のアプリケーションをホスティングする際のセキュリティと効率を上げるためのツールが必要です。SPARC M6-32サーバーのダイナミック・ドメイン（PDomなど）によって、IT組織は1つの大規模システムを複数の障害分離サーバーに分割し、それぞれのサーバーでOracle Solarisオペレーティング・システムの個別のインスタンスを実行できます。適切に構成されていれば、あるドメインでのハードウェア/ソフトウェア障害は分離されたままとなり、他のドメインの動作に影響することはありません。1つのサーバー・プラットフォーム内の各ドメインをOracle Solarisの別のバージョンで実行することもできるため、このテクノロジーは新規または変更したアプリケーションの本番前のテストに非常に便利です。PDomの最大数は4です。

PDomには、標準PDomと境界PDomの2種類があります。標準PDomは4～32基のプロセッサに拡大でき、DCU全体を組み合わせる構成されます。境界PDomは1台のDCUに限定されます。そのため、4基または8基のSPARC M6プロセッサしか含まれません。標準PDomの利点は、すべてのCMU/IOUボードを含めて拡張できることです。境界PDomの利点は、待機時間が少なくSSBの障害の影響を受けないことです。

標準PDomは、任意のDCUを組み合わせることで構成できます。DCU0～DCU3を組み合わせた1つのドメインとなる場合もあります。あるドメインのDCU0とDCU1、および別のドメインのDCU2とDCU3を組み合わせ、2つのPDomとなる場合もあります。DCUが隣接する必要はありません。DCU0とDCU3でPDomを構成することもできます。同様の組み合わせで、3つのPDomを構成することもできます。

図11は、異なる数のDCUが割り当てられた2つのPDomです。PDom 0には1台、PDom 1には2台のDCUが割り当てられています。



図11：2つのPDom

マルチスレッド・ハイパーバイザによってサポートされる論理ドメイン

前の世代のSPARCプロセッサと同様、SPARC M6プロセッサにはマルチスレッド・ハイパーバイザが搭載されており、1つのPDom内に論理ドメインを作成できます。ハイパーバイザは、プロセッサと緊密に統合された安定した仮想マシン・アーキテクチャを提供する小さなファームウェア・レイヤーです。ハイパーバイザは基盤となるマルチコア/マルチスレッド・プロセッサと直接やり取りするため、マルチスレッドは非常に重要です。このアーキテクチャは、シングル・コア内で複数のスレッドをコンテキスト切替えることができます。競合他社のアーキテクチャでこのようなタスクを行うと、追加のソフトウェアと多大なオーバーヘッドが必要となります。

図12のように、仮想化テクノロジーの対応するレイヤーは、ハイパーバイザ上に構築されます。オラクルのアプローチが優れているのは、完全にスレッド化されたJavaアプリケーション・モデルを使用するプロセッサからアプリケーションまで、アーキテクチャのすべてのレイヤーが完全にマルチスレッド化されていることです。これは決して新しいテクノロジーではなく、Oracle Solarisは1992年からマルチスレッドをサポートしてきました。この機能はあらゆるレベルで、すべてのOracle Solarisサービスに組み込まれています。

オラクルは、プロセッサとハイパーバイザの他、完全にマルチスレッド化されたネットワークとOracle Solaris ZFSファイル・システムを提供しています。Oracle VM Server for SPARC（旧称Sun Logical DomainsまたはLDM）、Oracle Solaris Zones、およびマルチスレッド・アプリケーションは、必要なリソースを的確に受け取ることができます。

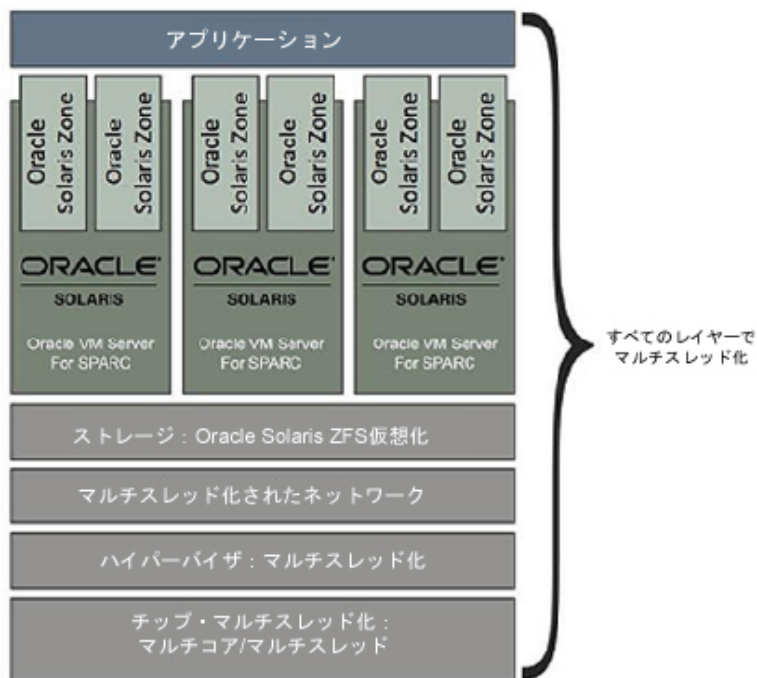


図12：オラクルは、テクノロジー・スタックのあらゆるレベルで並列化と仮想化を提供しています。

Oracle VM Server for SPARC

Oracle VM Server for SPARCは、オラクルのマルチコア/マルチスレッド・テクノロジーを使用するすべてのOracleサーバーでサポートされており、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ます。各オペレーティング・システム・インスタンスには、仮想化された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。Oracle VM Server for SPARCアーキテクチャでは、Oracle Solaris 10などのオペレーティング・システムがハイパーバイザに書き込まれます。この処理により、各ドメインのオペレーティング・システムには、基盤となるサーバー・ハードウェアの安定しており最適化された仮想化可能な内容が反映されます。各ドメインは完全に分離されており、1つのプラットフォーム上に作成される仮想マシンの最大数は、（システムにインストールされている物理的なハードウェア・デバイス数ではなく）ハイパーバイザの機能によって決まります。たとえば、1台のDCUに4基のSPARC M6プロセッサが搭載されているSPARC M6-32サーバーでは、最大128個のドメインがサポートされ、個別ドメインごとに独自のOSインスタンスを実行できます。

Oracle VM Server for SPARC 3.0では、ドメイン間のライブ・マイグレーションを実行できます。ライブ・マイグレーションという用語は、ソース・ドメインやアプリケーションの停止が不要であるということの意味します。Oracle VM Server for SPARC 3.0では、実行中のアプリケーションをドメイン間で移行できるようになりました。このため、SPARC M6-32サーバー上の論理ドメインを、同じサーバー上の他のPDom、別のSPARC M6-32サーバー、またはオラクルのSPARC T5/SPARC M5ベースのサーバーにライブ・マイグレーションできます。

組織はドメインを利用することで、複数のオペレーティング・システム・インスタンスを同時に1つのプラットフォーム上に柔軟に実装できます。また、管理者は仮想デバイス機能を利用することで、論理ドメイン上でホストされているソフトウェア・スタック全体を物理マシン間で転送できます。また、ドメインはOracle Solaris Zonesをホストすることで、両方のテクノロジーの分離、柔軟性、および管理の機能を得られます。Oracle Solarisは、Oracle VM Server for SPARCとSPARC M6プロセッサの緊密な統合により、柔軟性を高め、ワークロードの処理を分離し、サーバーの使用率の最大化を促進しています。

各PDom内で作成される論理ドメインについては、図13を参照してください。

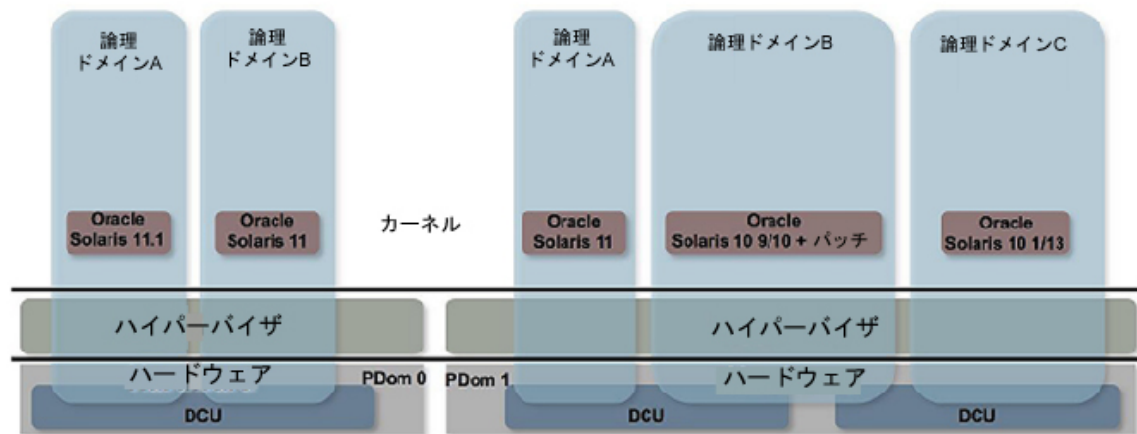


図13：各PDom内の論理ドメイン

Oracle Solaris Zones

Oracle Solaris 11には、Oracle Solaris Zonesという独自のパーティション化テクノロジーが搭載されています。これは、Oracle Solaris 10ではOracle Solaris Containersと呼ばれていたものです。このテクノロジーを使用して、アプリケーション実行用の分離されたセキュアな環境を作成できます。Oracle Solaris Zoneは、1つのOracle Solarisインスタンス内に作成された仮想オペレーティング・システム環境です。Oracle Solaris Zonesを使用すると、アプリケーションとプロセスをシステムの残りの部分から切り離すことができます。このように分離することで、Oracle Solaris Zone内のプロセスが別のOracle Solaris Zoneで実行されているプロセスによって妨げられることがなくなるため、セキュリティと信頼性が向上します。

