IBM @server BladeCenter および Topspin InfiniBand スイッチ・テクノロジー

- InfiniBand スイッチ機能を追加する eServer BladeCenter
- スループットの向上とコストの削減に役立つ構成
- Topspin VFrame および Tivoli Intelligent Orchestrator のデモンストレーション

Rufus Credle
Robyn McGlotten
Mark Welch

ibm.com/redbooks

Redpaper

International Technical Support Organization

IBM @server BladeCenter および TopSpin InfiniBand スイッチ・テクノロジー

SG88-8548-00
（英文原典：REDP-3949-00）

お願い : 本書および本書で紹介する製品をご使用になる前に、『特記事項』（vii ページ）に記載されている情報をお読みください。

本書は、BladeCenter、Topspin InfiniBand Switch Module、および Topspin InfiniBand ホスト・チャネル・アダプターに適用されます。

IBM 発行のマニュアルに関する情報のページ
http://www.ibm.com/jp/manuals/

こちらから、日本語版および英語版のオンライン・ライブラリーをご利用いただけます。また、マニュアルに関するご意見やご感想を、上記ページよりお送りください。今後の参考にさせていただきます。
(URL は、変更になる場合があります)

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： REDP-3949-00
International Technical Support Organization
IBM Eserver BladeCenter and TopSpin InfiniBand
Switch Technology

発行： 日本アイ・ビー・エム株式会社

担当： ナショナル・ランゲージ・サポート

第 1 刷 2006 年 7 月

注 : 本書は、GA 前の製品に基づいて作成されているので、製品が一般出荷可能になると適用されない場合があります。最新の情報については、製品の資料または本 Redbook の追加バージョンを参照することをお勧めします。

目次

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、または サービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の 製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの 評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権（特許出願中のものを含む）を保有している場合があります。本書の提供は、お客様にこれらの特許権について 実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。
〒106-0032 東京都港区六本木 3-2-31 IBM World Trade Asia Corporation Licensing

*以下の保証は、国または地域の法律に沿わない場合は、適用されません。*IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態で提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示 もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が 禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものでは ありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部では ありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確証できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者にお願いします。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

著作権使用許諾 :
本書には、様々なオペレーティング・プラットフォームでのプログラミング手法を例示するサンプル・アプリケーション・プログラムがソース言語で掲載されています。お客様は、サンプル・プログラムが書かれているオペレーティング・プラットフォームのアプリケーション・プログラミング・インターフェースに 準拠したアプリケーション・プログラムの開発、使用、販売、配布を目的として、いかなる形式においても、IBM に対価を支払うことなくこれを複製し、改変し、配布することができます。このサンプル・プログラムは、あらゆる条件下における完全なテストを経ていません。従って IBM は、これらのサンプル・プログラムについて信頼性、利便性もしくは機能性が あることをほのめかしたり、保証することはできません。お客様は、IBM のアプリケーション・プログラミング・インターフェースに準拠した アプリケーション・プログラムの開発、使用、販売、配布を目的として、いかなる形式においても、IBM に対価を支払うことなくこれを複製し、改変し、配布することができます。

商標

以下は、IBM Corporation の商標です。

BladeCenter™
DB2®
Domino®
Electronic Service Agent™
Enterprise Storage Server®
eServer™
@server®
@server®
FlashCopy®
IBM®
IntelliStation®
NetVista™
PowerPC®
Redbooks™
Redbooks（ロゴ）™
ServerGuide™
ThinkPad®
Tivoli®
TotalStorage®
xSeries®

他の会社名、製品名およびサービス名等はそれぞれ各社の商標です。

Java およびすべての Java 関連の商標およびロゴは、Sun Microsystems, Inc. の米国およびその他の国における商標または登録商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

Intel、Intel Inside（ロゴ）、および Pentium は、Intel Corporation の米国およびその他の国における商標です。

Linux は、Linus Torvalds の米国およびその他の国における商標です。

Topspin は登録商標です。Topspin ロゴ、TopspinOS、Topspin Switched Computing System、Grid-to-Go、および VFrame は、Topspin Communications, Inc. の商標です。他の会社名、製品名およびサービス名等はそれぞれ各社の商標です。

Mellanox は、Mellanox Technologies, Inc. の登録商標です。InfiniBlast、InfiniBridge、InfiniHost、InfiniRISC、InfiniScale、および InfiniPCI は、Mellanox Technologies, Inc. の商標です。Copyright 2001. Mellanox Technologies. All rights reserved.

