

ストレージ・エリア・ネットワークを
最適化する Sun StorEdge™ 9900 シリーズ



目次

スケーラブルなストレージ・エリア・ネットワークの必要性	1
Sun StorEdge™ 9900 シリーズ：世界初の真にスケーラブルなストレージ・システム	2
所有コストの削減	3
Hi-Star™ アーキテクチャ	3
キャッシュ・アーキテクチャ	4
ディスク・アレイ・アーキテクチャ	5
ディスク技術	6
Sun StorEdge 9900 シリーズのシステム容量	7
SANの管理	8
SANの拡張機能とSWAN	8
ファイバ・チャンネル・エクステンダ	7
ESCONチャンネル・エクステンダ	8
SWAN用のビジネス継続性ソリューション	8
Sun StorEdge 9900 True Copy	8
Sun StorEdge 9900 Shadow Image	8
まとめ	9

ストレージ・エリア・ネットワークを最適化する Sun StorEdge 9900 シリーズ

「今日の組織にとって、最も重要な問題はスピードです。つまり、顧客データ、製品データ、価格データ、販売データに、社内と社外のどちらからも迅速、安全にアクセスできなければなりません。スケーラブルなセキュア・ストレージ管理アーキテクチャがあれば成功が保証されるわけではありませんが、なければ生き残れないかもしれません」

- Robert Scheier、Computer World、2000年5月1日

スケーラブルなストレージ・エリア・ネットワークの必要性

今日のIT管理者が抱える最大の問題は、ノンストップの可用性と、コンピューター、場所、時間を問わないデータへの瞬時アクセスを提供しつつ、データの需要の増大に合わせてストレージ・システムを拡張する方法を探ることです。ニュー・エコノミーによって洪水のように情報があふれ、インターネット時代の到来で機敏性の要求が高まり、技術の急速な進歩もその需要にますます拍車をかけるばかりです。

IT管理者がこの問題に立ち向かうただ1つの手段は、サーバ接続ストレージからスケーラブルなストレージ・エリア・ネットワーク（SAN：Storage Area Network）に、ストレージ・パラダイムを切り替えることです。SANは、サーバからストレージを切り離し、需要に合わせて両方を独立して拡張できるようにします。ストレージ・リソースを有効に利用するための共有ストレージ・プーリングから、パフォーマンスの向上、セキュリティの強化、集中管理、高可用性に至るまで、SANにはさまざまな利点があります。

たとえば、接続されているすべてのサーバとストレージ・ノードの間で100MB/sのFC-ALファイバ・ケーブルを共有した場合、100MB/sの通信を一度に1つしか行うことができません。共有FC-ALループにデバイスをさらに追加すると、アービトレーションやループへのアクセスのためにデバイス間での競合が増え、パフォーマンスが低下します。ファイバ・チャンネル・スイッチ技術は、複数の100MB/sの非ブロック・パス接続を可能にすることによって、こうした状況をすっかり変えました。スイッチにポートを追加すると、帯域幅の総計が増え、それぞれのパスが100MB/sの通信を提供します。これこそが、スケーラビリティの本質です。

スイッチを使用してスケーラビリティを高めるという概念は、新しいものではありません。電話業界は、需要の増大に対応するためにクロスバー・スイッチに移行しました。ネットワークベンダーは、LAN/WANネットワークのスケーラビリティを高めるスイッチとルータを開発しました。SANのスケーラビリティを高めるために、Ancor、BROCADE、McDATAなどのベンダによってファブリック・スイッチが導入されまし

た。現在、サーバ・ベンダはInfiniband™のための標準を制定中です。この新しいアーキテクチャは、サーバの内部共有PCIバスの代替として、複数のマイクロプロセッサ、メモリ・モジュール、I/Oプロセッサを接続するスケーラブルなスイッチを実現することを目的としています。Infiniband 標準をベースにした製品は、向こう数年の間に利用可能になる見込みです。Sun Enterprise™ 10000 では独自のクロスバー・スイッチ・アーキテクチャをすでに実装しています。