マルチプロセッサ・システム内のCPU（またはSPARC M6プロセッサ内のスレッド）はプロセッサ・セットとして論理的にパーティション化して、リソース・プールにバインドできます。最終的には、そのリソース・プールをOracle Solaris Zoneに割り当てることができます。リソース・プールには、CPUリソースの消費が重複することを避けるためのワークロード分離機能があります。また、クラス割当てのスケジューリングとプロセッサ・セットに対する永続的な構成メカニズムも存在します。さらに、リソース・プールの動的な機能を使用すると、管理者はワークロード要求の変化に応じてシステム・リソースを調整できます。

Oracle Solaris Zonesテクノロジーは、古い環境を新しいプラットフォームに統合するための優れたツールです。これにより、アプリケーションは最新のCPU、メモリ、およびI/Oテクノロジーによるパフォーマンス向上のメリットを享受できます。また、より高いRASレベルのシステムにアプリケーションを構築できます。標準的な統合シナリオについては、図14を参照してください。

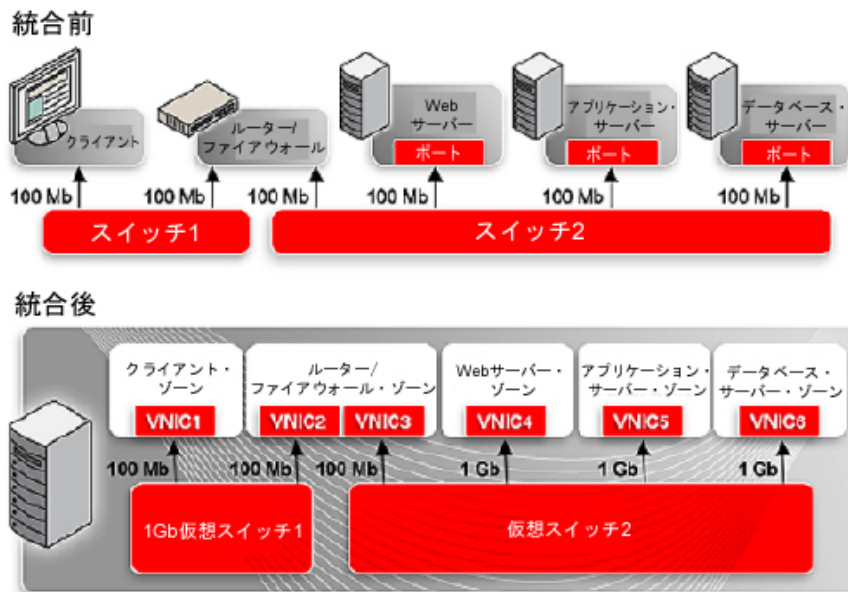


図14：サーバーとOracle Solaris Zonesを統合する利点

複数の仮想化テクノロジーを組み合わせることで、TCOの削減とビジネス目標を達成できます。たとえば、複数のOracle Solaris Zonesを各論理ドメイン内で実行できます。図15では、PDomごとに独自のハイパーバイザがあり、論理ドメインをサポートしています。論理ドメインごとに、異なるOracle Solarisリリースが分かれています。これは、最新OSの実行コストが異なる場合に一般的に行われます。次の手順として、アプリケーションをそれぞれのOracle Solaris Zone内で分離します。これにより、リソース割当てとプロセス分離をきめ細かく実行できます。Oracle Solaris Legacy Containersは、Oracle Solarisリリース8または9のみで認定されているアプリケーションのために使用します。管理者は新しいSPARC M6-32のパフォーマンスと機能を活用でき、ビジネス上の変更が必要とならないかぎり、すべてのソフトウェアのアップグレードを強制されることはありません。

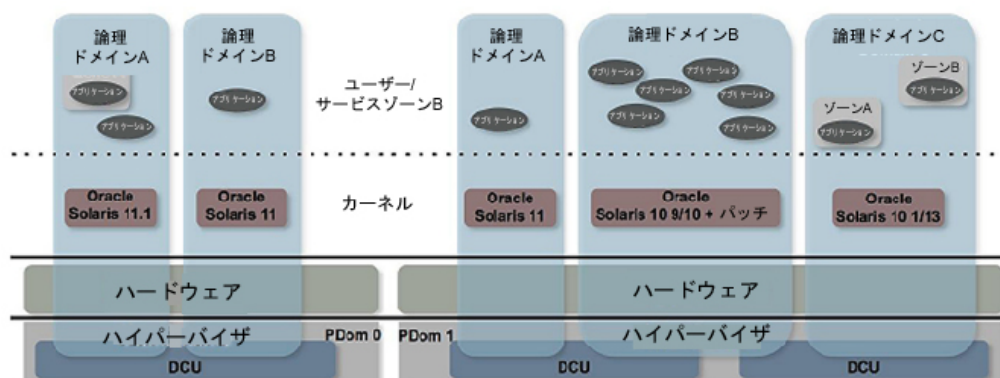


図15：SPARC M6-32サーバーの複数の仮想化テクノロジー

SPARC M6-32サーバーは、2つのPDomで構成されます。1つのPDomは1台のDCU、もう1つのPDomは2台のDCUで構成されます。1つめのPDomでハードウェアの問題が発生しても、2つめのPDomには影響しません。すべてのPDom間で、障害がハードウェア的/電氣的に完全に分離されているためです。

信頼性、可用性、および保守性

計画的/計画外の停止時間の短縮は、ITサービスにとって重要です。システム設計には、主要サービスの可用性に影響せずにフォルト・リジリエンス、すばやい修復、および迅速な拡張を可能にするメカニズムが含まれる必要があります。SPARC M6-32サーバーは、複雑なネットワーク・コンピューティング・ソリューションと厳しい高可用性（HA）要件をサポートできるよう特別に設計されており、冗長/ホットスワップ対応のシステム・コンポーネント、設計全体の診断/エラー・リカバリ機能、および組込みのリモート管理機能が含まれています。この信頼性の高いサーバーの高度なアーキテクチャにより、高レベルなアプリケーションの可用性と各種ハードウェア障害からの高速リカバリ、および企業のシステム運用の簡素化とコストの削減を実現できます。

高度な信頼性機能

SPARC M6-32サーバーのコンポーネント内に含まれる高度な信頼性機能により、プラットフォームの全体的な安定性が向上します。たとえばSPARC M6-32サーバーには、システム・バス内の冗長性を確保するために、複数のシステム・コントローラとデグレード可能なクロスバー・スイッチが含まれています。サーバー・アーキテクチャ内のコンポーネント数を減らして簡素化することで、信頼性が向上します。また、高度なCPU統合と確実なデータ・パスの統合により、SPARC M6プロセッサによる自律型のエラー・リカバリが可能となり、修正処理の開始時間の短縮によりアップタイムが長くなります。

Oracle Solarisの予測的自己修復ソフトウェアによって、SPARCサーバーの信頼性がさらに向上します。SPARC M6-32サーバーにはOracle Solarisの予測的自己修復ソフトウェアが実装されており、CPUとメモリが常に監視されます。エラーの性質によっては、スレッド、コア、またはCPU全体を自動的にオフラインにすることで、永続的なCPUソフト・エラーを解決できます。また、メモリ・ページ・リタイアメント機能によって、特定のDIMMメモリのデータが複数修正されると、メモリ・ページが事前予防的にオフラインになる機能がサポートされています。

エラー検出、診断、およびリカバリ

SPARC M6-32サーバーには、障害を早期に修正し、少数のコンポーネントによって停止時間が繰り返し発生しないようにする重要なテクノロジーが搭載されています。本質的な信頼性の向上につながるアーキテクチャの進歩は、サーバーのハードウェア・サブシステム内のエラー検出機能やリカバリ機能によって強化されます。つまり、次の機能が一緒に機能することでアプリケーションの可用性が向上します。

- エンド・ツー・エンドのデータ保護により、システム全体のエラーが検出/修正され、データの完全な整合性が維持されます。
- 最先端の障害分離によって、サーバーでコンポーネント境界内のエラーが分離され、（コンポーネント全体ではなく）関連するチップだけがオフライン化されます。チップに至るまでエラーを分離することで、安定性が向上し、最大限の処理能力を継続的に発揮できます。この機能は、CPU、メモリ・アクセス・コントローラ、クロス・バーASIC、システム・コントローラ、およびI/O ASICに適用されます。
- 継続的な環境監視によって、適切な環境/エラー条件の履歴ログが提供されます。