他の会社名、製品名およびサービス名等はそれぞれ各社の商標です。

前書き

BladeCenter™ 用の Topspin ソリューションは、IBM® @server BladeCenter シャーシとの 80 GB 接続、RDMA（Remote Direct Memory Access）、および単一の I/O ファブリック上でのシャーシからのクラスタリング、LAN、および SAN トラフィックの統合機能を提供して、スループットの増加とコストの削減を可能にします。

Topspin InfiniBand ソリューションの前段として、本書では、InfiniBand、BladeCenter、および IBM TotalStorage® のテクノロジーについて記述します。その後、Topspin InfiniBand のアーキテクチャー、スイッチ・モジュール、およびホスト・チャネル・アダプター・カード、ならびに Element Manager と Chassis Manager の使用について詳しく説明します。

本書では、Topspin InfiniBand ソリューション・コンポーネントを使用した IBM eServer™ BladeCenter の複数の構成について詳しく説明します。本書は、HPC または大規模エンタープライズ環境においてお客様が個別に BladeCenter InfiniBand ソリューションを構築される際の基礎を提供します。

redbook の作成チーム

本 redbook は、International Technical Support Organization、Raleigh Center で働く世界中から集まった専門家チームによって作成されました。

Rufus Credle は、International Technical Support Organization、Raleigh Center の Certified Consultant I/T Specialist です。彼は、研修を行い、ネットワーク・オペレーティング・システム、ERP ソリューション、音声テクノロジー、高可用性およびクラスタリング・ソリューション、Web アプリケーション・サーバー、パーベイシブ・コンピューティング、ならびに IBM と OEM e-business アプリケーション、つまり IBM eServer xSeries® と BladeCenter システムを実行するすべてのものについての Redbook™ を作成しています。Rufus の IBM でのキャリアには、業務管理と資産管理、システム・エンジニアリング、営業とマーケティング、および IT サービスが含まれています。彼は、Saint Augustine’s College で経営管理の学士号を取得しています。Rufus は IBM に 25 年間在職しています。

Robyn McGlotten は、NC、RTP の IBM eServer BladeCenter Development グループの開発サポート・エンジニアです。彼女は、IBM Personal Computing Division から初めて、IBM に 3 年間在職しています。IBM BladeCenter ネットワーク製品の開発とインプリメンテーションのテクニカル・サポートを行っています。Robyn は、Florida A&M University で電気工学の学位を取得しています。

Mark Welch は、NC、RTP の IBM eServer BladeCenter Development グループの Advisory Developer です。Mark は、IBM Networking Hardware Division、IBM Global Services、および IBM eServer xSeries servers におけるネットワーキングに 15 年以上の経験を有しています。彼の専門分野は、ネットワークのインターオペラビリティーとテストです。彼は、Florida Atlantic University でコンピューター・プログラミングの応用科学学士を取得しています。彼の取得免許には、Cisco Systems Network Associate、Nortel Networks Certified Design Specialist、および Nortel Networks Certified Account Specialist があります。

次の方々による本プロジェクトへの貢献に感謝いたします。

Tamikia Barrow および Jeanne Tucker
International Technical Support Organization, Raleigh Center

Ishan Sehgal, BladeCenter Marketing Manager - Networking
IBM Research Triangle Park NC

Chris LaGrego, Senior Sales Engineer
Topspin Communications Inc. Mountain View CA

Immani Venkat, Fibre Channel Customer Support Engineer
Topspin Communications Inc. Mountain View CA

Ben Eiref, Director of Product Marketing
Topspin Communications Inc. Mountain View CA

Stuart Aaron, Vice-President of Marketing
Topspin Communications Inc. Mountain View CA

Robert Starmer, Manager Corporate Consulting
Topspin Communications Inc. Mountain View CA