こうした努力はすべて、ネットワーク、SAN、将来のサーバのスケーラビリティ向上を目的としています。今までこのデータ食物連鎖の中で見過ごされてきた1つの重要なリンク、それがストレージ・システムでの共有バス・アーキテクチャです。サンが発表した Sun StorEdge 9900 シリーズは、この欠点を解消しています。

Sun StorEdge 9900 シリーズ：世界初の真にスケーラブルなストレージ・システム

スイッチト・ファブリック SAN は、共有 SCSI バスのボトルネックを解消しますが、一方でストレージ・システムにこれまでにない負担をかけます。スイッチによって、多数のホストが個々のストレージ・ポートにアクセスできるようになります。ストレージ・ポートにホストが追加されると、I/O レートが増大し、アクセス・パターンが分散するようになります。さらに、内部共有バスが過負荷状態になり、キャッシュ・ヒット率が低下して、バックエンドのディスクに多くの I/O 要求が直接送られる結果になります。

e-business 業界向けに設計された Sun StorEdge 9900 シリーズは、IT コストを劇的に低減させると同時に、ノンストップのスケーラブルなパフォーマンスと容量を提供し、スケーラブル・スイッチングの概念をストレージ・システムに導入した革新的なアーキテクチャにより、SAN 環境の需要を満たします。スイッチト SAN インフラストラクチャと Sun StorEdge 9900 シリーズを組み合わせることによって、一貫したハイパフォーマンスを維持しながら、きわめて大規模なストレージを多数のオープン・システムやメインフレーム・サーバで統合できます。

SAN 対応をうたう競合ストレージ・システムでも、実際にはホストが追加されたりワークロードが増大したりすると速度が低下します。SAN 内でファイバ・チャネル・スイッチを使用しているにもかかわらず、コントローラ内では共有内部バスを使用して、ストレージ・ポートをキャッシュ・メモリやバックエンドのディスク・アレイと接続しています。こうしたストレージ・コントローラは、各ストレージ・ポートにたった1台のサーバを接続していた時期に設計されたもので、各ポートに数十台のサーバがアクセスする現在の SAN 環境にはふさわしくありません。

所有コストの削減

SAN がコスト削減に役立つ大きな要因は、ストレージ・リソースのプールに多数のサーバを接続する能力です。IDC などの業界アナリストの調査によると、直接接続されたディスク・スペースの 50% 以上は使用されていません。SAN 上でストレージ・リソースをプーリングすることによって、複数のサーバが未使用のディスク・スペースを共有できます。ディスク・スペースの使用率が 50% から 75% 以上に増えれば、経費にその効果がすぐに現れます。e-business は、多数の高性能プロセッサを使用して Web サービスの需要をサポートしますが、これらの Web サーバはすべて同じデータ・プールにアクセスする必要があります。SAN は、このデータ・プーリングを容易にします。さらに Sun StorEdge 9900 シリーズは、フロントエンドのネットワークからバックエンドのディスクまで、スイッチの完全なスケーラビリティを実現することによって、SAN のスケーラビリティを高めます。

キャッシュ・アーキテクチャ

Hi-Star は、キャッシュのパフォーマンスと容量を高めるためのコネクティビティを備えています。ますます多くのユーザが大量のデータにアクセスする e-business 業界では、一層大容量のキャッシュが必要になっています。さらにこのキャッシュは、複数の並行アクセスに十分対応できるスケーラブルな構成をサポートしていなければなりません。共有バス・アーキテクチャにキャッシュ・モジュールを追加してもバスの競合が悪化するだけですが、Hi-Star スイッチに複数のキャッシュ・モジュールを取り付ければ、より大きな容量と帯域幅が得られます。Sun StorEdge 9900 シリーズは、最大 32GB のデータ・キャッシュを提供します。この容量は、Hi-Star スイッチを経由してストレージ・ポートとディスク・アレイ・ポートに交差接続される、4 つのキャッシュ・モジュール（基本 2、拡張 2）から成ります。すべての書き込みはキャッシュ・ヒットとして扱われ、ディスクに書き込まれるまでの保護のために、2 つの別個のキャッシュ・モジュールにミラーリングされます。