- ホストのウォッチドッグ機能により、ソフトウェアの動作（ドメインのオペレーティング・システムを含む）が定期的にチェックされます。またこの機能では、SPファームウェアを使用してエラー通知/リカバリ機能がトリガーされます。
- CPUリソースの動的な縮退により、プロセッサの障害を検出、分離、リカバリできます。この機能では、動的再構成によって、（実行中のアプリケーションの中断なしで）運用システムにCPUリソースが動的に再割当てされます。SPARC M6-32サーバーは、CPUの動的な縮退をサポートします。
- コンポーネント・ステータスの定期的なチェックを実行することで、多くのシステム・デバイスのステータスを確認して差し迫った障害の兆候を検出します。リカバリ・メカニズムのトリガーにより、システムとアプリケーションの障害が回避されます。
- エラー・ロギング、マルチステージ・アラート、現場交換可能ユニット（FRU）の電気的な識別情報、およびシステム障害のLEDインジケータにより、問題を迅速に解決できます。

冗長/ホットスワップ対応コンポーネント

今日のIT組織は、不休のビジネス運用という課題に直面しています。ネットワーク化されたグローバル経済では昼夜を問わない運用が行われており、計画的な停止時間枠を短縮するか、場合によっては完全になくする必要があります。SPARC M6-32サーバーでは、これらの要望に応えるため、組み込みの冗長/ホットプラグ対応/ホットスワップ対応のハードウェアを採用して、個々のコンポーネント障害やシステム構成の変更による中断を最小限に抑えています。実際にこれらのシステムでは、（場合によってはユーザーやシステムの機能に影響を与えずに）ハードウェア障害からリカバリすることが可能です。

SPARC M6-32サーバーには、冗長性のあるホットスワップ対応の電源/ファン・ユニットが搭載されており、複数のCPU、メモリDIMM、およびI/Oカードの構成オプションがあります。管理者は、ホットプラグ対応のディスク・ドライブとディスクのミラー化ソフトウェアを組み合わせ、内部の冗長ストレージを作成できます。また、SPARC M6-32サーバーは、冗長性のあるホットスワップ対応のサービス・プロセッサとデグレード可能なスケラビリティ・スイッチ・ボードをサポートしており、冗長クロック・ボードも搭載しています。障害が発生しても、これらの重複コンポーネントによって引き続き運用できます。コンポーネントやエラーの種類によっては、システムの動作がデグレード・モードで続行されたり、再起動されたりする場合があります。この場合、障害が自動的に診断され、関連コンポーネントが自動的にシステム外で構成されます。また、これらのサーバー内のホットスワップ対応ハードウェアによりサービスが高速化され、システムを停止しなくてもコンポーネントの交換や追加を簡単に実行できます。

ホットプラグの場合、デバイスをOracle Solarisから取り外す準備をしたり、Oracle Solarisにデバイスの追加を明示的に指示したりする必要があります。ホットスワップの場合、コンポーネントの削除/追加はSPにだけ通知する必要があります。いずれの場合も、デバイスの追加/削除にOracle Solarisの停止/起動は不要です。SPARC M6-32サーバーの各種コンポーネントについては、表4のリストを参照してください。

表4：ホットスワップ/ホットプラグ対応のコンポーネント

コンポーネント	ホットスワップ対応	ホットプラグ対応
電源とファン	あり	なし
PCIカード	なし	あり
ディスク・ドライブ	なし	あり
サービス・プロセッサ	あり	なし
サービス・プロセッサ・プロキシ	あり	なし

システム管理テクノロジー

ほとんどの組織にとって、サーバー・システムで、ハンズオンのローカルなシステム管理を行うことは現実的ではありません。不休のシステム運用、ディザスタ・リカバリのホット・サイト、および地理的に分散した組織では、システムのリモート管理が必要です。Oracleサーバーの多くのメリットのうちの1つとして、完全自動データセンターのサポートがあります。これにより、高額なサポート・スタッフが、ネットワーク・アクセスにより場所を選ばず作業できます。SPARC M6-32サーバーの設計は、Oracle Integrated Lights-Out Management (Oracle ILOM) ソフトウェアを実行する強力なサービス・プロセッサ (SP) と組み合わされています。この設計とOracle Enterprise Manager Ops Centerソフトウェアのおかげで、管理者はハードウェアに物理的にアクセスしなくても、ほぼすべてのタスクをリモートで実行/制御できます。これらの管理ツールやリモート機能によって管理者の負荷が減り、組織の時間や運用コストを削減できます。

Oracle ILOMとサービス・プロセッサ

SP上のOracle ILOMソフトウェアは、SPARC M6-32サーバーのリモート監視/管理機能にとって非常に重要です。SPはサーバー・システムから独立した専用プロセッサで構成され、Oracle ILOMソフトウェア・パッケージを実行します。SPとドメイン間の通信に使用される内部の100BaseTネットワークがあり、サーバーに入力電力が供給されている間は、すべてのドメインが非アクティブでも、SPによってシステムが継続的に監視されます。

SPによって環境センサーが定期的に監視され、潜在的なエラー状態の事前警告が出されたり、必要に応じて事前予防的なシステム・メンテナンス手順が実行されたりします。たとえば、システムの物理的な損傷につながりうる温度状態の場合、SPによってサーバーが停止される場合があります。管理者は、SPで実行されるOracle ILOMソフトウェア・パッケージを使用して、ドメインやプラットフォーム自体をリモートで制御/監視できます。

オペレータはSPとのネットワーク/シリアル接続を使用して、ネットワーク上の任意の場所からサーバーを効率的に管理できます。SPとのリモート接続はオペレーティング・システムとは別に実行され、システム・コンソールの完全な制御/権限実行が可能です。

SPARC M6-32サーバーでは、1つのSPがアクティブ、もう1つのSPがスタンバイとして構成されます。2つのSP間のSPネットワークによって、システムの管理情報を簡単にやり取りできます。障害が発生しても、SPは同期済みでロールを変更できる状態になっています。

すべてのOracleサーバーでOracle ILOM SPがシステム・コントローラとして機能していれば、SPARC M6-32サーバーのリモート管理は簡単です。このSPはフル機能で、オラクルの他のモジュラー・サーバーやラックマウント型サーバーで使用される実装に似ています。このため、サーバーと既存の管理インフラストラクチャを簡単に統合できます。Oracle ILOM SPには、効率的なシステム管理に不可欠な次の機能があります。

- IPMI 2.0準拠のSPの実装により、サーバーのファームウェア、OS、アプリケーション、および（Oracle ILOM 3.2のイーサネット管理インタフェース経由でアクセスする）IPMIベースの管理ツールに対して、IPMI管理機能を提供します。また、SPによってシャーシ内のサーバー・モジュールその他の場所にある環境センサーを確認できます。
- CPU、DIMM、電源を含むサーバーのインベントリと環境制御を管理し、HTTPS、CLI、およびSNMPによってこのデータにアクセスできるようにします。
- テキスト表示のリモート・コンソール・インタフェースを提供します。
- すべてのシステム・ファームウェアのアップグレードをダウンロードするための手段を提供します。

また、管理者はOracle ILOM SPによって、プラットフォームで実行されているオペレーティング・システムに関係なく、システム・アクティビティを妨げずにサーバーをリモート管理できます。Oracle ILOMによって、サーバー関連のハードウェア障害、警告、およびその他のイベントに関する電子メール・アラートが送信される場合があります。この電子回路は、サーバーのスタンバイ電源を使用して、サーバーから独立して稼働します。そのため、ILOM 3.2のファームウェアとソフトウェアは、サーバーのオペレーティング・システムがオフラインになった場合や、サーバーの電源がオフになった場合でも、動作し続けます。Oracle ILOMは、サーバーの次の状態を監視します。

- CPUの温度状態
- ハード・ドライブの有無
- エンクロージャの温度状態
- ファンの速度とステータス
- 電源のステータス
- 電圧の状態
- Oracle Solarisの予測的自己修復、ブート・タイムアウト、およびサーバーの自動再起動イベント

すべてのPDom構成（CMUボードのホットプラグを含む）は、SP上で、CLI経由でOracle ILOMコマンドを使用するかWebブラウザで実行されます。

電源管理

サーバーの電力/冷却コストは増大しており、これらのコストの削減が企業のデータセンターの一番の課題です。使用可能な電力やデータセンターの拡張スペースは限られているため、お客様はサーバーの電力効率に敏感になっています。所定の消費電力を超えると違約金が発生する契約を電力業者と締結している場合、顧客がサーバーの消費電力を自分で制御して制限できる必要があります。電力効率と二酸化炭素の排出量は、顧客がサーバーを評価する際の要素となってきました。

オラクルのマルチコア/マルチスレッド設計の本来の効率性に加え、SPARC M6およびT5プロセッサには独自の電力管理機能がプロセッサのコア・レベルとメモリ・レベルの両方で組み込まれています。電力消費を抑えるため、これらの機能には、命令速度の低下、アイドル状態にあるスレッドとコアの保留、コアとメモリの両方のクロック・オフ機能が含まれています。