Kevin Deierling, V.P. of Product Marketing
Mellanox Technologies, Inc. Santa Clara, CA

Gene Crossley, Director of Field Applications
Mellanox Technologies, Inc. Raleigh, NC

Doug Vassello, BladeCenter Infrastructure Solution Center（BISC）Team Lead
IBM Research Triangle Park NC

Dexter Monk, WW Level3/PFE, Solution Central - BISC Technical Team Lead
IBM Research Triangle Park NC

Robert Jakes, BISC Team Member, BladeCenter Infrastructure Solution Center
IBM Research Triangle Park NC

Bill Holland, BladeCenter Development - I/O（Storage and Networking）
IBM Research Triangle Park NC

Bill Vetter, BladeCenter Architecture & Strategy
IBM Research Triangle Park NC

Chris Verne, Blade Server Development and Operations Manager
IBM Research Triangle Park NC

寄稿のお誘い

2 週間から 6 週間の研修プログラムにご参加ください。個々の製品またはソリューションを扱う IBM Redbook を作成すると共に、最新のテクノロジーを実地体験します。IBM の専門技術者、ビジネス・パートナー、お客様との共同作業です。

皆さんの努力が、製品の受容やお客様の満足度の向上に役立ちます。さらに、IBM 開発研究所での人脈を築き、生産性とキャリアを高めることができます。

研修プログラムの詳細情報、研修プログラムの索引、およびオンラインの申し込みは、次の Web サイトをご覧ください。

ibm.com/redbooks/residencies.html

ご意見の送付方法

ご意見をお寄せください。

Redbooks をできる限り有益なものにしたいと考えています。この またはその他の Redbook に関するご意見を、次のいずれかの方法でお送りください。

- 次の Web サイトにあるオンラインの **Contact us** review redbook

  ibm.com/redbooks

- 電子メールの宛先 :

  redbook@us.ibm.com

- 郵送先 :

  IBM Corporation, International Technical Support Organization
  Dept. HQ7 Building 662
  P.O. Box 12195
  Research Triangle Park, NC 27709-2195

1

要約

InfiniBand は、科学、技術、および金融アプリケーションで見られるハイパフォーマンス・コンピューティング（HPC）要件に対応するクラスターを作成するための業界標準のファブリックとして認められています。InfiniBand 高帯域幅ファブリックにより、クラスター・サーバー間の高速相互接続が可能になり、高速パフォーマンスが有効になります。BladeCenter InfiniBand ソリューションを使用すると、組織は InfiniBand 標準を有効に利用して、費用対効果の高い BladeCenter ブレード・クラスターに基づく HPC ソリューションを作成できます。

Topspin InfiniBand スイッチとドーター・カードを備えた BladeCenter は、高接続帯域幅とノード間の低遅延との強力な組み合わせを実現します。その結果、InfiniBand ソリューションを搭載した BladeCenter クラスターは、単一のハイエンド・モノリシック・サーバーに匹敵するパフォーマンスを、低いコストで提供することができます。

また、InfiniBand ソリューションを搭載した BladeCenter は、分散データベース処理に必要な拡張容易性と高可用性も提供できます。InfiniBand ソリューションを搭載した BladeCenter は、データ・センターをオンデマンド・コンピューティング環境に変換することができます。この環境は、HPC と分散データベース処理と共に、エンタープライズ・ビジネス・アプリケーションに対する費用対効果の高いサポートを提供します。オンデマンド・コンピューティングは、インフラストラクチャーの複雑さを軽減し、管理を単純化し、可用性を高め、リソースの使用率を向上させることができます。

InfiniBand ソリューションを搭載した BladeCenter クラスターは、複数の拡張機能を組み合わせて非常に高速なパフォーマンスを実現します。1 つの BladeCenter シャーシ内に格納された複数のブレード・サーバー全体で、高速かつ効率的な通信を行うための高帯域幅接続を行います。また、最高 40 Gbps のアップリンク帯域幅、10 Gbps 4X ポート 1 つ、および 30 Gbps 12X ポート 1 つを提供して、大規模な非ブロッキング BladeCenter クラスターを構築するために BladeCenter シャーシ全体の高速通信を可能にします。さらに、BladeCenter InfiniBand スイッチ・ソリューションは、よく知られた Message Passing Interface（MPI）などのオープン・スタンダード・プロトコルを使用して、ノード間遅延を 6 マイクロ秒未満に短縮して、パフォーマンスをさらに向上させることができます。