Sun StorEdge 9900 シリーズは、業界で最もインテリジェントなキャッシング・アルゴリズムを使用して、システム・パフォーマンスを劇的に高めます。レコード・レベル・キャッシングとともに、インテリジェント学習アルゴリズムが備わっていますが、このアルゴリズムは、データのアクセス・パターンを監視して、ランダムにアクセスされるデータのヒット率ができるだけ高くなるように、ステージング基準を動的に更新します。シーケンシャル・アクセスが検出されると、「先読み後削除」アルゴリズムが使用されて、他の要求のためにキャッシュを解放しながら、高速なシーケンシャル・パフォーマンスが得られます。

キャッシュは、データ・アクセスの最適化に使用されるだけでなく、アドレッシング・テーブルなど、制御情報へのアクセスの最適化にも使用されます。このことは特に、ストレージ・ポートとディスク・アレイ間の「any-to-any」接続を実現するグローバル・キャッシュ・システムでは重要です。制御ストアへのアクセスとデータへのアクセスの競合を防ぐために、Sun StorEdge 9900 シリーズは制御情報用に別個の冗長なキャッシュ・モジュールのセット（1.5GB）を提供しています。この追加キャッシュ・メモリは、データに割り当てられる最大 32GB のキャッシュには含まれません。他のベンダは、データと制御情報に同じキャッシュを使用しています。キャッシュ・サイズの小さいアクティブ・システムでは、キャッシュ容量の 50% もの部分が制御情報に占有されることがあります。Sun StorEdge 9900 シリーズの多重キャッシュ・モジュール、分離した制御メモリ、スイッチ接続、そしてインテリジェントなキャッシング・アルゴリズムが、パフォーマンスと可用性の高いスケーラブルなキャッシュ・アーキテクチャを実現し、Hi-Star スイッチ・アーキテクチャを補完しています。

キャッシュ・メモリは揮発性メモリなので、電源が失われるとキャッシュの内容は消えます。電源障害の発生時に、電源復旧までキャッシュ・メモリ内のデータを保持するために、Sun StorEdge 9900 シリーズは 48 時間のバッテリー・バックアップを用意しています。このバッテリー電源は冗長性があり、障害が 48 時間を超える場合にはホットプラグ可能です。一部のストレージ・ベンダは、電源異常時にバッテリー電源を使用して、キャッシュ内のデータをディスクにダウンロードする方式をとっていますが、電源ダウンや継続的な電源障害の発生時にデータを移動するよりも、キャッシュ内のデータを凍結した方が安全で、二重の書き込みキャッシング、バッテリー・バックアップ、複数のスイッチ・キャッシュ・モジュール、データ用と制御用のメモリ・バンクの分離により、Sun StorEdge 9900 シリーズは最高の信頼性と最高のパフォーマンスを兼ね備えたスケーラブル・キャッシュを提供します。

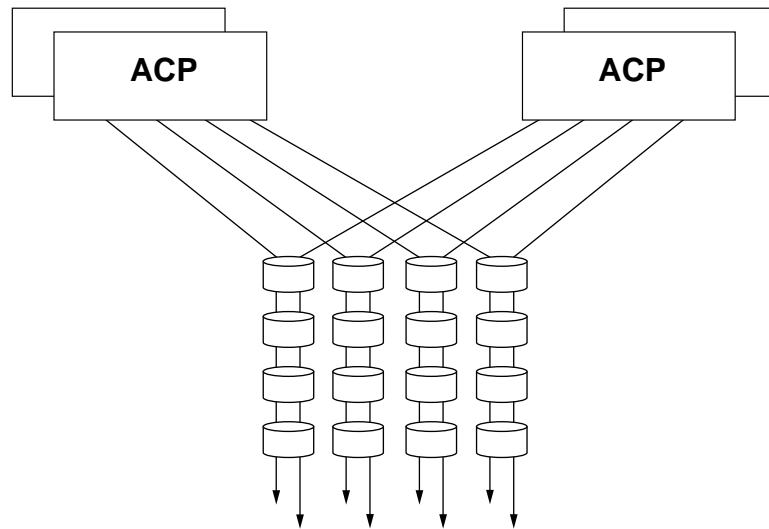
ディスク・アレイ・アーキテクチャ

コストの低い大容量ストレージの需要に応えるには、大容量のディスクが必要です。ディスク技術をリードする Sun StorEdge 9900 シリーズには、18GB、73GB の 10,025RPM ディスクを搭載しています。スケーラビリティのための拡張機能がなければ、ディスク容量の増大はボトルネックになる可能性があります。コントローラに効率の高い大容量キャッシュが備わっていても、データ・アクセスのパフォーマンスは最終的にはディスク・アレイによって決まります。

