

ハイエンド・データセンター・ストレージ

高度なスケーラビリティのために：
Sun StorEdge™ 9900 シリーズ



目次	
目的	1
対象読者	1
はじめに	1
Sun StorEdge™ 9900 シリーズ - オープン・システム設計	2
前例のない内部帯域幅	2
Hi-Star アーキテクチャ	2
共有バス・ストレージ・システムとの比較	5
制御メモリ - 階層型スター・ネットワーク	6
高速バックエンド設計	7
新しいファイバ・チャンネル・ディスク・ドライブ	9
追加パフォーマンス拡張機能	10
Hi-Star アーキテクチャの真のスケーラビリティの証明	10
スケーラビリティの比較	11
テスト構成	11
テストの実施明細	13
テスト結果 - 共有バス・アーキテクチャ	14
テスト結果 - Sun StorEdge 9900 シリーズの Hi-Star アーキテクチャ	15
共有バス・アーキテクチャと Hi-Star アーキテクチャのテスト結果の比較	16
結論	16
用語集	17

高度なスケーラビリティのために： Sun StorEdge 9900™ シリーズ

目的

このホワイト・ペーパーでは、Sun StorEdge™ 9900 シリーズを支える革命的な Hi-Star™ アーキテクチャの概要を紹介し、Sun StorEdge 9900 シリーズの特徴と機能について説明するとともに、この新しいエンタープライズ・ストレージ・システムのパワフルなスケーラビリティを明らかにします。

対象読者

このホワイト・ペーパーは、次世代エンタープライズ・ストレージ・システムについて、十分な情報に基づく意思決定を行なうための参考資料として作成されています。また、この新しいアーキテクチャに関する説明書を作成したり、その高度なスケーラビリティを大いに活用する方法や I/O を多用するオープン・システム・アプリケーション、ソリューションのアドバイザを務める人々にも役立ちます。ストレージ製品に関する予備知識があれば、Sun StorEdge 9900 シリーズのエンタープライズ・ストレージ・システムとしての革命的な設計の精密さを理解、評価しやすくなります。

はじめに

ニュー・エコノミーでの成功は、エンタープライズの情報インフラストラクチャによって決定されます。このインフラストラクチャは、電子商取引アプリケーションの大量のトランザクションを処理し、膨大な量のテキスト、グラフィックス、オーディオ、ビデオを無数のインターネット・ユーザに日々配信しなければなりません。

このようなビジネス環境では、電子商取引サイトがダウンしたり、応答に時間がかかったりすると販売機会の喪失につながり、広く知られた場合には、株価が急落し、時価総額が大幅に減少することもあります。そのため、エンタープライズの情報インフラストラクチャを支えるストレージ・システムは、容量とパフォーマンスを不断に拡大し、アクセスの急増に耐え、幾何級数的なデータの拡大に対応する一方で、昼夜を問わずいつでも使用できるものである必要があります。

Sun は、世界で最も先進的なエンタープライズ・ストレージ・システムの Sun StorEdge 9900 シリーズを市場に投入することで、この難題に答えています。ニュー・エコノミーを念頭に入れて設計されている Sun StorEdge 9900 シリーズは、現時点で可能な最大の容量（37TB）を提供し、コンポーネントを増設しても速度が落ちない、ボトルネックのないパフォーマンスを提供する革命的なアーキテクチャを備えています。Sun StorEdge 9900 シリーズを使用すれば、ニュー・エコノミーの中でビジネスの拡張

と成功の拡大を支援する、自社の情報インフラストラクチャの能力に自信を持つことができます。

Sun StorEdge™ 9900 シリーズ - オープン・システム設計

Sun StorEdge™ 9900 シリーズは、今日の業界で、最もパワフルなエンタープライズ・ストレージ・システムです。Sun StorEdge 9900 シリーズで採用されている新しい内部交換アーキテクチャは、オープン・システム環境の苛酷な I/O デマンドを容易に処理できるように特別に設計されています。非常に高い内部帯域幅、高速バックエンド設計、スケーラブルな内部パスウェイ、プロセッサ速度の向上とプロセッサ数の増大、拡大されたキャッシュ・サイズ、改善された新しい高性能 RAID-1 (RAID-1+)。これらによって Sun StorEdge 9900 シリーズは、他のすべてのストレージ・システムと一線を画するものとなっています。以下の項では、Sun StorEdge 9900 シリーズ・ストレージ・システムの物理コンポーネントと論理コンポーネントについて説明します (アーキテクチャの概要は、図 1 を参照してください)。

前例のない内部帯域幅

Sun StorEdge 9900 シリーズのスループットのための巨大な能力の心臓部は、階層型スター・ネットワーク (HSN : Hierarchical Star Network) すなわち Hi-Star™ アーキテクチャです。これは、サブシステム・コンポーネントに相互接続された高速リンクのネットワークで、6.4GB/s という破格の内部総帯域幅を提供します。これは、データ転送用として 3.2GB/s、個別の内部制御情報の転送用として 3.2GB/s です。旧世代のストレージ・システムの共有共通バス・アーキテクチャとは違って、Hi-Star アーキテクチャは、まさしく最適なシステム・パフォーマンスに合わせてスケーリングできるように設計されています (共有共通バス・アーキテクチャの例は、図 2 を参照してください)。

Sun StorEdge 9900 シリーズ・エンタープライズ・ストレージ・システムのメイン・コンポーネントであるアレイ・コントロール・プロセッサ (ACP : Array Control Processor) とクライアント・ホスト・インターフェイス・プロセッサ (CHIP : Client-Host Interface Processor) は、それぞれ独自のネットワーク・パス (および帯域幅) をキャッシュ・HSN と制御メモリ・HSN に提供することで、Hi-Star アーキテクチャの帯域幅全体に貢献します。このことは、Sun StorEdge 9900 シリーズは、ACP の 4 つすべてのペア (合計 8 モジュール) と CHIP の同じく 4 つすべてのペア (合計 8 モジュール) によって完全に構成されたときにのみ、6.4GB/s の内部帯域幅を実現できることを意味します。またこれは、Sun StorEdge 9900 シリーズでは、高性能の構成も費用効果の高い構成も可能である、という意味でもあります。

Sun StorEdge 9900 シリーズの拡張やスケーリングで接続用のファイバ・チャンネルを増設する場合も、容量増大やパフォーマンス向上のための ACP を増設する場合も、内部帯域幅は、コンポーネントを増設するたびに拡大します。規定の帯域幅はパスの定格帯域幅であり、使い方のタイプと帯域幅の消費量によって実効速度は異なります (つまり、「マイルージ」は変化することがあります)。

Hi-Star アーキテクチャ

キャッシュ - 階層型スター・ネットワーク

内部 Hi-Star アーキテクチャは、キャッシュ - HSN と制御メモリ - HSN の 2 つの独立したネットワークから構成されています。キャッシュ - HSN (C-HSN) は、メイン・グローバル・キャッシュとの間のデータ転送用のネットワークです。C-HSN は、さら