InfiniBand ソリューションを搭載した BladeCenter は、RDMA（Remote Direct Memory Access）を使用して、クラスター内の CPU を通信処理オーバーヘッドから解放します。RDMA は、リモート・プロセッサーの CPU を必要とすることなく、あるサーバー上のプロセスが別のサーバー上のメモリーに直接アクセスできるようにします。この方法で通信処理

をオフロードすると、CPU 時間が解放され、より生産性の高い処理タスクが可能になります。

高帯域幅と低遅延を組み合わせることにより、費用対効果の高いクラスターの作成が可能になり、サーバー・クラスター間だけでなく、サーバー全体での高速プロセス間通信が実現されます。また、高帯域幅と低遅延により、IBM DB2® Parallel Edition および Oracle Real Application Clusters（RAC）などの分散データベースの能力を有効に利用して、低コストのサーバー・プラットフォームに基づいて構築されたクラスターによる高い拡張容易性と可用性を実現することもできます。

数千個のノードへの拡張が容易に実現します。単一の外部スイッチ・モジュールは、非ブロッキング構成で 300 個を超えるブレードを接続できます。さらに、複数のスイッチ・モジュールを組み合わせると、数千個のブレードを伴う非常に大規模な非ブロッキング・クラスターを構築できます。

データ・センターの管理者は、InfiniBand ソリューションを搭載した BladeCenter のサーバー統合およびサーバー I/O 仮想化機能を利用して、高速パフォーマンスを可能にする高帯域幅、低遅延サーバー相互接続機能を備えたオンデマンド・コンピューティング環境を作成できます。このオンデマンド環境では、サーバーのハードウェアを過剰に購入することなく、さまざまなワークロードに対応できます。また、オンデマンド・コンピューティングは、インフラストラクチャーの複雑さを大幅に軽減して、管理を単純化し、スペース所要量、消費電力、必要な冷却機能を減らし、ケーブル配線の煩わしさを軽減します。

Topspin 外部 I/O シャーシ（Topspin 社から入手可能）を、イーサネットおよびファイバー・チャネル・ゲートウェイに接続すると、オンデマンド・コンピューティング環境を拡張して、ストレージおよびローカル・エリア・ネットワーク（LAN）を組み込むことができます。これらのゲートウェイは、システム管理またはインフラストラクチャーを変更することなく、既存の LAN および SAN を備えたハイパフォーマンス InfiniBand サーバー・ファブリックを容易に相互接続できます。単一の外部 I/O シャーシは、LAN と SAN の両方の接続を提供できます。I/O とストレージは、CPU 全体で共用されるので、データ・センター内のアダプター、ケーブル、およびスイッチ・ポートの数を削減できます。さらに、VFrame Manager ソフトウェア（Topspin 社から入手可能）は、ポリシーおよびプロビジョニング・インテリジェンスをソリューションに組み込むので、さらに真のデータ・センター・リソース仮想化を押し進めることができます。これにより、管理者は、中央から BladeCenter クラスター用のサーバー、I/O、およびストレージを管理できるので、大幅に管理を簡素化できます。さらに、管理者は、中央から I/O やストレージ帯域幅だけでなく、サーバーも追加または削除できるので、サービスの提供を中断することなく、オンデマンドで調整できます。

本書（Redpaper）では、InfiniBand ソリューション・コンポーネントを搭載した BladeCenter の構成を詳しく説明します。本書は、お客様が HPC または大規模エンタープライズ環境に固有の BladeCenter InfiniBand ソリューションを構築する基盤になります。

2

IBM eServer BladeCenter および TotalStorage テクノロジー

この章では、IBM @server BladeCenter および IBM TotalStorage をインプリメントして得られるテクノロジーと利点について説明します。