Sun StorEdge 9900 シリーズのディスクは、ファイバ・チャンネル・ディスクを改良したもので、100MB/s の二重アクティブ・ファイバ・チャンネル・ループを介してディスク・アレイに接続されます。一方、ファイバ・チャンネル・ディスクはポートを1つしか備えておらず、二重ファイバ・チャンネル・ループへの接続に、ループを一度に1つしか利用できないバイパス・スイッチを介しています。ACPのペアにより、8つのファイバ・チャンネル・ループを最大32台のディスクに同時接続できます。デュアル・ポートのファイバ・チャンネル・ディスクが、2つのACPに交差接続されます。

Sun StorEdge 9900 シリーズは4ペアまでのACPをサポートでき、ペアごとに8つのファイバ・チャンネル・ポートがあるので、合計32の100MB/s同時データ転送がサポートされます。

Four 100 MB/s Fibre loops per ACP



A logical disk is striped across four physical disks.

図2：Sun StorEdge 9900シリーズのディスク・アレイ構成

冗長性とパフォーマンスを高めるために、ファイバ・チャンネル・ループの別々のペアに接続した4つのディスク間で、論理デバイスがストライプ化されます。さらにこれらのループは、2つのACPに交差接続されます。これにより、論理デバイスは100MB/sの4倍、つまり400MB/sの帯域幅を使用できます。データをRAID-5構成でストライプ化した場合は、ドライブの1つにパリティが格納されるので、使用可能なデータ帯域幅は300MB/sになります。データをRAID-1構成(2つのディスクが他の2つのディスクにミ

ラーリングされる)でストライプ化した場合、使用可能なデータ帯域幅は200MB/sです。RAID-1を使用すると、先にデータの転送準備ができたRAID-1ペアに対してアクセスが行われるので、ランダムI/Oのパフォーマンスが向上します。

ディスク技術

ディスクのパフォーマンスを決定する基本パラメータは、I/Oキュー時間、シーク時間(読み取り/書き込み部をディスク上のターゲット・トラックに移動するために要する時間)待ち時間(トラックのターゲット・セクターを読み取り/書き込み部まで回転するために要する時間)およびデータ転送時間の4つです。Sun StorEdge 9900シリーズは、これらの数値を革新的な方法で低減しています。ディスクにある2つのポートはそれぞれ、4つまでの要求をキューに入れることができます。このサポートにより、ディスクが1回転する間に複数の要求にサービスを提供できるように要求をソートできます。従来のディスク・アレイは、コントローラのレベルで要求をキューに入れ、ディスクに一度に出す要求は1つだけでした。

シーク時間は、読み取りについては5.7ms、書き込みについては6.5msに短縮されました。研究によれば、シークの80%はいわゆる「ゼロ・シーク」で、読み取り/書き込みアームの移動を必要としません。シーク時間は統計の平均値です。Sun StorEdge 9900シリーズを使用してRAID-1を実装すると、2組の読み取り/書き込みアームを使用して前述のターゲット・トラックがシークされるので、この平均値を半減させることができます。さらに、ディスク・レベルでのキューイングも、シーク間の距離が最小になる順序に要求をソートすることによって、シーク時間の短縮に役立っています。

他の2つのパフォーマンス・パラメータ、待ち時間とデータ転送は、I/Oでは必ず生じるもので、ディスクの回転速度に依存します。10,025RPMのディスクを備えたエンタープライズ・ストレージ・システムを提供しているのは、Sun StorEdge 9900シリーズだけです。つまり、Sun StorEdge 9900シリーズのディスクは、現時点で最短の待ち時間、最速のデータ転送速度を実現しています。