Oracle ILOMでの電源管理サポートの他、Oracle Solaris 11.1には電源マネージャが搭載されています。このようにOracle Solarisでは、poweradm設定に基づいて、有効にする節電機能が決定されます。この設定は、システム (Oracle ILOM) のポリシーに基づいてプラットフォームにより設定されますが、Oracle Solarisの管理者がオーバーライドできます。

以下の領域で大幅なイノベーションが達成されています。

- 実行されない条件分岐など、推測に対する制限
- データ・パス、制御ブロック、アレイにおける広範なクロック・ゲーティング
- 余分なストール・サイクルをデコード段階に注入できるようにする電源調整

Oracle VM Server for SPARCを使用する仮想化環境では、論理ドメイン・ゲストを管理する際、電源管理マネージャによって次のタスクが実行されます。

- 電源管理ポリシーに基づき、有効にする節電機能を決定する
- 電源管理エンジンを呼び出してリソース上で電力状態の変更を開始し、電力の調整/使用率レベルを達成する (Oracle Solaris 11.1ゲストが所有していないリソースの場合)、またはハイパーバイザに指示して、ハイパーバイザ/ハードウェア管理の電力状態を有効/無効にする。電源管理ピアがあるのは、Oracle Solaris 11.1ゲストだけです。

1つの物理ドメインのみをサポートするシステムでは、既存のインタフェース (/SP/powermgmtポリシー) 経由でシステム全体に電源ポリシーが設定されます。

複数の物理ドメインが含まれるシステムでは、物理ドメインごとに電源ポリシーを設定できますが、シャーシ全体に設定することはできません。制御と監視は、PDomのゴールドデン・サービス・プロセッサ・プロキシで実行できます。

- /SP/powermgmt/budgetによって、物理ドメイン予算が制御されます。
- /SYS/PWRBSによって、物理ドメイン予算のステータスが監視されます。

または、プライマリSPで/Servers/PDomains/PDomain_<#>/SP/powermgmt/budgetを使用します。<#>は、特定の物理ドメイン番号です。物理ドメイン予算を構成するには、プラットフォームの管理者権限/ロールが必要です。

電力利用制御の重要な要素は、消費電力を制限できることです。"ハード・キャップ"（電力制限の猶予期間なし）と"ソフト・キャップ"（電力制限の猶予期間あり）の両方がサポートされています。物理ドメインで有効なパワー・キャップは、物理ドメインが完全に所有するボードの消費電力に対して測定されます。ソフト・キャップの場合、論理ドメインの電源マネージャによって物理ドメイン・リソースの電力状態が調整され、キャップに収束されます。ハード・キャップの場合、Oracle ILOMによってキャップが強制されます。キャップを超過するとシステムをブートできず、違反アクションHardPowerOffが有効になります。SPARC M6ベースのサーバーの場合、Oracle ILOMをボード単位のField-Programmable Gate Array（FPGA）と一緒に使用することで、電力のハード・キャップに対応できます。

SPARC M6-32サーバーの場合、Oracle ILOMがパワー・キャップに適合するための唯一の方法は、ボードをPDomに追加しないことです。

Oracle Enterprise Manager Ops Centerを使用したSPARC M6-32サーバーの管理

SPARC M6-32サーバーでは、インフラストラクチャ・スタック全体の管理を統合する集約型のハードウェア管理ソリューションとしてOracle Enterprise Manager Ops Centerを使用します。高度な仮想化管理機能およびレポート作成機能、アプリケーションからディスクまでの管理、インテリジェントな構成管理などができるため、インフラストラクチャ管理の煩わしさが軽減され、合理化、簡素化が促進されます。すべてのSPARC M6-32サーバーにOracle Enterprise Manager Ops Centerが搭載されているため、データセンターの管理者は、ストレージ、ネットワーク、サーバー、Oracle Solaris、および仮想化環境を単一のインタフェースから監視および管理できます。これにより、運用の効率性が向上し、運用コストが削減されます。

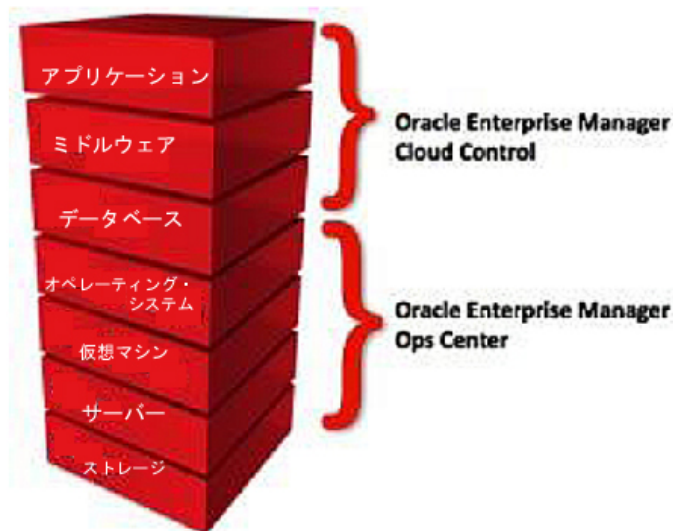


図16：Oracleスタックの管理

Oracle Enterprise Manager Ops Centerは、オラクルのサーバーおよびエンジニアド・システム・インフラストラクチャ向けのもっとも包括的な管理ソリューションです。複数のサーバー・アーキテクチャと無数のオペレーティング・システムの管理に単一のコンソールを使用するOracle Enterprise Manager Ops Centerでは、資産の検出、ファームウェアおよびオペレーティング・システムのプロビジョニング、パッチの自動管理、パッチおよび構成の管理、仮想化の管理、包括的なコンプライアンス・レポートの作成機能を使用して、SPARC M6-32サーバーのコンポーネントを管理できます。

Oracle Enterprise Manager Ops Centerを使用することにより、ポリシー・ベースの管理を通じてワークフローの自動化とコンプライアンスの徹底が可能です。しかも、このすべてを直感的な単一のインターフェースを使用して実現できます。Oracle Enterprise Manager Ops Centerを使用することで、ビジネス要件に合致したインフラストラクチャを効率的に構築しながら、データセンターを標準化し、ベスト・プラクティスを実装し、法令順守およびセキュリティ・ポリシーを徹底できます。図17は、直感的なOracle Enterprise Manager Ops Centerのブラウザベースのユーザー・インターフェースを示しています。

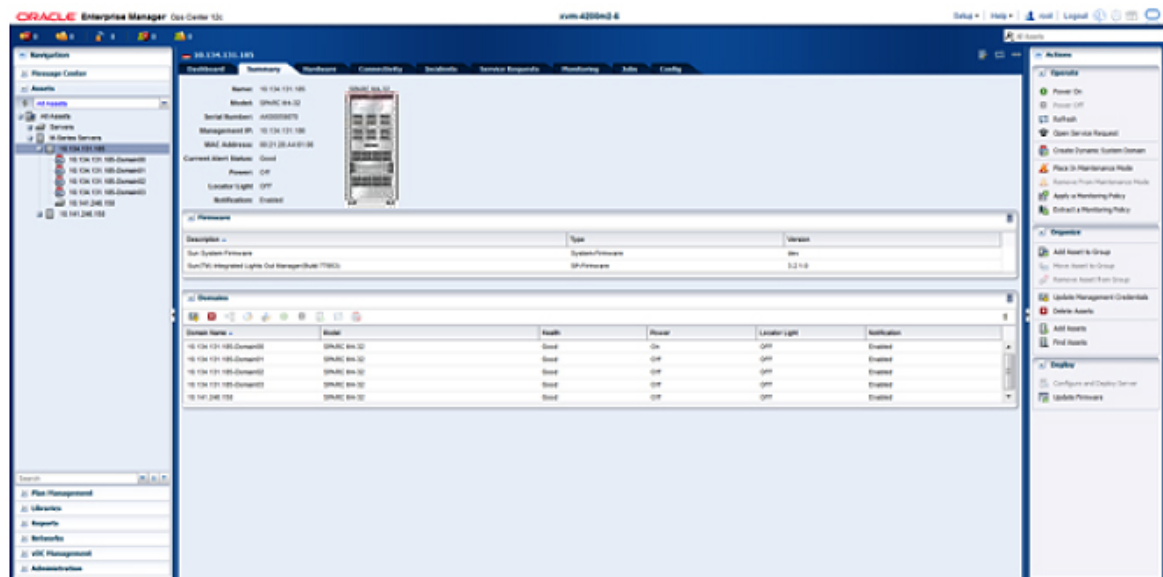


図17：Oracle Enterprise Manager Ops Centerのインターフェース

Oracle Solaris 11オペレーティング・システム

SPARC M6-32サーバーは、Oracle Solaris 10とOracle Solaris 11の両方をサポートします。Oracle Solaris 11は、すべてのドメインで使用できます。Oracle Solaris 10は論理ドメインのゲストのみに使用できます。制御論理ドメインでは使用できません。