2.1 IBM eServer BladeCenter

BladeCenter は、革新的なモジュラー・テクノロジー、卓越した密度、および可用性を備え、多数の実世界における問題の解決に役立つように設計されています。

サーバー統合を必要とする組織に対して、BladeCenter は、柔軟性の向上、保守の容易さ、コストの削減、人的資源の合理化を図るためにサーバーを一箇所に集めます。新しい e-commerce および e-business アプリケーションのデプロイを必要とする企業は、柔軟性、拡張容易性、および可用性を確保すると同時に、高速化を実現できます。ファイル & プリントやコラボレーションなどのエンタープライズ要件に対して、BladeCenter は、信頼性、拡張のための柔軟性、および費用対効果を提供するように設計されています。可用性の高いクラスタリングを必要とする計算主体のアプリケーションを使用するお客様は、BladeCenter を使用すると、高度な拡張容易性とパフォーマンスを実現することができます。

BladeCenter ファミリーの製品はモジュラー設計を特徴としています。このモジュラー設計は、複数のコンピューティング・リソースを費用対効果の高い高密度格納装置に統合し、次のようなプラットフォームを提供します。

- インストール、デプロイメント、および再デプロイメントの時間を短縮する
- 有用な管理ツールを使用して管理コストを削減する
- 最高レベルの可用性と信頼性を達成する
- XpandonDemand スケールアウト機能を提供する
- 1U ソリューションと比較して、必要なスペースと冷却能力を削減する

BladeCenter 内における Topspin InfiniBand Switch Module の動作について理解を深めるには、引き続き以下の節を読んで BladeCenter のアーキテクチャーを理解することをお勧めします。BladeCenter とそのコンポーネントの詳細については、IBM Redpaper「*The Cutting Edge: IBM @server BladeCenter*」（REDP-3581）をお読みになることをお勧めします。この Redpaper は次の Web サイトで入手できます。

http://www.redbooks.ibm.com/redpapers/abstracts/redp3581.html

5 ページの図 2-1 は、BladeCenter シャーシ、HS40、HS20、および JS20 を示しています。各製品の説明は次のとおりです。

- BladeCenter シャーシ

  BladeCenter は、アプリケーション・サービス提供、ストレージの柔軟性、および長期にわたる投資保護のために最大限のパフォーマンス、可用性、および管理容易性を提供する高密度ブレード・ソリューションです。

- HS40

  HS40 は、4 プロセッサー SMP 機能を必要とするハイパフォーマンス・エンタープライズ・アプリケーション用の 4way ブレード・サーバーです。BladeCenter シャーシは、最大 7 枚の 4way サーバーをサポートし、エンタープライズ・リソース・プランニング（ERP）およびデータベース・アプリケーションに理想的です。

- HS20

  この IBM の効果的な 2way ブレード・サーバーは、サーバーのパフォーマンスを犠牲にすることなく、高密度な設計を実現しています。Domino®、Web サーバー、Microsoft® Exchange、ファイル & プリント、アプリケーション・サーバーなどに理想的です。

- JS20

  JS20 は、64 ビット・コンピューティングを必要とするアプリケーション用の 2way ブレード・サーバーです。計算主体のアプリケーションやトランザクション Web サービス提供に理想的です。

注 : BladeCenter シャーシ用の将来のブレードが開発中です。

図 2-1 IBM eServer BladeCenter とブレード

BladeCenter テクノロジーの詳細については、次の Web サイトをご覧ください。

http://www.ibm.com/servers/eserver/bladecenter/index.html

BladeCenter ストレージ・ソリューション

IBM は、お客様の厳しいビジネス・ニーズに対応するために、BladeCenter 用にさまざまな取り付けやすいテスト済みの大容量ストレージ製品を用意しています。これにより、次のような一連の IBM TotalStorage ストレージ・ソリューション製品の中から選択できます。

- ファイバー・チャネル製品と Storage Area Network
- Network Attached Storage
- エンタープライズ・ストレージ・サーバー ®

IBM TotalStorage は、お客様固有の要件に合わせて設計された、接続済みの保護された完全なストレージ・ソリューションを提供するので、お客様のストレージ環境の管理を容易にし、コストを削減し、ビジネス効率とビジネスの継続性を提供します。