データ転送速度は、ターゲット・トラック上のビットの記録密度にも依存し、密度が高いほどビット転送速度は速くなります。データは、ディスク上の3つのバンドに記録されます。ビット密度は記録バンドの円周によって決まり、円周が長いほど、1回転の間に転送できるデータの量は多くなります。Sun StorEdge 9900シリーズの73GBディスクは50MB/sを超える速度でデータをストリーミングします。これに比べ、高性能ディスクの業界平均は約13~15MB/sです。これらのストリーミング・データ転送速度は、ディスクの面からディスク・バッファまでのデータ転送速度を表しています。バッファからアレイ・コントローラまでは、データは100MB/sに近いファイバ・チャンネル速度で移動します。

Sun StorEdge 9900 シリーズのシステム容量

最大構成	制御1、アレイ・フレーム6
フロント・エンド・ポート数	並行ファイバ32基
内部アーキテクチャ	Hi-Star スイッチ・ファブリック
内部帯域幅	6.4GB/s
最大容量	37テラバイト
最大キャッシュ	32GB
データ論理デバイス数	4096

SANの管理

SAN内のストレージ・システムの管理要件は、共有バス接続ストレージの管理要件とは異なります。SAN内のシステムの第1の管理要件は、ネットワーク内の単一制御ポイントからストレージ・システムを管理できることです。Sun StorEdge 9900シリーズは、IPアドレスとSNMP (Simple Network Management Protocol) エージェントを備えています。SNMP エージェントは、Computer Associates の Unicenter TNG[®]、Tivoli[®] Management Suite、HP OpenView[®]、VERITAS[®] などの管理プログラムによる、アウト・オブ・バンドLANネットワーク上での管理をサポートしているため、Sun StorEdge 9900シリーズのディスクバリ、SAN内でのマップ、高水準管理ソフトウェアによるアラートの処理が可能です。

SANの第2の要件は、複数のホストによる共有が可能であることです。SAN内のストレージ・システムは多数のホストによって共有されるので、データ接続を経由したインバンド管理では、SANのストレージ・システムのうち特定のホストに直接接続されている部分しか見ることができず、システム全体が見えません。Sun StorEdge 9900シリーズは、SNMP エージェントによるアウト・オブ・バンド管理を提供しています。さらにこのエージェントは、ストレージ・ポートの固有の名称など、SAN管理に必要な情報をすべて指定する管理情報ブロック (MIB: Management Information Block) を提供します。

前述のとおり、SANの重要な要件は、それぞれのストレージ・ポートに多数のホストを接続する能力です。共通のストレージ・ポートに接続するすべてのホストは、そのポートに対して設定されたすべてのLUN (論理装置番号) にアクセスできます。各ホストのビューを特定のLUNのサブセットに制限しなければ、この同じアクセス性が大きな混乱を招く可能性があります。Sun StorEdge 9900シリーズは、ユーザがリモート・コンソールから開始できるLUN マスキング機能によって、この問題を解決しています。各ホスト内にあるホスト・バス・アダプタの固有の名称を、このコンソールのグラフィカル・ユーザ・インターフェイス (GUI: graphical user interface) に入力すれば、ユーザはこれらの固有の名称にLUNを個別に割り当てることができます。これにより、Sun StorEdge 9900シリーズのコントローラ内にマスキング・テーブルが作成され、各ホスト・バス・アダプタはストレージ管理者によって割り当てられたLUNだけを「見る」ことができます。

SANの拡張機能とSWAN

ファイバ・チャネルの仕様は、最長10kmのノード間距離をサポートしています。この距離で使用するには、長波レーザー送信機と、シングルモードの9ミクロン・ファイバ・ケーブルを使用する必要があります。現在、大部分のファイバ・チャネル・ストレージ・システムには、最大500m (マルチモードの50ミクロン・ファイバ・ケーブルを使用) の距離をサポートする短波レーザー送信機が付属しています。Sun StorEdge 9900シリーズは、ローカルな用途のための短波送信機と、最長10kmの距離に対応する長波レーザー送信機の両方をサポートします。