Oracle Solarisには次の機能が含まれます。

- **高度な信頼性** - 総合ソリューション・スタック全体の包括的なテストと、ハードウェア障害およびソフトウェア障害を対象とした予測的自己修復やOracle Solaris ZFSの持つデータ整合性機能、Oracle Solaris DTraceによるリアルタイムの可観測性などにより、アップタイムを向上させています。

- **卓越したパフォーマンス** - Oracle Solarisは最新のSPARCプロセッサ・テクノロジーに合わせてスループットとスケーラビリティが最適化されており、Transaction Processing Performance Council (TPC) のTPC-HおよびTPC-Cベンチマーク、オラクルのPeopleSoftアプリケーション、Oracle Business Intelligence Enterprise Editionなどの多くで優れたパフォーマンスを達成しています。
- **組み込みの仮想化機能** - OSの仮想化機能とネットワークの仮想化機能に加え、Oracle Solaris ZonesとOracle VM Server for SPARC (旧称Sun Logical Domains) が組み込まれているため、少ないオーバーヘッドで効率的に統合でき、柔軟性とパフォーマンスが向上します。
- **幅広いセキュリティ・インフラストラクチャ** - Oracle Solarisはマルチテナント環境に必要な区分化機能と制御機能を備えており、政府機関や金融機関の厳密な要件を満たすことができます。
- **確かなサポート** - オラクルは、システムを運用するお客様がいる限り、Oracle Solarisの各リリースのサポートを持続し、ビジネス上の合理性がある限りソフトウェア・インフラストラクチャを使用し続けられるようにしています。

Oracle Solarisの障害管理、サービス管理機能、および予測的自己修復

Oracle Solarisは、障害管理および予測的自己修復が可能なシステムとサービスを構築、導入するためのアーキテクチャを備えています。

- 予測的自己修復機能は、さまざまなハードウェア障害とアプリケーション障害の診断、分離、復旧を自動的に実行します。これにより、ビジネスに不可欠なアプリケーションと基本的なシステム・サービスは、ソフトウェアや主要ハードウェア・コンポーネントに障害が発生した場合や、ソフトウェアの設定ミスがあった場合でも、中断することなく継続できます。
- ハードウェアおよびソフトウェアのエラーに関するデータは、Oracle Solaris Fault Managerアーキテクチャによって収集されます。この機能は、さまざまなエージェントを使用して根本的な問題を自動的にかつサイレントに検出および診断し、障害が発生したコンポーネントをオフラインにすることで自動的に対応します。
- Oracle SolarisのOracle Solaris Service Manager Facility機能は、アプリケーション・サービスを管理者が一定の方法で確認および管理できるファーストクラス・オブジェクトに変換することで、アプリケーション・サービス用の標準化された制御メカニズムを構築します。これで、管理者が誤ってサービスを終了した場合や、ソフトウェアのプログラミング・エラーの結果サービスが中止された場合、または根本的なハードウェアの問題によってサービスが中断した場合に、サービスを自動的に再起動させることができます。

予測的自己修復と障害管理アーキテクチャを使用すると、障害が発生しているプロセッサのスレッドまたはコア、リタイアしたと想定されるメモリ・ページ、I/Oのログ・エラーまたは障害、システムによって検出されたその他のあらゆる問題をオフラインにすることができます。

結論

データセンターのスケラビリティ、信頼性、管理性に対する高レベルな要求をサポートするには、インフラストラクチャをより簡単に導入/構成/管理できるようにしながら、パフォーマンスと機能を向上させる必要があります。SPARC M6-32サーバーにはSPARC M6プロセッサと大容量メモリが搭載されており、アーキテクチャの信頼性が本質的に高いため、企業にとってかつてないレベルのパフォーマンス、可用性、使いやすさを実現しています。また、このサーバーは、ダイナミック・ドメイン、Oracle VM Server for SPARC、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ラブルなSPARC M6-32サーバーを導入することで、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。

追加情報

オラクルのSPARC M6-32サーバーおよび関連するソフトウェアとサービスの詳細については、表5の参照をご覧ください。

表5：参照

SPARCシステム	http://www.oracle.com/jp/products/servers-storage/servers/sparc/overview/index.html
Oracle Solaris	http://www.oracle.com/jp/products/servers-storage/solaris/solaris11/overview/index.html
Oracle Solaris Cluster	http://www.oracle.com/jp/products/servers-storage/solaris/cluster/overview/index.html
Oracle Enterprise Manager Ops Center ソフトウェア	http://www.oracle.com/technetwork/jp/oem/ops-center/index.html
Oracle Support	http://www.oracle.com/jp/support/index.html

付録A:SPARC M6プロセッサのアーキテクチャ

SPARC M6プロセッサは、アプリケーションに真のパフォーマンスを提供する洗練された堅牢なアーキテクチャにより、オラクルのマルチコア/マルチスレッド・イニシアチブを拡張します。図18に、SPARC M6プロセッサのブロック・レベルの図を示します。

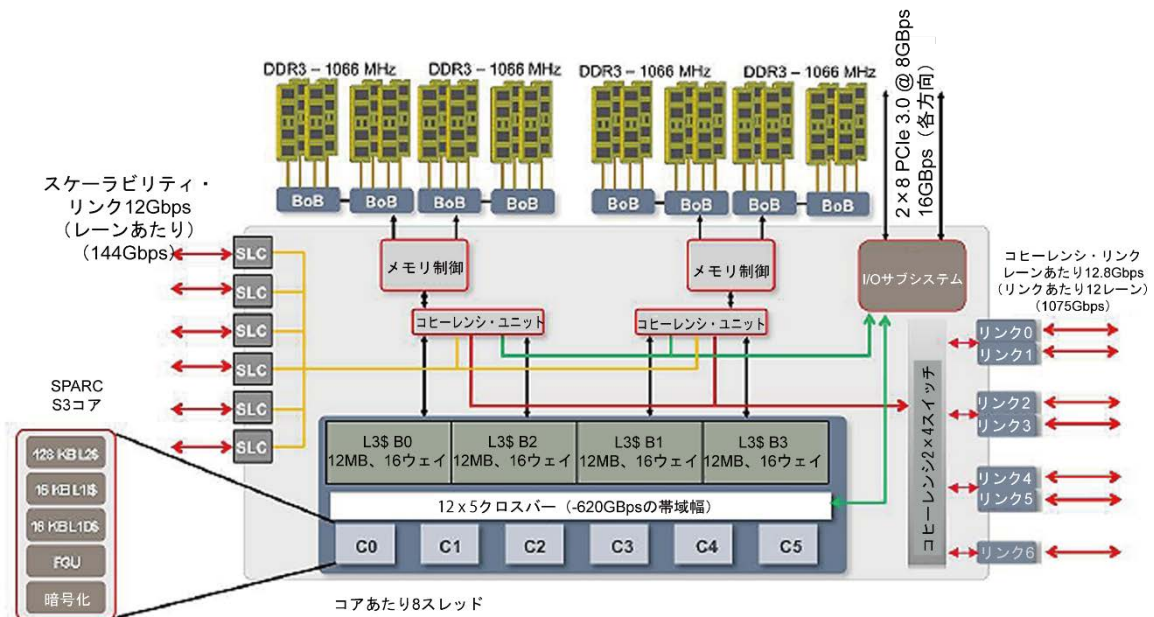


図18：SPARC M6プロセッサ

SPARC M6プロセッサは1チップのマルチプロセッサ（CMP）で、12個の物理プロセッサ・コアが搭載されています。物理プロセッサ・コアはそれぞれ、ストランド×8、整数実行パイプライン×2、浮動小数点パイプライン×1、およびメモリ・パイプライン×1をサポートします。

SPARC M6プロセッサにはアウトオブオーダー、デュアルイシューの堅牢なプロセッサ・コアがあり、このコアは8個のストランド間でしっかりとスレッディングされています。また、16段階の整数パイプラインによる高い動作周波数と高度な分岐予測によって、深いパイプラインや、スレッドへのプロセッサ・リソースの動的な割当ての影響を減らすことができます。このため、SPARC M6プロセッサでは非常に高いスループット・パフォーマンスを実現しながら（以前のSPARC64プロセッサの約6倍）、非常に高レベルのシングル・スレッド実行に拡張できます。

各物理コアには16KB/4ウェイの連想命令キャッシュ（32Bライン）、16-KB/4ウェイの連想データ・キャッシュ（32Bライン）、64エントリの完全連想命令の変換索引バッファ（TLB）、および128エントリの完全連想データTLB（8個のストランドで共有）が搭載されています。また、プライベート、128KB/8ウェイのインクルーシブ・ライトバックL2キャッシュ（32Bライン）も含まれます。さらに、各物理コアには、暗号化アクセラレーション・ハードウェア（ユーザー・レベルの命令でアクセス可能）も含まれます。

SPARC M6プロセッサにはコヒーレンシ・リンク・インターフェースが搭載されているため、DCU内で、外部ハブ・チップを必要とすることなく、最大8基のSPARC M6チップ間で通信できます。各方向にそれぞれ12レーンのコヒーレンシ・リンクが7つあり、153.6Gb/秒で動作します。SPARC M6プロセッサには、コヒーレンシ・リンク・ユニット×7、コヒーレンシ・ユニット×2、およびコヒーレンシ・ユニットとCLU間のクロス・バー（CLX）が搭載されています。