にC-HSNへのプロセッサ・パスと実際のC-HSNの2つの側面に分割できます。このプロセッサ・パスとC-HSNの詳細を図3に示します。

C-HSNの心臓部は、C-HSNの内部スイッチト・ファブリック・ネットワークを形成する4つのキャッシュ・スイッチ(CSW: Cache Switch)です。CSWは、コンビネーションMUX、パス、パス・アービトラータおよび非ブロック・ネットワーク・スイッチとして機能する特別設計のクロスバー・スイッチです。CSWはMUXとして機能し、各スイッチのプロセッサ側へのパスが8つあり、キャッシュ・モジュールへのパスが4つあります(8-4ルーティング)。

ファブリック・ネットワークのプロセッサ側には合計で32のパスがあり、キャッシュ・モジュールからファブリック・ネットワークのキャッシュ側へのパスは16あります。これとは別に、CSWはアービトラータとしても機能します。キャッシュ・メモリアービトラータ(CARB: Cache Memory Arbitrator)は、キャッシュ・モジュールとの間のアクセスとアドレッシングを処理します。さらに、CSWはプロセッサとキャッシュ間の非ブロック・パスも管理します。CSWの詳細なダイアグラムは、図4を参照してください。

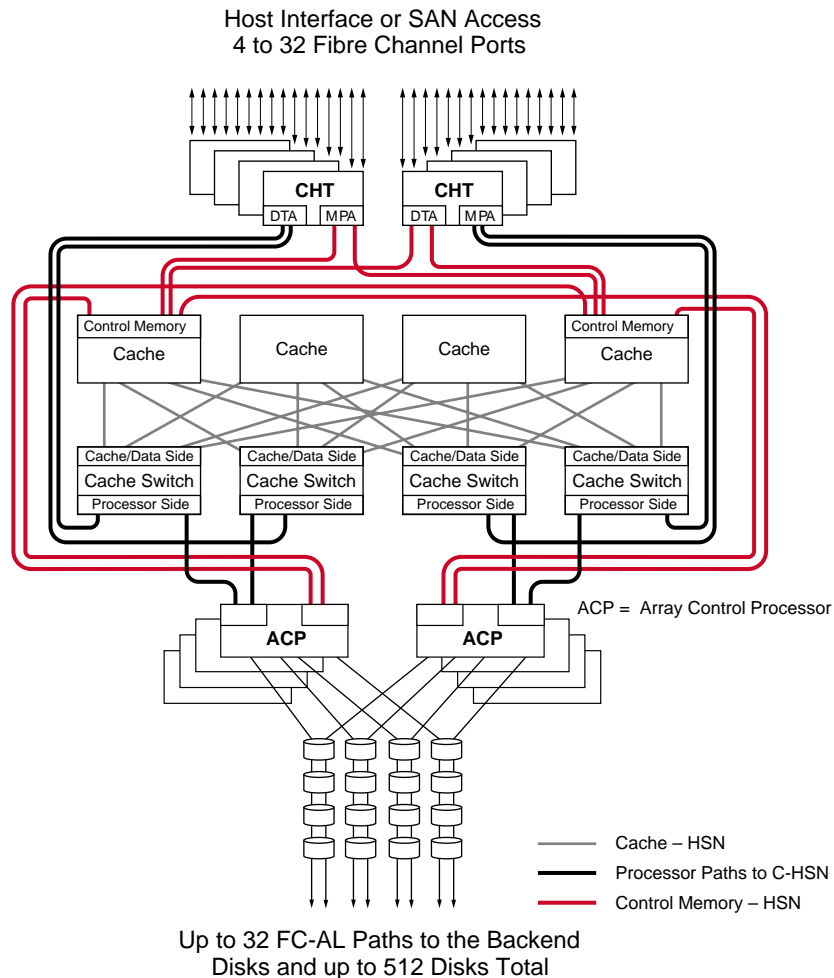


図1: Sun StorEdge 9900 シリーズの内部Hi-Starアーキテクチャの概要

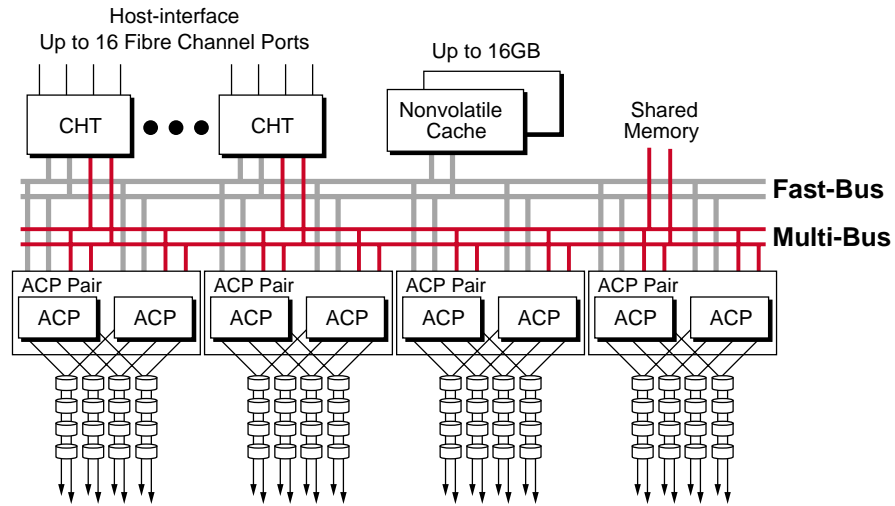
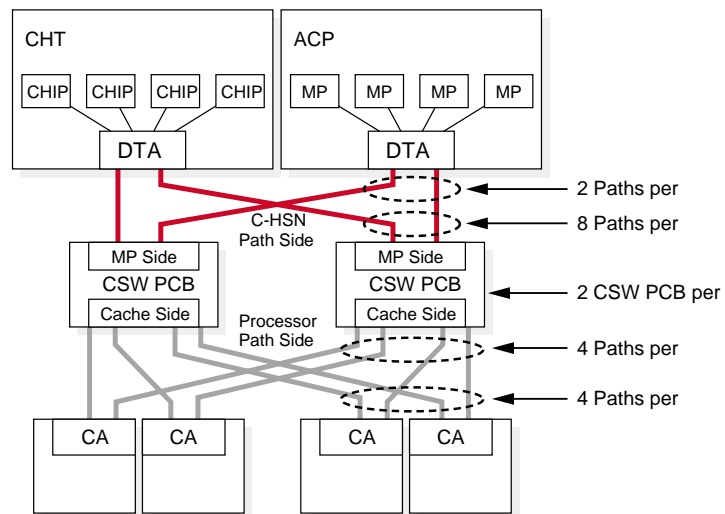


図2：多くのストレージ・アレイに共通する、内部共有バス設計



Switched Point-to-Point Connections through the Cache Switch

C-HSN
16 bit x 100 MHz
x 16 Paths = 3.2 GB/s Total C-HSN
200 MB/s per Path
x 2 Paths = 400 MB/s per PCB
32 Paths x 200 MB/s = 6.4 GB/s to the C-HSN