BladeCenter ストレージ・ソリューションの詳細については、次の Web サイトをご覧ください。

http://www.pc.ibm.com/us/eserver/xseries/storage.html

BladeCenter システム管理

お客様の BladeCenter への投資をそのライフ・サイクル全体で最大限に活用するには、高可用性と低コストを保持するための効果的な高性能システム管理機能が必要です。

管理基盤

評価の高い業界標準ベースのワークグループ・ソフトウェアである IBM Director は、xSeries、IntelliStation®、NetVista™、および ThinkPad® ハードウェアの包括的な管理を実現し、コストの削減と生産性の向上に役立ちます。

IBM Director

IBM Director は、高機能なシステム管理用に設計されたハードウェアです。業界で最良のツールを提供し、可用性の向上、資産の追跡、パフォーマンスの最適化、およびリモート保守の使用可能化によって時間と費用を節約することができます。BladeCenter には、IBM Director 4.2 以降を使用してください。

拡張サーバー管理

以下のソフトウェア・ユーティリティーの集合は、高度なサーバー管理機能と最大限の可用性を提供します。

- Server Plus Pack
- Application Workload Manager
- Scalable Systems Manager
- リアルタイム診断
- Electronic Service Agent™
- Tape Drive Management Assistant

デプロイメントとアップデートの管理

IBM デプロイメント・ツールは、サーバーとクライアントを稼働可能にするために必要な面倒な作業を最小限に抑えるのに役立ちます。このようなツールには次のものがあります。

- リモート・デプロイメント・マネージャー
- ソフトウェア配布 Premium Edition
- ServerGuide™
- UpdateXpress

BladeCenter システム管理の詳細については、次の Web サイトをご覧ください。

http://www.ibm.com/servers/eserver/xseries/systems_management/xseries_sm.html

2.2 BladeCenter アーキテクチャー

ここでは、BladeCenter シャーシとコンポーネントのアーキテクチャー設計について詳しく説明します。

2.2.1 ミッドプレーン

7 ページの図 2-2 は、BladeCenter のミッドプレーンを示しています。このミッドプレーンには、冗長機能を提供する、同じような 2 つのセクション（上部と下部）があります。プロセッサー・ブレード（ブレード・サーバー）は、ミッドプレーンの前面に接続されます。それ以外の主要コンポーネントはすべて、ミッドプレーンの背面に接続されます。

プロセッサー・ブレードには 2 つのコネクターがあります。ミッドプレーンの上部セクションに接続されるコネクターと、下部セクションに接続されるコネクターです。それ以外のコンポーネントはすべて、一方のセクションのみ（上部または下部）に接続されます。ただし、