SPARC M6プロセッサには6つのスケーラビリティ・リンクが搭載されており、スケーラビリティ・スイッチ・ボード(SSB)と通信できます。このため、あるDCU内のSPARC M6プロセッサが別のDCU内のSPARC M6プロセッサと通信できます。スケーラビリティ・リンクには各方向に4つのレーンが含まれ、72Gb/秒で動作します。

SPARC M6プロセッサは外部のDDR3 DIMMと、独自の一方方向高速リンクを使用する外部のBuffer-on-Board (BoB) チップ経由でインターフェースされます。SPARC M6プロセッサには、2つのメモリ・リンクがあります。各メモリ・リンクは、12レーンの下り方向と12レーンの上り方向で、12.8Gb/秒で動作します。各メモリは、カスケード構成で2つのBoBと通信します。各BoBチップに2個、SPARC M6プロセッサあたり合計で16個のDDR3チャンネルがあります。各DDR3チャンネルに2枚、SPARC M6プロセッサあたり最大32枚のDDR3 DIMMがあります（BoBあたり4枚のDIMM）。

SPARC M6プロセッサのキャッシュ・アーキテクチャ

SPARC M6プロセッサには3レベルのキャッシュ・アーキテクチャが備わっています。レベル1（L1）とレベル2（L2）は各コアに固有となっています。つまり、これら2つのレベルのキャッシュは他のコアと共有されません。レベル3（L3）は特定のプロセッサの全コアで共有されます。プロセッサが同じ物理システム内にある場合でも、キャッシュは別のプロセッサと共有されません。SPARC M6プロセッサには、個別のデータ・キャッシュと命令キャッシュからなるL1キャッシュがあります。両方ともサイズは16KBで、コアごとに存在します。1つのL2キャッシュもコアごとに存在し、サイズは128KBです。L3キャッシュは、SPARC M6プロセッサの12個すべてのコアで共有され、サイズは48MBです。4つのバンクを備え、16ウェイ・セット・アソシエイティブとなっています。図19に、L2キャッシュとL3キャッシュの関係を、5×12クロス・バーで接続された状態で示します。

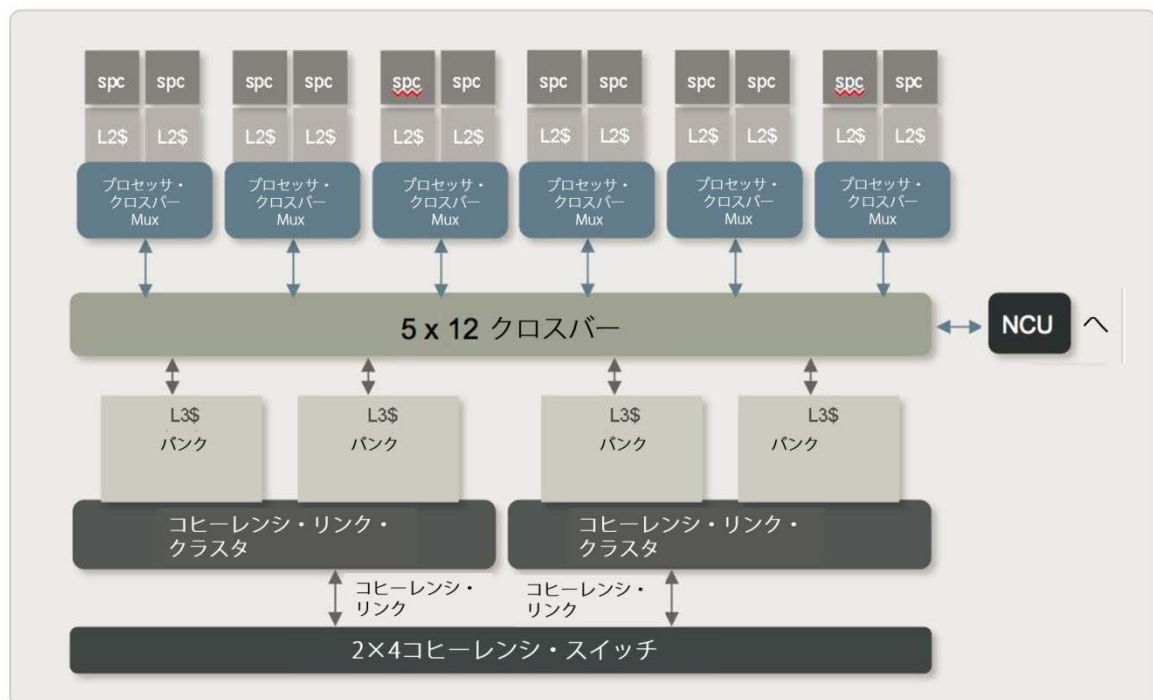


図19：レベル2キャッシュとレベル3キャッシュの関係

SPARC M6コア・アーキテクチャ

SPARC M6プロセッサでは、SPARC64マルチコア・アーキテクチャとSPARC M5プロセッサの延長として、コアが基本的に再設計されています。コア内に、従来からスーパースカラ設計に関連する次の機能が含まれるようになりました。

- アウトオブオーダー（OoO）命令の実行
- 高度な分岐予測
- 命令とデータの両方のプリフェッチ
- より深いパイプライン（Sun/オラクルのマルチコア・プロセッサの前のバージョンと比較した場合）
- 3レベルのキャッシュ
- より大型のメモリ管理ユニット（MMU）のページ・サイズ（2GB）のサポート
- 複数の命令の発行

SPARC M6プロセッサのこれらすべての特性により、スループットのパフォーマンスがSPARC64 VII+の12倍に向上しています。

SPARC S3コア内には、数多くの機能ユニット、パイプライン、および関連する詳細がありますが、本書の範囲には含まれません。ただし、SPARC S3コアには斬新な特性や機能があるため、本書では（SPARC M6-32システムのプログラマーやユーザーが確認できる）おもな機能や特性については説明します。

SPARC M6アーキテクチャの設計者がチップの物理的スペースを節減できた理由の1つは、特定のコアの多数の物理的部分を再利用して、幅広い機能を実現したことです。たとえば、各コア内にある4つの主要なパイプラインでは、各パイプラインの最初の14段階が実際に共有されています。それぞれ最初の14段階を同一にすることで、スペース使用が大きく効率化されています。したがって、2つの整数命令の1つ、浮動小数点グラフィックス命令、またはロード/ストア命令で使用できます。図20では、最初の6ブロックが14個の同一段階を示しています。これについては、特に図22で定義されています。

ダイナミック・スレッディング

SPARC M6プロセッサは動的にスレッド化されます。ソフトウェアではコアごとに最大8個のストランドを同時にアクティブ化できますが、ハードウェアではコア・リソース（命令、データ、L2キャッシュ、TLBなど）やアウトオブオーダー実行リソース（コア中の128エントリ・リオーダ・バッファなど）が動的/シームレスに割り当てられます。これらのリソースは、アクティブなストランド間に割り当てられます。ソフトウェアでは、停止したストランドへの割込みの送信により、ストランドがアクティブ化されます。ソフトウェアでは、非アクティブ化する各ストランド上でHALT命令を実行することで、ストランドが非アクティブ化されます。ストランドには、特別なハードウェア特性はありません。すべてのストランドのハードウェア機能は同一です。

コアではアクティブなストランド間にリソースが動的に割り当てられるため、ソフトウェアをアクティブ化/非アクティブ化するための明示的なシングルスレッド・モードやマルチスレッド・モードはありません。ソフトウェアで、クリティカル・スレッドの最適化（前述）によって、コア上の1個を除くすべてのストランドが効率的に停止されると、コアではすべてのリソースが単一の実行中のストランドに使用されます。このため、このストランドはできるだけ迅速に実行されます。同様に、ソフトウェアで8個のストランドのうち6個が非クリティカルと宣言されると、2個のアクティブなストランドでコア実行リソースが共有されます。

ストランドでコア・リソースの競争が発生する範囲は、その実行特性によって決まります。これらの特性には、キャッシュとTLBのフットプリント、実行ストリームにおける命令間の依存性、分岐予測の有効性などが含まれます。コア上で単独で実行する際に、キャッシュのフットプリントが小さく、分岐予測率が高く正確なプロセスを考えた場合、このプロセスではサイクルあたり2つの命令が実行されます（SPARC M6プロセッサの命令実行の最大速度）。これは、高IPCプロセスと呼ばれます。同じコア上の別のストランドで、特性の似た他のプロセスがアクティブ化された場合、各ストランドではサイクルあたり約1つの命令が実行される可能性が高くなります。つまり、各プロセスのシングル・スレッドのパフォーマンスは半分になります。目安として、N個の高IPCストランドをアクティブ化すると、各ストランドは最大速度の1/Nで実行されます（各ストランドでサイクルあたり約2つの命令を実行できると仮定した場合）。