図3：キャッシュ - 階層型スター・ネットワーク (C-HSN) と4基のキャッシュ・スイッチ (CSW) のうちの2基

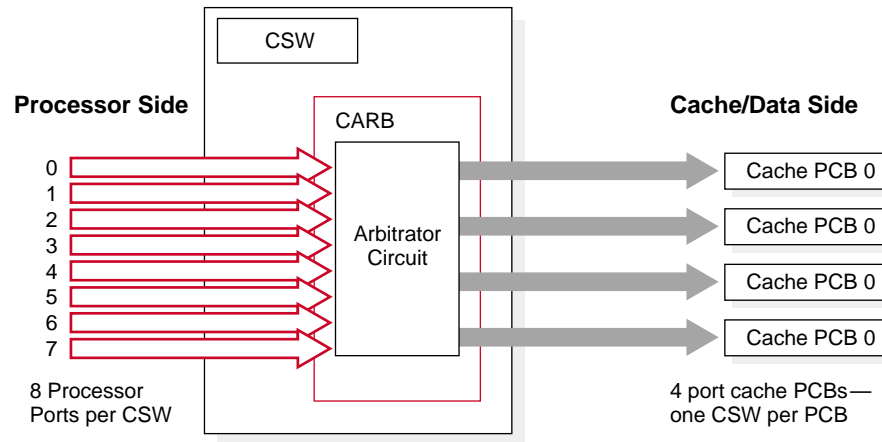


図4：CSWとキャッシュ・メモリ・アービトラータ（CARB）

プロセッサを（データ・アダプタ（DTA：Data Adapter）を介して）CSWに接続し、CSWをキャッシュ・モジュールに接続するバスはそれぞれが16ビット幅で、これにパリティ用の2ビットが加えられます。各バスの帯域幅は200MB/sです。これは、クロック周波数が100MHzの16ビット幅バス（シリアルではなく、パラレル）のデータ転送速度です（100MHzの16ビットは200MB/sに相当します）。先に述べたように、Sun StorEdge 9900シリーズのプロセッサ・コンポーネント（ACPとCHIP）からCSWのポートへのバスは32あります。これは、（16個のコンポーネント・モジュールがすべて取り付けられている場合）スイッチト・ファブリックのプロセッサ側への6.4GB/sの帯域幅

（200MB/sの32のバス）に相当します。ただし、スイッチト・ファブリックのキャッシュ側とキャッシュ・モジュール間の帯域幅は合計で3.2GB/s（200MB/sの16のバス）です。この帯域幅は、4基のCSWと4つのキャッシュ・モジュールが取り付けられて機能している場合、常に3.2GB/sになります。

共有バス・ストレージ・システムとの比較

共有バス・アーキテクチャの構成を検討する場合は、バス上で同時に存在できるオペレーションの数を考慮する必要があります。共有バスや複数共有バス設計に基づくストレージ・コントローラでは、発生可能な内部同時プロセッサ・オペレーションの数は、プロセッサ、キャッシュ、制御メモリ（もしくは共有メモリ）の3者間の交信に用いられるバスの数と同じです。（図2の場合）は、その2つの完全なバス（データおよび制御情報）により、この数は2です。4つの完全なバス（同じバスで共有される制御とデータ）から構成される他のベンダのエンタープライズ・ストレージ・アレイでは、その数は4です。他方、Sun StorEdge 9900シリーズでは、プロセッサとキャッシュを結ぶ非ブロック・スイッチト・ファブリック技術により、その数は16（CSWとキャッシュ・モジュール・ポート間のバスの総数）になります。C-HSNの詳細は、図5を参照してください。ストレージ・コントローラのスケーラビリティを検討するときには、この概念を理解することが大切です。

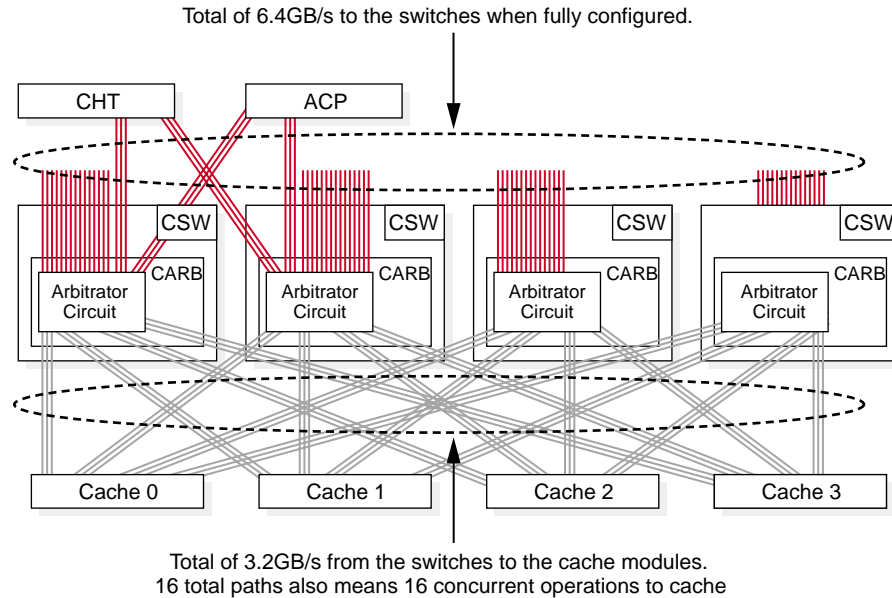


図5：キャッシュ・スイッチとキャッシュ・モジュールの相互接続の全容

制御メモリ - 階層型スター・ネットワーク

Hi-Star ネットワークのもう1つのコンポーネントは、制御メモリ - HSN (CM-HSN) です。このネットワークは、プロセッサと制御メモリとの間の制御情報のやりとりを処理します。制御メモリには、システムに関するメタデータ、すなわちキャッシュの状態、位置、構成、キャッシュ内のデータ、Sun StorEdge 9900 シリーズの構成、ストレージ・システムの運用状態などに関する情報が格納されます。相互のミラー・イメージである、2つの制御メモリ領域があります (制御メモリの論理的、物理的位置は、図1を参照してください)。制御情報は、本質的に個別のメモリ領域とネットワークの両方を介して、データから帯域外で処理されます。

CM-HSNは、キャッシュ - HSN よりも一段と簡素な設計で、いずれの接続もポイント・ツー・ポイント接続です。キャッシュ - HSN (データ・バス) だけが、その相互接続ネットワークのためにスイッチト・ファブリック・トポロジを使用しています。また、CM-HSNは、キャッシュ - HSN よりも狭隘なバスを多用します。CM-HSNのネットワーク・トポロジの詳細は、図6を参照してください。図1のダイアグラムでは、プロセッサと制御メモリを結ぶCM-HSNパスが2つありますが、図6のダイアグラムでは、プロセッサ・モジュールごとに4つのバスがあり、プロセッサと制御メモリを結ぶ4ビットのバスが64あります。また、図1では、完全な8ビット (これにパリティ・ビットが加えられます) のバスに統合された4ビットのバスを示していますが、拡大した図6では、制御メモリへのポートがすべて「分割」構成になっています。さらにまた、制御メモリはミラーリングされます。このバス・トポロジは、このミラーリング・アーキテクチャを容易にするために用いられています。