冗長性を確保するために他方のミッドプレーン・セクションに接続できるマッチング・コンポーネントがあります。

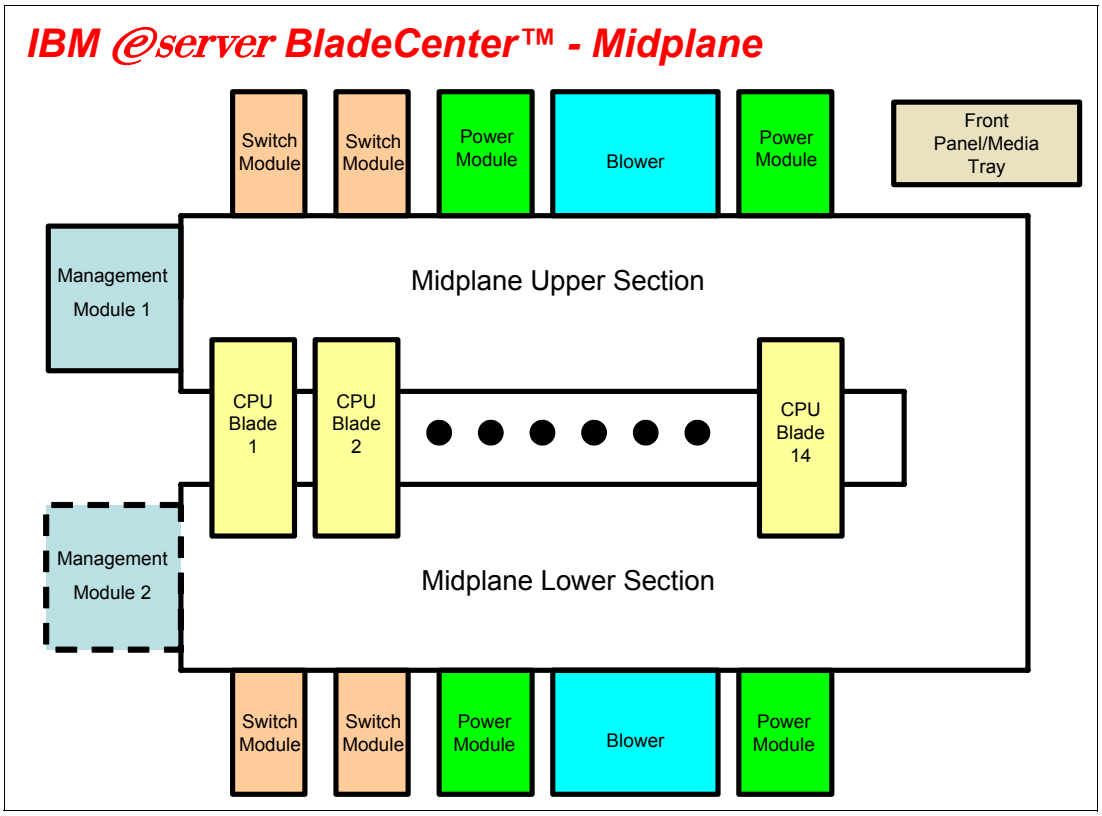

図 2-2　ミッドプレーン図

2.2.2 管理モジュールのイーサネット接続

8 ページの図 2-3 は、管理モジュール（MM）のインターフェースを示しています。スイッチ・モジュールは、100 Mb イーサネット・インターフェースを使用して、アクティブな管理モジュールによって構成されます。各管理モジュールには、スイッチ・モジュールごとに 1 つずつ、4 つの 100 Mb イーサネット・インターフェースがあります。各スイッチ・モジュールには、管理モジュールごとに 1 つずつ、2 つの 100 Mb イーサネット・インターフェースがあります。次のリストは、経路指定を示しています。

- 管理モジュール 1 イーサネット 1 → スイッチ・モジュール 1 イーサネット 15
- 管理モジュール 1 イーサネット 2 → スイッチ・モジュール 2 イーサネット 15
- 管理モジュール 1 イーサネット 3 → 拡張スイッチ・モジュール 3 イーサネット 15
- 管理モジュール 1 イーサネット 4 → 拡張スイッチ・モジュール 4 イーサネット 15
- 管理モジュール 2 イーサネット 1 → スイッチ・モジュール 1 イーサネット 16
- 管理モジュール 2 イーサネット 2 → スイッチ・モジュール 2 イーサネット 16
- 管理モジュール 2 イーサネット 3 → 拡張スイッチ・モジュール 3 イーサネット 16
- 管理モジュール 2 イーサネット 4 → 拡張スイッチ・モジュール 4 イーサネット 16