大部分がメモリバインドされているプロセスを考えてみましょう。このプロセスのネイティブIPCは小さく（場合によっては0.2に）なります。このプロセスをコア上のあるストランドで（他のストランドで他のクローン・プロセスを実行したまま）実行すると、両方のストランドで目立ったパフォーマンスの低下がなく、コア・スループットが0.4 IPCに向上する可能性が高くなります。あるストランドで低IPCプロセスを（他のストランドで高IPCプロセスを実行したまま）実行すると、いずれかのストランドのIPCに大きな悪影響が出る可能性が高くなります。高IPCストランドのパフォーマンス低下は、わずかである可能性があります（低IPCストランドによって、高IPCストランドのキャッシュ/TLBミス率が実質的に増えない場合）。

上記のガイドラインは、一般的な経験則にすぎません。あるストランドが他のストランドのパフォーマンスに与える影響の範囲は、多くの要素によって決まります。プロセス自体は問題なく動作しても、他のストランドで実行した場合は有害なキャッシュやTLBの干渉により、予想外のパフォーマンス低下が発生する場合があります。同様に、複数のストランドを一緒に実行することで一緒にパフォーマンスが上がる場合もあります。たとえば、1つのコア共有コードやデータでストランドを実行した場合です。この場合、あるストランドで（他のストランドで近い将来に使用される）命令やデータをプリフェッチできます。

同じことが、チップで実行されるコア間にも適用できます。コア間でL3キャッシュとメモリ・コントローラが共有されるため、あるコアでのアクティビティが、別のコア上のストランドのパフォーマンスに影響する可能性があります。

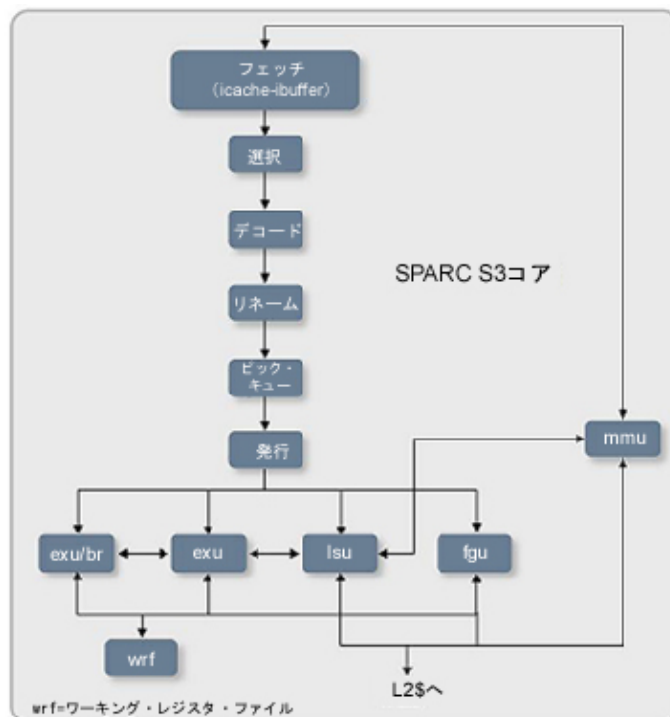


図20：SPARC M6プロセッサのシングル・コアのブロック・レベル図

各コアに実装されているコンポーネントには、次のようなものがあります。

- **トラップ論理ユニット**（図には表示されていません） - トラップ論理ユニット（TLU）によってマシンの状態が更新され、例外や割込みが処理されます。
- **命令フェッチ・ユニット** - 命令フェッチ機能によって、スレッドの選択、選択したスレッドの命令キャッシュ（icache）からの命令のフェッチ、および選択段階のすべてのサイクルへの最大4つの命令の提供が行われます。次の主要機能が実行されます。
 - フェッチするスレッドを選択する。
 - 選択したスレッドのicacheからの命令をフェッチし、選択ユニットの命令バッファに入れる。
 - フェッチ中のスレッド上の遅延制御転送命令（DCTI）の方向とターゲットを予測する。

- icacheミス時にL2キャッシュ (L2\$) からデータをフェッチし、事前デコードしてicacheに格納する。
- **選択ユニット** - 選択ユニットのおもな機能は、各サイクルのプロセッサのパイプラインでの、スレッドの実行をスケジューリングすることです。計8つのスレッドのうち、各サイクルで最大1つのスレッドの実行を選択できます。スレッドの状態は、準備完了が待機です。スレッドは、同期後の状態、分岐予測ミス、有効な命令の欠如、およびその他の命令関連の待ち状態の場合に待機状態となる場合があります。サイクルごとに、選択ユニットによって、準備完了スレッドの中から実行するスレッドが1つ選択されます。この際、Least-Recently-Used (LRU) アルゴリズムによって適正かどうか判断されます。選択したスレッドでは、サイクルあたり最大2つの命令がデコード・ユニットに送信されます。
- **デコード・ユニット** - SPARC M6プロセッサのデコード・ユニットの機能は次のとおりです。
 - 不正な命令の識別
 - サイクルあたり最大2つの命令の整数/FPソースおよびシンクのデコード、およびソース/シンクの依存性の検出
 - 整数/FP登録のフラット・マッピングの生成
 - 条件コードのソースと宛先のデコード
 - 複雑な命令のmicro-opsの生成
 - 命令のスロット割当ての生成
 - DCTI (遅延転送制御命令) の組合せの検出
 - 例外/取消し検出時のNOOPの作成
 - ウィンドウ・レジスタの推論コピーの維持、および特定のウィンドウ・レジスタ命令の実行
 - サイクルごとに最大2つの命令のデコード
 - データの準備および論理マップ表 (LMT) のアドレス指定 (LMTはリネーム・ユニット (RU) の一部)
- **リネーム・ユニット** - リネーム・ユニットの機能は、命令の宛先のリネームと、スレッド内の命令間の宛先とソースの依存性の解決、および発行スロットに基づくエイジ・ベクターの依存性の作成です。リネームにはR1、R2、およびR3の3つのサイクルがあります。各サイクルでは、D2サイクルの最後に、リネーム・ユニットがデコード・ユニットから最大2つの命令を受け取ります。各命令グループはデコード・グループと呼ばれます。リネーム・ユニットによって、デコードから受信した命令のデコード・グループが分割されることはありません。
- **ピック・ユニット** - ピック・ユニットによって、40エントリのピック・キュー (PQ) のうち、サイクルあたり最大3つの命令がスケジューリングされます。R3の第2フェーズの間に、最大3つの命令 (2つの命令と1つのストア・データ取得op) がPQに書き込まれます。PQは、ピック・サイクルの第1フェーズ中に読み取られます。
- **発行ユニット** - 発行ユニットのおもな機能は、命令のソースとデータを実行ユニットに提供することです。図21のとおり、SPARC M6プロセッサには3つの発行スロットに対して6個の実行ユニットがあります。

発行スロット	ユニット
0	ロード/ストア・ユニット 整数実行ユニット0
1	整数実行ユニット1 ブランチ・ユニット FGU SPU
2	ストア・データ操作

図21：発行スロットと実行ユニットの関係

- **浮動小数点/グラフィックス・ユニット** - 浮動小数点/グラフィックス・ユニット (FGU) は各コア内にあり、コアに割り当てられている8つのスレッド全部で共有されます。スレッドあたり、32個の浮動小数点レジスタ・ファイル・エントリがあります。浮動小数点のFused Mul/Add命令が実装されています。また、SPARC64 VII命令セットから整数のFused Mul/Add命令も追加されました。これにより、実行されるアルゴリズムに基づいて一部の暗号計算も実行されます。

SPARC M6プロセッサのS3コアには、16段階の整数パイプライン、20段階のロード/ストア・パイプライン、および27段階の浮動小数点グラフィックス・パイプラインが実装されています。これらすべてがSPARC M6プロセッサの12コアそれぞれに備わっています (図22)。

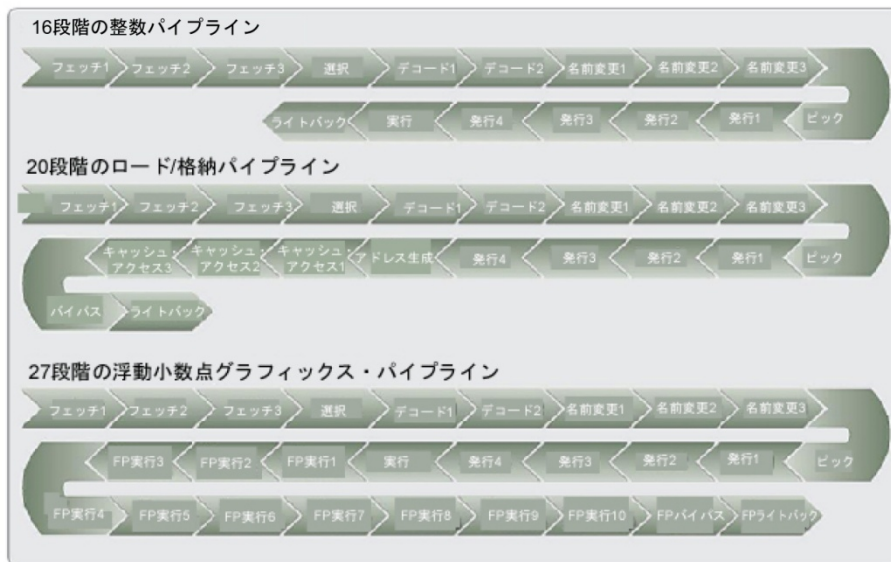


図22：16段階の整数パイプライン、20段階のロード/ストア・パイプライン、および27段階の浮動小数点グラフィックス・パイプラインが、各プロセッサ・各コアに備わっています。