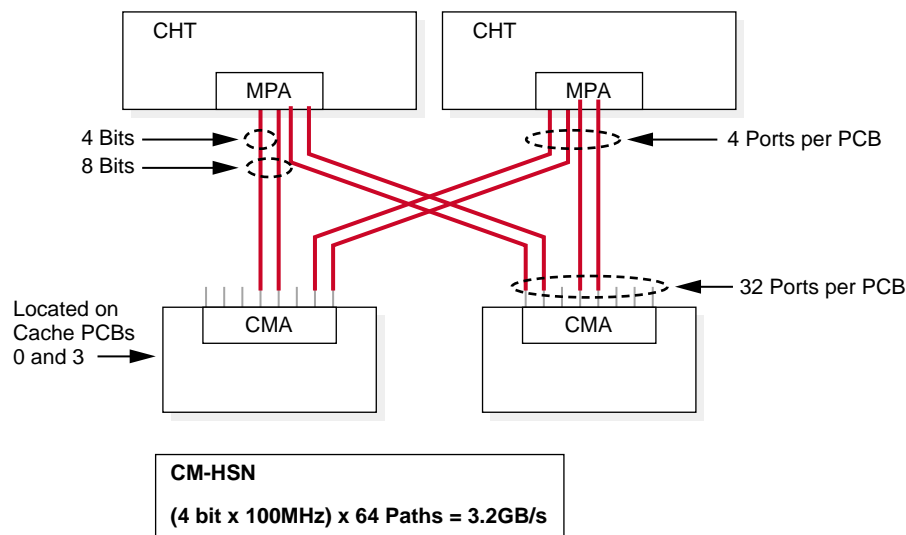


図6：制御メモリ - 階層型スター・ネットワーク (CM-HSN)

CM-HSMの帯域幅をいずれの方法で計算しても、プロセッサ・モジュールがすべてインストールされていれば、総帯域幅は3.2GB/sになります。たとえば、クロック周波数が100MHzの4ビットは、1パスにつき50MB/sに相当します。これにポート数の64を乗じると、3.2GB/sになり、クロック周波数が100MHzの完全な8ビットのパスに制御メモリ「ペア」の複合ポート数32を乗じた場合も、速度は3.2GB/sになります。

高速バックエンド設計

新しい高性能バックエンド・ドライブが、FC-AL (Fibre Channel-Arbitrated Loop) を介して取り付けられています。8つの100MB/s FC-ALによって、アレイ・グループとACPの各ペアのデュアル・ポート・ドライブに接続されます。構成されたACPの4つのペアが(1ペアごとに800MB/sで)インストールされているため、物理ドライブへのバックエンド帯域幅の合計は3.2GB/sになります。

各FC-ALパスは、専用プロセッサとデータ復旧再生 (DRR : Data Recovery and Regeneration) 回路によって駆動されます。Sun StorEdge 9900 シリーズのACPとACPペアのダイアグラムは、図7～9を参照してください。

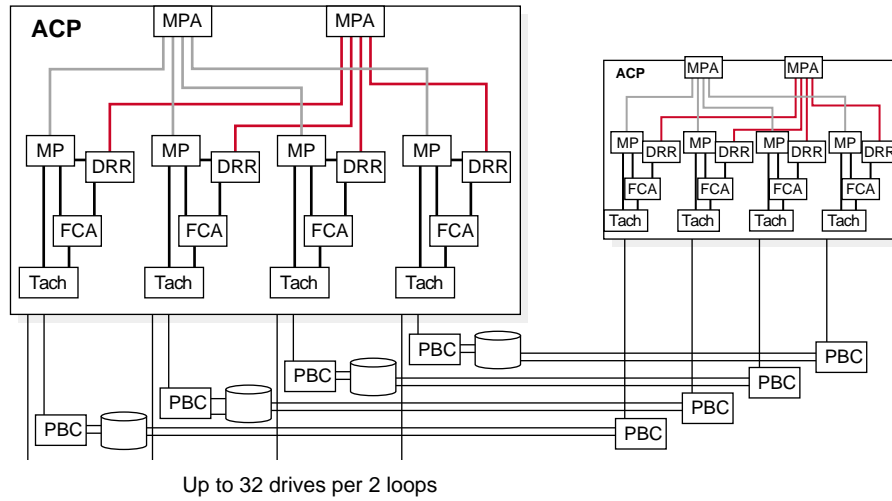


図7 : ACPペア

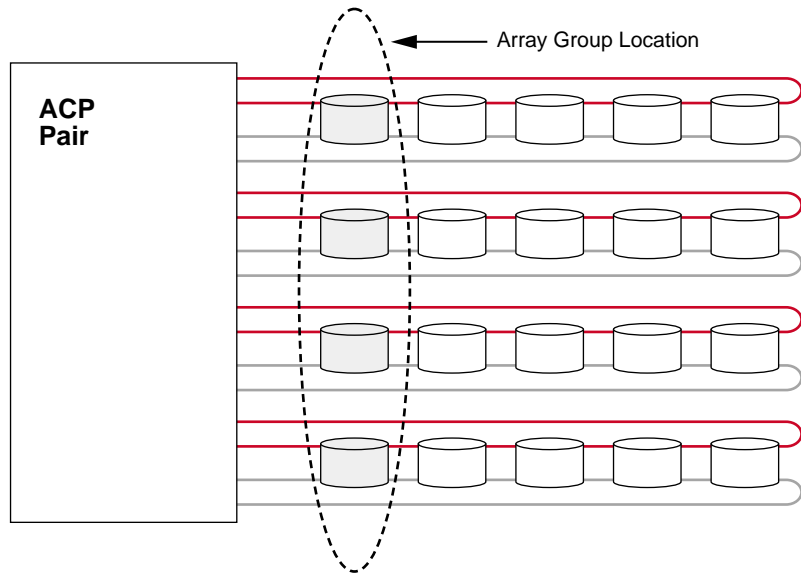


図8 : ACPペアとアレイ・グループの位置

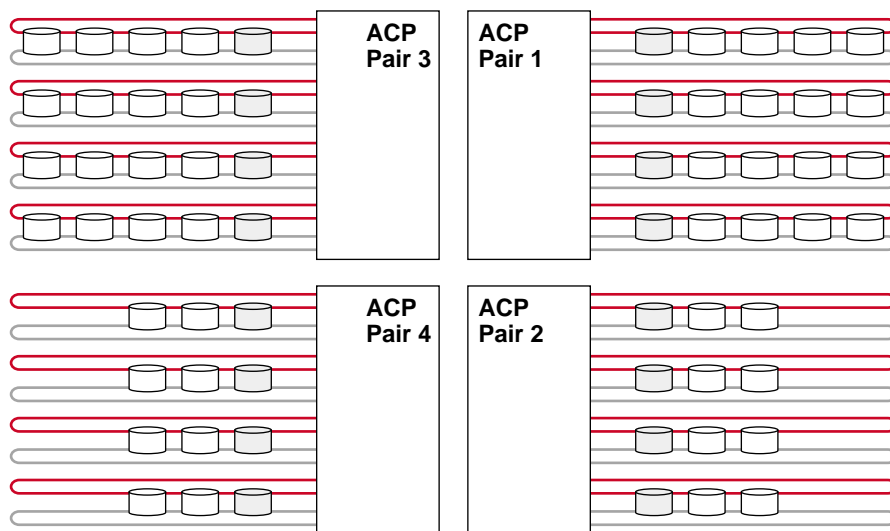


図9：4つのすべてのACPペアと32のすべてのFC-AL（3.2GB/s）

新しいファイバ・チャンネル・ディスク・ドライブ

Sun StorEdge 9900 シリーズのバックエンド・パフォーマンスは、新しいデュアル・ポート/デュアル・アクティブ・ファイバ・チャンネル・ディスク・ドライブ技術によって増強されています。フォーム・ファクタが3インチのこれらの高密度ディスク・ドライブは、