短波送信機を備えたストレージ・システムの距離が500mを超える場合は、長波送信機をサポートする中間のファイバ・チャネル・スイッチまたはハブに、ストレージ・システムを接続する必要があります。Sun StorEdge 9900シリーズは、10kmまでの距離に対応するネイティブの長波送信機をサポートできるので、中間のハブやスイッチは不要です。さらにSun StorEdge 9900シリーズは、10kmを超えるストレージ広域ネットワーク (SWAN: Storage Wide Area Network) 距離をサポートしています。SWANは、ファイバ・チャネル・エクステンダとESCONエクステンダによって実現する、長い距離を隔てて相互接続されたSANです。

ファイバ・チャンネル・エクステンダ

Sun StorEdge 9900 シリーズは、エクステンダを使用して、90km までの距離でテストされました。エクステンダは、1本のファイバ・ケーブル上で最大32の光波長を提供し、リング上の両方向で32波長、合計64波長を提供します。また、両方向に同じ32波長を送信することによって冗長性も提供できます。エクステンダは、2Gb/s を超える速度でデータを転送し、ESCON、ファイバ・チャンネル、高速イーサネット、ギガビット・イーサネット、ATMなどの伝送プロトコルをネイティブの速度で組み合わせることができます。また、冗長性を得るために、リング構成内で8つまでのエクステンダと通信することもできます。

ESCON チャンネル・エクステンダ

Sun StorEdge 9900 シリーズは、オープン・システムと S/390 用の ESCON チャンネルを経由した、同期および非同期のリモート・コピー機能も備えています。ESCON チャンネル・エクステンダは、非同期モードで数千キロメートルまでの距離をサポートできます。

SWAN 用のビジネス継続性ソリューション

サンは、SWAN を経由したデータ移動機能を実行するための総合ソフトウェア・ツールも提供しています。このデータ・コピー機能は、ビジネス・インテリジェンス・アプリケーション、リモート・サイトにあるアプリケーションのテスト、あるいは災害時回復などのビジネス継続性のために使用できます。

Sun StorEdge 9900 True Copy

True Copy は、オープン・システム用の同期、もしくは非同期リモート・コピー機能です。True Copy を実行するには、ESCON、またはファイバ・チャンネル・リンクを使用する必要があります。非同期リモート・コピー機能では、ほぼ無制限の距離でリモート・コピーを実行できます。

Sun StorEdge 9900 ShadowImage

ShadowImage は、1台の Sun StorEdge 9900 シリーズ・システム内で最大10、複数の Sun StorEdge 9900 シリーズシステム間で最大20のボリュームのコピーを作成できます。グラフィックまたはコマンド行のインターフェイスが、データのレプリケーションと論理ボリュームの高速再同期を制御します。

オープン・システム環境では、True Copy と ShadowImage を使用して、単独のシステム内で定義された LUN のグループに対して無中断スナップショットを実行できます。このためには、単独の Sun StorEdge 9900 シリーズ・システム内でシーケンシングを使用します。

無中断スナップショットの利点は、数時間かかる復旧時間を分単位に短縮できることです。アプリケーションを停止せずにスナップショットを実行できるので、1日数回ではなく1時間に数回に、スナップショットの頻度を高めることができます。その結果、ログを適用して最終スナップショットから復旧するための時間が大幅に短縮されます。

まとめ

Sun StorEdge 9900 シリーズとストレージ・エリア・ネットワークを組み合わせることによって、きわめて変動の激しい企業の要件に適合する速度とスケーラビリティが得られます。Sun StorEdge 9900 シリーズに導入された2つの新技術は、Hi-Star スイッチ・アーキテクチャと、デュアル・ポートのファイバ・チャンネル・ディスク・アレイです。Hi-Star スイッチ・アーキテクチャは、ストレージ・コントローラの共有バスの制約を解消し、ストレージ・システムの真のスケーラビリティを実現します。デュアル・ポートのファイバ・チャンネル・ディスク・アレイにより、ますます大容量化するディスクをスケーラブルに活用できます。SAN 管理と SWAN 機能に関しては、災害発生時の迅速な復旧を可能にする ShadowImage ボリュームの無中断スナップショットなど、全機能が Sun StorEdge 9900 シリーズに搭載されています。

SAN 特有の要件と e-business の需要をサポートするために根本から設計された業界初のストレージ・システム、Sun StorEdge 9900 シリーズは、これからのストレージ・システムの標準となることでしょう。