- **ストリーム処理ユニット** - 各コアには、暗号処理を行うストリーム処理ユニット（SPU）が搭載されています。この機能は、SPARC M6プロセッサのコア・パイプライン内に実装されており、29個の新しいユーザー・レベルの命令でアクセスできます。
- **ロード/ストア・ユニット** - ロード/ストア・ユニット（LSU）の機能は、すべてのメモリ参照命令を処理し、すべてのメモリ参照を適切に順序付けることです。ロード/ストア命令がピック・ユニットでピックされると、LSUではこれらの命令が順序付けされない状態で受信されます。他のロードについてはロードが、他のロードやストアに関してはストアが、それぞれ順序付けされない状態で発行される場合があります。ただし、ロードが前のストアより先に発行されることはありません。LSUには、命令セットに必要なメモリ参照の他、（検出されたアクセス・パターンに基づいてデータをL1キャッシュにプリフェッチする）ハードウェア・プリフェッチャも含まれます。
- **メモリ管理ユニット** - メモリ管理ユニットにハードウェア・テーブル・ウォーク（HWTW）が搭載されており、8KB、64KB、4MB、256MB、および2GBのページをサポートしています。
- **整数実行ユニット** - 整数実行ユニット（EXU）では、サイクルあたり最大2つの命令を実行できます。単一サイクルの整数命令は、EXU0（スロット0）またはEXU1（スロット1）のパイプラインで実行されます。ロード/ストアのアドレス操作はEXU0（スロット0）で実行されます。分岐命令はEXU1（スロット1）で実行されます。浮動小数点、マルチサイクル整数、およびSPU命令は、EXU1（スロット1）のパイプラインを経由します。ストア・データ操作はEXU0（スロット2）で実行されますが、EXUでは別の命令とは見なされません。アドレスの格納操作も、同じ命令で実行される必要があるためです。

デュアル整数パイプラインの機能を説明するため、図23ではワーキング・レジスタ・ファイル（WRF）、浮動小数点レジスタ・ファイル（FRF）、および整数レジスタ・ファイル（IRF）を含むデュアルEXUと各種データ・パスを示しています。

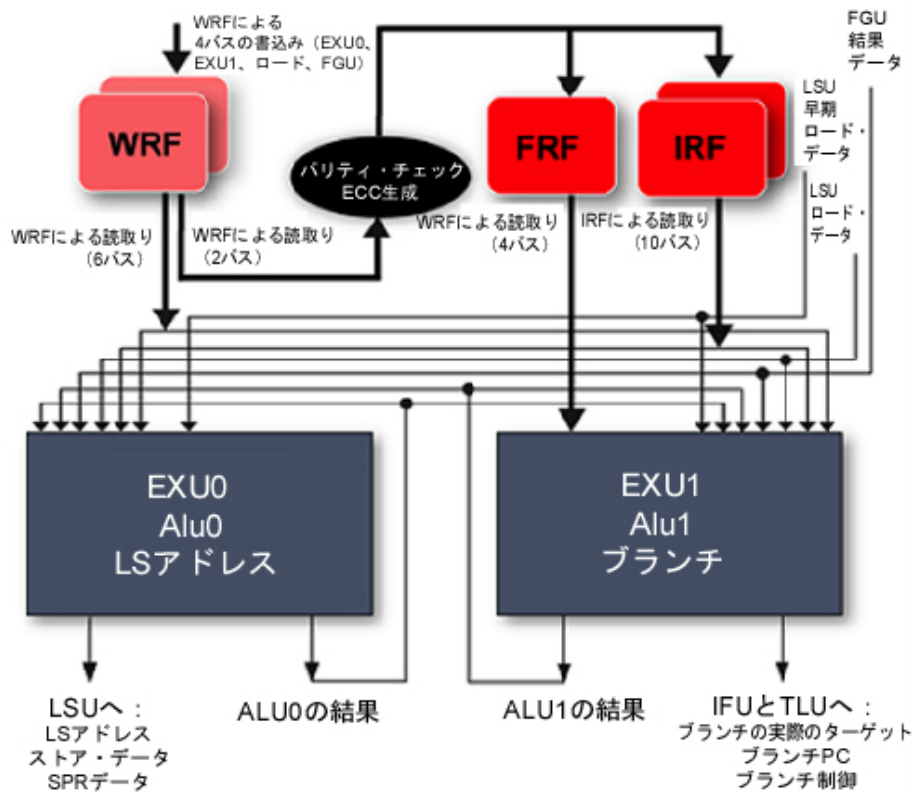


図23：スレッドは2個の整数パイプライン間に交互に配置され、実行される整数演算の種類によってEXU0またはEXU1に限定されます。

ストリーム処理ユニット

各コア上のSPUは、パイプライン自体の一部としてコア内に実装され、コアと同じクロック速度で動作します。SPARC M6プロセッサは、次の暗号アルゴリズムをサポートします。

- DH、DES/3DES
- AES-128/192/256
- Kasumi、Camellia
- CRC32c、MD5
- SHA-1、SHA-224、SHA-256、SHA-384、SHA-512
- MPMUL/MONTMUL/MONTSQR命令を介したRSA

暗号化アルゴリズム（上記のグループのハードウェアでサポート）は実際には、FGUと整数パイプラインの一部を使用します。SPUの基本的な論理パイプラインについては、図24を参照してください。

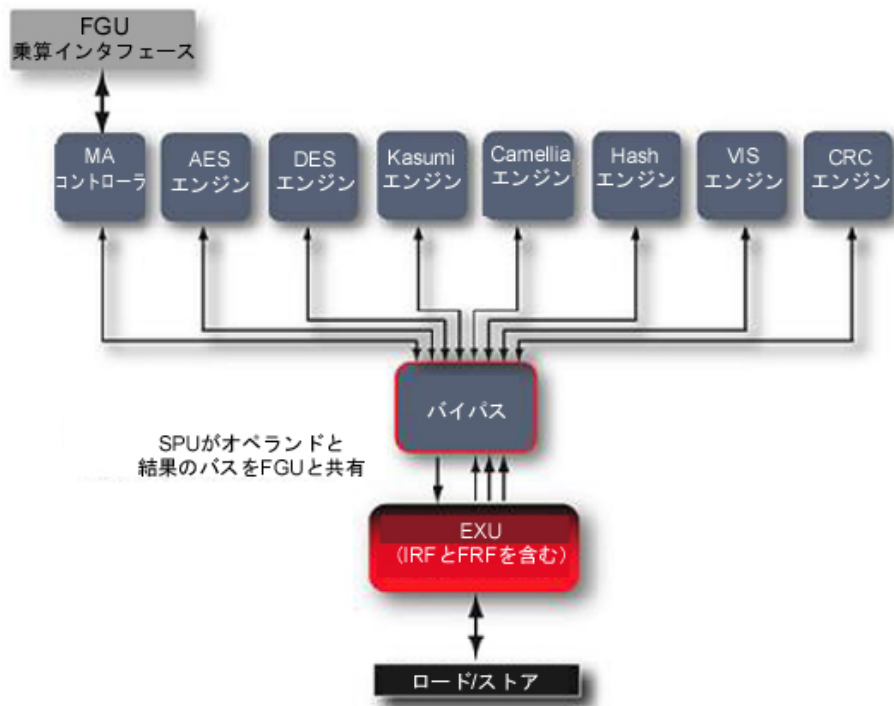


図24：各コア内のSPUパイプラインの論理的な図

内蔵PCIe Generation 3のサポート

SPARC M6プロセッサは、デュアル・オンチップPCIe Gen3インターフェースを搭載しています。各インターフェースは、ポイント・ツー・ポイントのデュアルシンプレックス・チップのインターコネクトを介して、x1レーンごとに8Gb/秒で双方向に動作します。つまり、各x1レーンは2つの単指向性の1ビット幅の接続で構成されており、1つは上り方向のトラフィック、もう1つは下り方向のトラフィック用となっています。内蔵の入力/出カメモリ管理ユニット (IOMMU) はI/Oの仮想化をサポートし、PCIe BUS/Device/Function (BDF) の番号を使用して、デバイスの分離を処理します。理論上のI/O帯域幅の合計は (x8レーンの場合) 16GB/秒となり、最大のペイロード・サイズはPCIe Gen3インターフェースあたり256バイトになります。実現可能な実際の帯域幅は約14.8GB/秒となる可能性が高く、オフチップPCIeスイッチとの統合用にx8 SerDesインターフェースが1つ装備されています。



SPARC M6-32サーバーのアーキテクチャ
2013年9月、バージョン2.0

著者：Gary Combs
Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

海外からのお問い合わせ窓口：
電話：+1.650.506.7000
ファクシミリ：+1.650.506.7200

oracle.com



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

Copyright © 2013, 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、The Open Groupの登録商標です。

1012

Hardware and Software, Engineered to Work Together