Sun StorEdge 9900 シリーズ用に特別に設計、製造されたものです。容量は18GB、73GBの3種類で、10,025 RPMで回転します。これらの新しいディスクの転送速度は60MB/sに達します。物理トラック1つ当たりのセクタ数が増大すると、7700Eで使用できる15GB、12,030 RPMのディスクと比較した場合でも、ヘッドの移動が減少し、1回転当たり転送されるデータ量が増加することになります。

デュアル・ポート/デュアル・アクティブ・ドライブをACPペアに組み込まれている技術と共に使用することで、Sun StorEdge 9900 シリーズのバックエンドは、8つすべてのFC-ALパスをACPペアで使用して、最適なパフォーマンスとフォールト・トレランスを実現します。ただし、1つのディスクもアレイ・グループも、両方のパスを同時に使用することはできません。ACPは、パスのアクティビティと使用率を監視し、その情報に基づいて、ディスクへのアクセスに最適なパスを判別します。1つのアレイ・グループが、それに接続されている8つのFC-ALパスをすべて同時に使用することはできませんが、ACPペアのビジー状態にある多くのアレイ・グループが、ACPペアの8つのパスをすべて利用することは可能です。

さらにSun StorEdge 9900 シリーズでは、コマンドタグ・キューイング(CTQ: Command-Tag Queuing)をバックエンド・ドライブに導入しています。この新しい機能は、シーク最適化機能の多くをディスク・ドライブ自体にオフロードし、より多くのバックエンドI/Oオペレーションを発生できるようにして、バックエンドのパフォーマンスを大幅に向上します。

追加パフォーマンス拡張機能

Sun StorEdge 9900 シリーズ・ストレージ・システムには、それを業界の最先端システムたらしめる追加パフォーマンス拡張機能が用意されています。以下、Sun StorEdge 9900 シリーズに組み込まれている追加のパフォーマンス拡張機能について簡単に説明します。

- フロントエンドで使用可能な最大 32 基のファイバ・チャンネル・ポート。これは、文字通り 32 基のポートです。これらのポートを収容しても、サブシステムの別の場所でカード・スロットの数は減りません（つまり、ACP ペアを取り外すと、新たに 8 つのポートを取り付ける余地が生まれます）。
- 最大 32GB のキャッシュ。
- Sun StorEdge 9900 シリーズのパフォーマンスを管理するための IOPS ポート・コントロール。ポート使用率（MB/s または IOPS）に基づいて、パフォーマンスを動的に増減できます。これにより、Sun StorEdge 9900 シリーズのフロントエンド・ポートで Quality of Service（QOS）を実現し、帯域幅、パフォーマンス、リソースを管理することが可能になります。この機能は、Storage On Demand（SOD）サービスを提供しているストレージ・サービス・プロバイダ（SSP）には非常に役立ちます。Sun StorEdge 9900 シリーズを使用すると、SSP は、自社の顧客に Performance On Demand（POD）という別のサービスを提供できます（この機能については、後続のホワイト・ペーパーで独立したトピックとして説明します）。
- Sun StorEdge 9900 シリーズの発表と共に、新しい RAID レベルである RAID-1+ が紹介されています。RAID-1+ は、ツイスト（回転コピー）を伴う RAID-1+0 構成の 2Data + 2Data ディスク配列（合計 4 ディスク）です。これには、パフォーマンスと信頼性の面で、通常の RAID-1 や後続のホワイト・ペーパーで検討する RAID-1+0 を上回るいくつかの利点があります。

Hi-Star アーキテクチャの真のスケラビリティの証明

Hi-Star アーキテクチャの設計に関する説明から、Sun StorEdge 9900 シリーズが、今までに作成されたどのエンタープライズ・ストレージ・システムよりも優れたスケラビリティを備えていることは明らかです。基本的に、このアーキテクチャのコンポーネントを多く使用すればするほど、速度が向上します。Sun StorEdge 9900 シリーズの容量を増設し、接続を増やすにつれて、総パフォーマンスが向上します。完全な 6.4GB/s の内部帯域幅は、16 個のモジュール（CHIP および ACP）がすべてインストールされ、機能しているときにしか達成されませんが、フロントエンド・ポートがそれぞれ 100MB/s であれば、関係ありません。

このアーキテクチャのもう 1 つの課題は、これをどう使用するかということです。簡単に言うと、このアーキテクチャは、使用しないと得られないということです。つまり、1 基のファイバ・チャンネル・ポートを通じて 1 つの LUN を実行する環境の場合には、Sun StorEdge 9900 シリーズを完全に構成し、6.4GB/s の内部帯域幅を完備しても無益だということです。Sun StorEdge 9900 シリーズで 6.4GB/s または 3.2GB/s で転送されるデータはありません。また、6.4GB/s で送信される単一の I/O 要求もありません。帯域幅の定格は、内部ネットワーク（C-HSN または CM-HSN）の個々のパスの総計です。Sun StorEdge 9900 シリーズ内で最高速のパスは 200MB/s で、16 のパスがこの速度で動作するという事です。

これがスケラビリティの本質です。Sun StorEdge 9900 シリーズの既存のワークロードに新たなワークロードを追加した場合、どの作業負荷にも、内部リソースと帯域幅を共有した結果としてのパフォーマンスの劣化が生じません。この点を、以下の一連

のテストで証明します。コントローラのみを共有したときに、ある1つのワークロードが別のワークロードに及ぼす影響を測定することで、このスケーラビリティの現象を説明することができます。

スケーラビリティの比較

この一連のテストでは、今日のエンタープライズ・ストレージ・システムで使用できる伝統的な共有バス・アーキテクチャと Sun StorEdge 9900 シリーズのスイッチト・ファブリック・アーキテクチャを使用したときに1つのアプリケーション・ワークロードが別のワークロードに及ぼす影響を比較します。

この比較の目的は、「速度と供給量」を向上することではありませんが、使用可能なリソースに基づいてアプリケーションを調整し、高性能を実現するために多大な努力が払われています。2つのアプリケーションベースのワークロードを組み合わせることで、ホスト・システムでスループットを測定することにより、1つのワークロードがもう1つのワークロードに及ぼす影響を分析することができます。

テスト構成

テスト構成は、ディスク・ドライブをスケーリングしないことに基づいています。言い換えると、各アプリケーションがそれぞれ固有のレイ・グループに常駐することです。アプリケーションによって共有されるリソースは、ストレージ・コントローラ（キャッシュ、パスなど）と ACP の2つのペア（合計で4ペア）に限られます。比較は、共有バス・ベースのアーキテクチャと、この種類で唯一のものである、スイッチト・ファブリック・アーキテクチャを表す Sun StorEdge 9900 シリーズとの間で行なわれます。

アプリケーション・ワークロードは、オンライン・トランザクション処理（OLTP：On-Line Transaction Processing）システムと意思決定支援システム（DSS：Decision Support System）という非常に異なる2種類のデータベース環境から構成されています。OLTP ワークロードは、両方のストレージ・システムが次の構成になっている、他社プラットフォームで実行されます。

- 他社製デュアル・プロセッサ・サーバ 180 MHz、RAM 768MB、7700E と Sun StorEdge 9900 シリーズのいずれにも4基のファイバ・チャンネル HBA（FC-AL トポロジ）を直接接続。
- Oracle® Database Server 8.0.4 50MB の SGA（ストレージ・レイへの I/O の駆動用）付属、データベース・ブロック・サイズ 2KB。
- OLTP データベース（TPC-C ベンチマークと同様の）擬似ユーザ・ドライバ付属。
- 共有バス・アーキテクチャ型ストレージおよび Sun StorEdge 9900 シリーズ 1GB キャッシュ、ファイバ・チャンネル（FC-AL トポロジ）ポート4基、2つの ACP ペア、6GB 12,030RPM ドライブ、RAID-5+ フォーマット済み。OLTP ワークロードでは、共有バス・アーキテクチャ型ストレージと 9900 の間で次の点が異なります。
 - 共有バス・アーキテクチャ型ストレージでは、合計48台のディスク・ドライブ（合計12のレイ・グループ）を使用。
 - Sun StorEdge 9900 シリーズでは、OPEN-9 ボリューム（これにより、テーブルのダブル・アップが発生）、40台のディスク・ドライブ（合計10のレイ・グループ）、および 18GB 10,025 RPM ドライブを使用。

DSS ワークロードは Sun™ プラットフォームで実行され、次の構成になっています。

- 2台の Sun E3000 デュアル・プロセッサ・サーバ それぞれ 250MHz、RAM 512MB、2つの 32 ビット JNI HBA (ポイントツーポイント・ファブリック・トポロジ)、それぞれファイバ・チャンネル・スイッチ (Ancor MKII 16ポート) で接続。
- Sun Cluster 2.1。
- Oracle Parallel Server 8.0.5 データベース・ブロック・サイズ 16KB (256KB I/O ブロック・サイズ)
- 1つの大きなテーブルを持つ DSS データベース それぞれ 112 バイトで 600037899 ロー、複数インスタンスにより完全な並列テーブル・スキャンを実行。
- 共有バス・アーキテクチャ型ストレージと Sun StorEdge 9900 シリーズは、1GB キャッシュ、4つのファイバ・チャンネル (ポイントツーポイント・ファブリック・トポロジ) ポート、4つの ACP ペア、6GB 12,030RPM ドライブ、RAID-5+ フォーマット済みの合計 64 のディスク (16 のアレイ・グループ) を搭載。DSS ワークロードでは、共有バス・アーキテクチャ型ストレージと Sun StorEdge 9900 シリーズの間で次の点が異なります。
 - 共有バス・アーキテクチャ型ストレージは、すべての 6GB ドライブを使用。
 - Sun StorEdge 9900 シリーズは、18GB ドライブと 47GB ドライブを併用。

図 10 と図 11 は、これら 2 つのアーキテクチャの比較に使用したテスト構成の図です。注) 下記のテスト構成は、Hitachi Data Systems のホワイトペーパーをもとにしており、Sun で正式にサポートしている構成とは異なりますのでご注意ください。

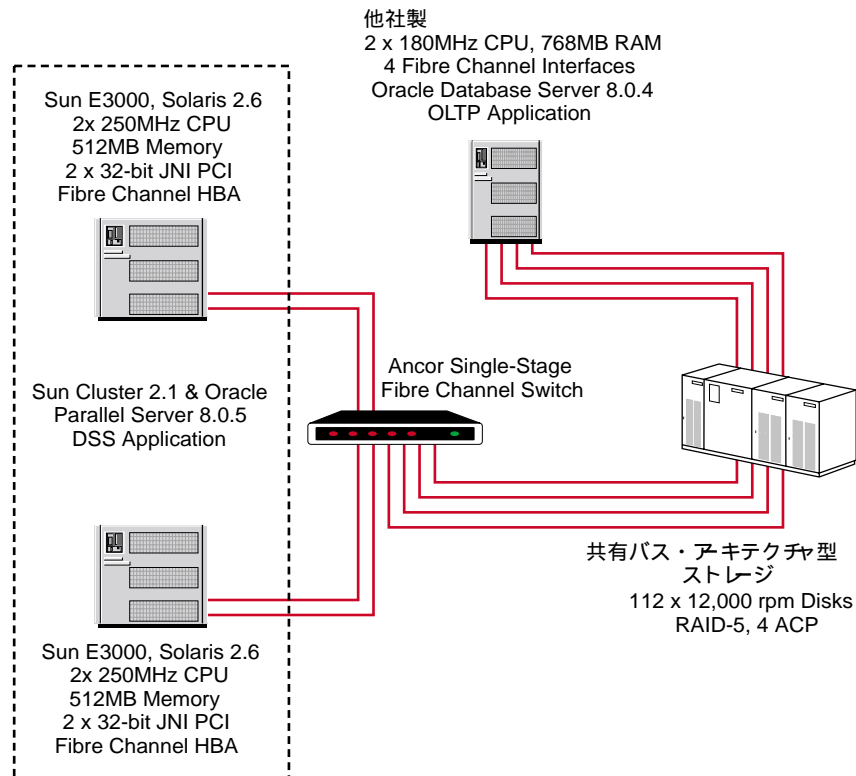


図 10 : 共有バス・アーキテクチャ型ストレージ・スケーラビリティ・テスト環境の構成図

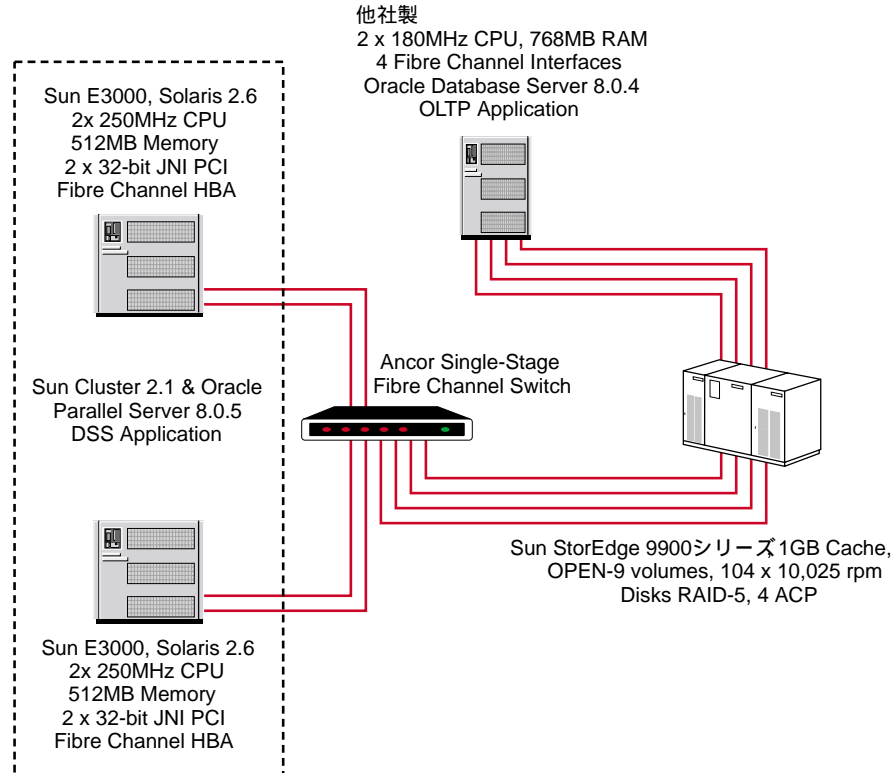


図 11 : Sun StorEdge 9900 シリーズ・スケーラビリティ・テスト環境の構成図

テストの実施明細

テストを実施するための一連のイベントを以下に列記します。

1. すべてのシステムをリブート（Sun Cluster と他社製サーバを起動）した後、OLTP アプリケーションを起動し、ストレージ・システムで3時間動作できるようにします。
2. OLTP テストの間、Oracle チェックポイントをテスト・スクリプトによって最初の10分後に強制実行し、テストが完了するまで60分ごとに強制実行します。
3. OLTP アプリケーションがストレージ・システム上で単独で45分間動作したら、DSS アプリケーションを起動します。
4. DSS アプリケーションは、約3時間にわたって完全なテーブル・スキャンを実行し続けます。

テスト結果 - 共有バス・アーキテクチャ

共有バス・アーキテクチャのスケラビリティ・テストの結果を図12に示します。

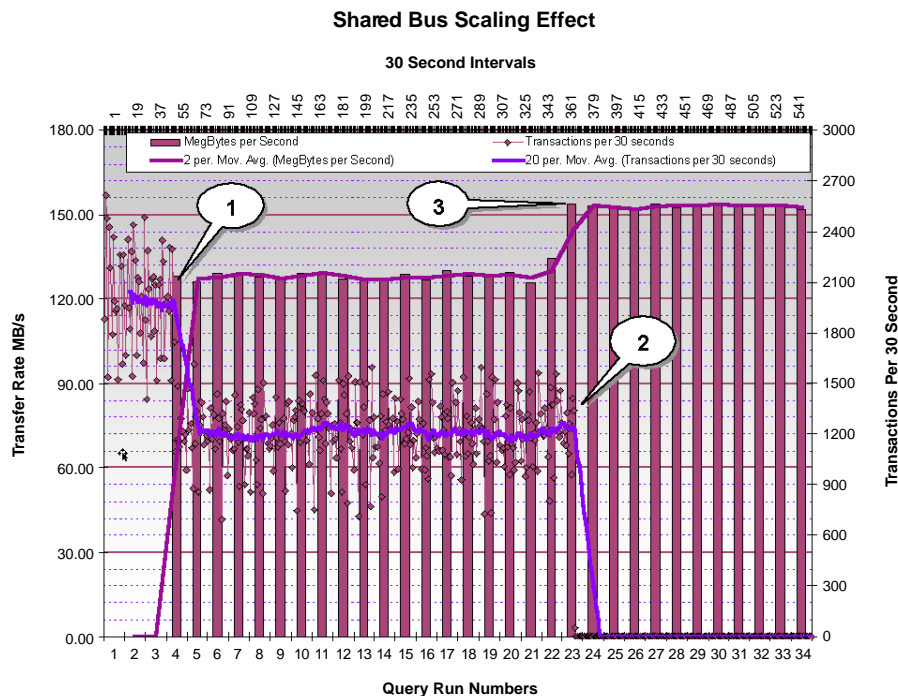


図12：共有バス・スケラビリティ・テストの結果

図12に示されている結果について、次の3つのステップ（チャート上のポイント）で説明します。

1. 約1,900トランザクション/分（TPM：Transactions per Minute）で45分間動作しているOLTPアプリケーションが変化し始め、DSSアプリケーション・ワークロードが開始されます。その後、トランザクション・レートは約1,200TPMに低下します（約63%低下）。この間、DSSアプリケーションは、約128MB/s（毎秒約130万ロー）の転送速度を維持しています。
2. この時点で、OLTPアプリケーションは、テストの複合ワークロード段階の間、1,200TPMの速度を維持しながら完了します。OLTPアプリケーション・ワークロードにおける結果の高低（負荷なしと負荷あり）は、それぞれ1,900TPMと1,200TPMです。
3. OLTPアプリケーションが完了すると、DSSアプリケーションのスループットが128MB/s（毎秒130万ロー）から約153MB/s（毎秒160万ロー）に変化（約20%上昇）し、この速度をテストが終了するまで維持します。スループットの結果の高低（負荷なしと負荷あり）は、それぞれ128MB/s（毎秒130万ロー）と153MB/s（毎秒160万ロー）です。

ポイント1とポイント3でのパフォーマンスの低下は、共有バス設計をコントローラで使用したときに、1つのワークロードが別のワークロードに及ぼす影響を表しています。共有バス・アーキテクチャ型ストレージは、業界全体を通じて、スケラビリティに優れたストレージ・アレイであると見なされているのが普通ですが、この種の設計には限界があります。さらにまた、このアーキテクチャを採用しているストレージ・アレイはいずれも同じような動作を示します。

テスト結果 - Sun StorEdge 9900 シリーズのHi-Star アーキテクチャ
 Sun StorEdge 9900 シリーズの内部Hi-Star アーキテクチャのスケラビリティ・テストの結果を図13に示します。

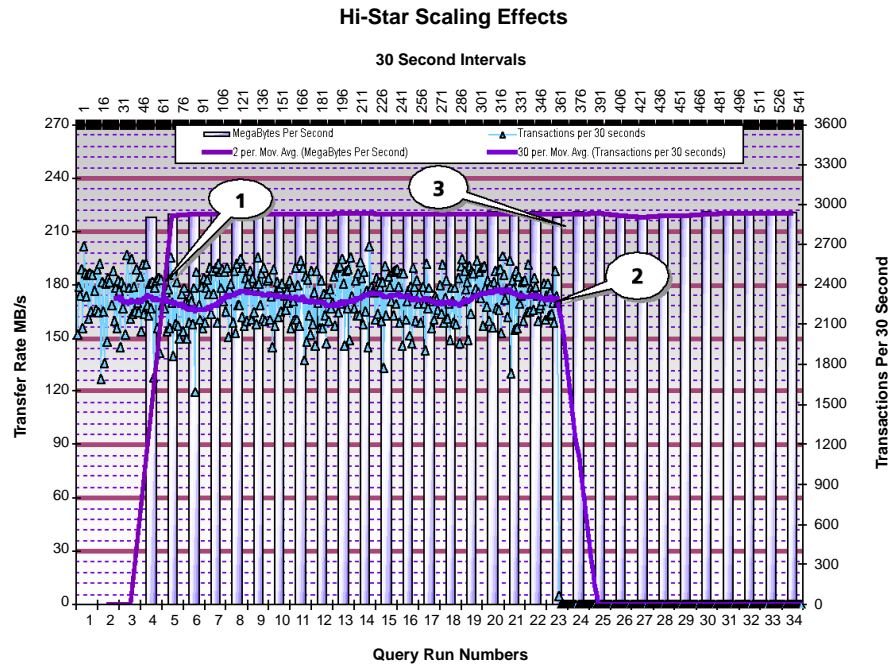


図13：Sun StorEdge 9900 シリーズのHi-Star スケラビリティ・テストの結果

図13に示されている結果について、次の3つのステップで説明します。

1. これは、DSS アプリケーションがその完全なテーブル・スキャンの実行を開始し、同時にOLTPアプリケーションが単独で約45分間動作してきたポイントです。OLTPアプリケーションは、約2,350TPMで開始し、DSSアプリケーションが開始してから終了するまでこの速度を維持します。DSSアプリケーションは、約220MB/s (毎秒230万ロー以上) で開始します。TPMは低下しません。
2. OLTPアプリケーションのテストは、約2,350TPMのトランザクション・レートを維持したまま終了します。そのため、結果の高低(負荷なしと負荷あり)はいずれも2,350TPMです。言い換えると、DSSアプリケーションは、図12の共有バス・コントローラの結果とは違って、OLTPアプリケーションに影響しないということです。
3. OLTPアプリケーションが完了した(約3時間)後、DSSアプリケーションは220MB/s(毎秒230万ロー以上)の転送速度を維持します。この転送速度は、DSSアプリケーションのテスト段階を通じて、一貫して変わりません。ここでも、MB/s(ロー数/秒)は低下しません。

これが従来の共有バス・ストレージ・アレイであれば、ワークロードを追加したポイント1とポイント3でTPMは低下したでしょう。しかし、Sun StorEdge 9900 シリーズのHi-Starアーキテクチャでは、各ワークロードが本質的に独自の内部ルート・パスを持つため、従来の帯域幅制限が適用されません。キャッシュ・レベルでは一定の競合が生じることもありますが、キャッシュには、16の同時オペレーションが並行して存在します。この内蔵バス・フェイルオーバは、たいいていのアプリケーションの要求をはるかに上回り、今日の複数共有バス・アーキテクチャとの対比で測定することができます。

共有バス・アーキテクチャとHi-Starアーキテクチャのテスト結果の比較
共有バス・アーキテクチャとスイッチト・ファブリック・アーキテクチャの2つのテスト結果を図14に示します。繰り返しますが、このテストは、「速度と供給量」の性能数値を得ることを目的としていませんでしたが、Sun StorEdge 9900 シリーズの機能からは、同じテストで共有バス・アーキテクチャよりも高い数値が得られています。

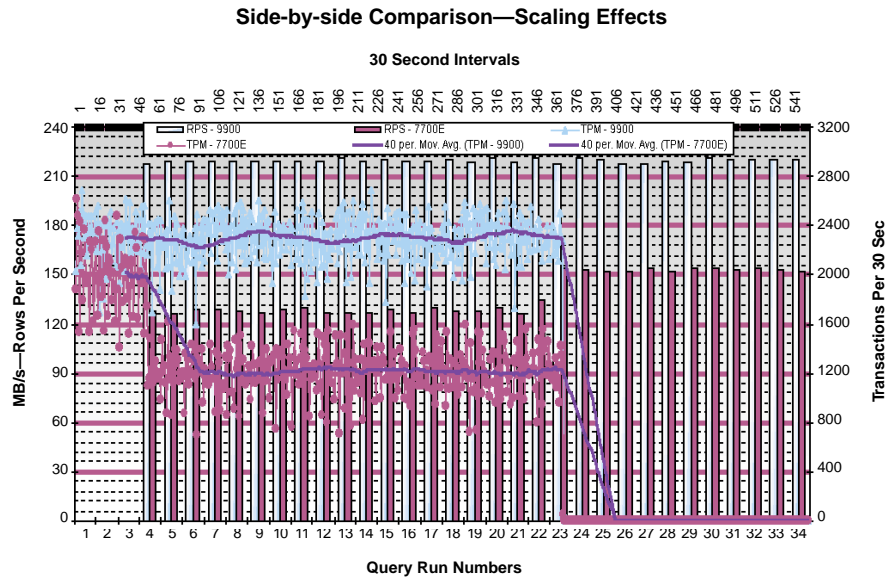


図14：スケーラビリティ・テスト結果の比較

結論

本書の詳細なテスト結果と分析から、新しいSun StorEdge 9900 シリーズ・インテリジェント・ストレージ・システムが、今日、業界一のスケーラビリティを備えた最強のエンタープライズ・システムであるという確証が得られます。Sun StorEdge 9900 シリーズで採用されている新しいHi-Starスイッチト・ファブリック・アーキテクチャは、混合アプリケーションを実行しているオープン・システム環境の苛酷なI/O要求に対応するものとして特別に設計されています。その革命的な内部アーキテクチャと進化した機能により、Sun StorEdge 9900 シリーズはまさしく他に類を見ないものです。

用語集

ACP	アレイ・コントロール・プロセッサ 4つのFC-ALを通じたディスク・ドライブへの物理的アクセスを実行する、PCB上のプロセッサのグループ。
ACPペア (ACP Pair)	2つのACPの組み合わせ。冗長性とパフォーマンスを向上するように設計されています。
アレイ・グループ (Array Group)	使用される RAID レベルに依存しない、ディスク・ドライブの物理的配列。RAID-5+ では、3データ+1パリティ、RAID-1+ では、2データ+2データ。
バックエンド (Back-end)	ストレージ・アレイに関して、バックエンドは、コントローラ、ディスク・ドライブ、ディスク・ドライブへのパスから構成され则认为られています。Sun StorEdge 9900 シリーズのバックエンドは、ACPペア、FC-AL、ファイバ・チャンネル・ディスクから構成されています。
CA	キャッシュ・アダプタ。
CARB	キャッシュ・メモリ・アービトラータ回路。Hitachi が設計した、キャッシュへのアクセスを調停するためのCSW内のロジック。
CHIP	クライアント・ホスト・インターフェイス・プロセッサ。
C HSN	キャッシュ 階層型スター・ネットワーク。
CHIP	クライアント・ホスト・インターフェイス・プロセッサ。
CMA	コントロール・メモリ・アダプタ。
CM HSN	制御メモリ 階層型スター・ネットワーク。
CSW	キャッシュ・スイッチ Sun StorEdge 9900 シリーズ用として特別に設計されたクロスバー・スイッチ。ファイバ・チャンネル・スイッチではありません。
CTQ	コマンドタグ・キューイング。
DRR	データ復旧再生。
DSS	意思決定支援システム。一般にデータ・ウェアハウスとデータ・マートによって実現されます。
DTA	データ・アダプタ。
ファブリック (Fabric)	スケーラビリティと接続性を向上するように配列された、スイッチのグループまたはネットワーク。
FC-AL	Fibre Channel-Arbitrated Loop。
フロントエンド (Front-end)	ストレージ・アレイに関して、フロントエンドは、「実世界」へのインターフェイスまたはポート、そのポートを使用するプロセッサ、さらに場合によってはキャッシュ・メモリから構成され则认为られています。Sun StorEdge 9900 シリーズのフロントエンドは、CHT から構成されています。
Hi-Star	Sun StorEdge 9900 シリーズの階層型スター・ネットワーク・アーキテクチャを示すマーケティング用語。

HSN	階層型スター・ネットワーク。Sun StorEdge 9900 シリーズの複合内部ネットワーク (C-HSN および CM-HSN) を示す専門用語。
Sun StorEdge 9900 シリーズ	世界一の速度とスケーラビリティを備えたストレージ・システム。
メタデータ (Meta-data)	データに関するデータ。CM-HSM と制御メモリ・サブシステムによって使用されます。
MP	マイクロプロセッサ。
MPA	マイクロプロセッサ・アダプタ。
OLTP	オンライン・トランザクション処理。
POD	Performance On Demand。
PCB	プリント基板。
QOS	Quality of Service。
SOD	Storage On Demand。