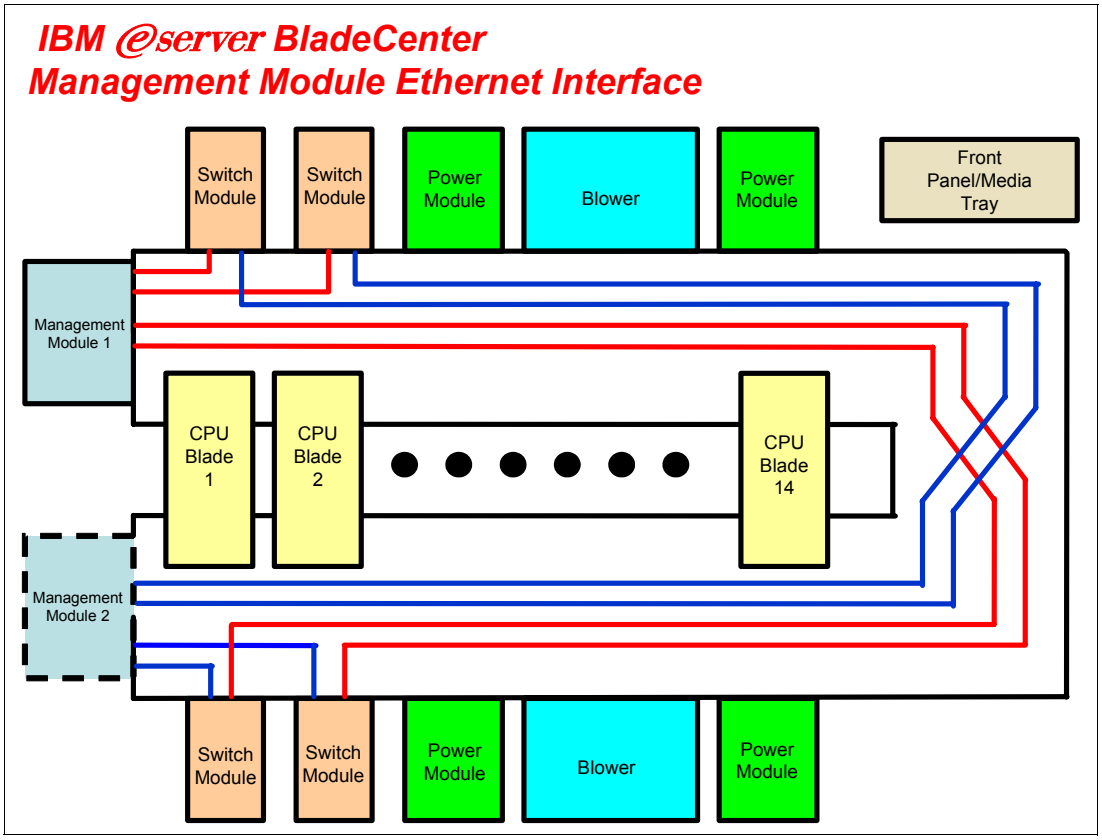

図 2-3　管理モジュールのイーサネット・インターフェース

2.2.3 Gigabit Ethernet パス

9 ページの図 2-4 は、Gigabit Ethernet のパスを示しています。各プロセッサー・ブレードには、最小 2 つから最大 4 つの EtherLAN インターフェースがあります。特に、BladeCenter HS20 プロセッサー・ブレードには、ミッドプレーン・コネクターごとに 1 つずつ、2 つの SERDES ベース Gigabit Ethernet インターフェースがあります。ドーター・カードを取り付けると、さらに 2 つのネットワーク・インターフェースを追加できます。各スイッチ・モジュール（SW Module）は、各プロセッサー・ブレードから 1 つの LAN 入力を受信し、合計で 14 個の入力を受信します。次の部分的なリストは、経路指定を示しています。

- プロセッサー・ブレード 1 LAN 1 → スイッチ・モジュール 1 入力 1
- プロセッサー・ブレード 1 LAN 2 → スイッチ・モジュール 2 入力 1
- プロセッサー・ブレード 1 LAN 3 → 拡張スイッチ・モジュール 3 入力 1
- プロセッサー・ブレード 1 LAN 4 → 拡張スイッチ・モジュール 4 入力 1
- プロセッサー・ブレード 2 LAN 1 → スイッチ・モジュール 1 入力 2
- プロセッサー・ブレード 2 LAN 2 → スイッチ・モジュール 2 入力 2
- プロセッサー・ブレード 2 LAN 3 → 拡張スイッチ・モジュール 3 入力 2
- プロセッサー・ブレード 2 LAN 4 → 拡張スイッチ・モジュール 4 入力 2

プロセッサー・ブレードでは、LAN 1 と LAN 2 は、オンボード SERDES Gigabit Ethernet インターフェースであり、すべてのプロセッサー・ブレードのスイッチ・モジュール 1 とスイッチ・モジュール 2 にそれぞれ経路指定されます。LAN 3 と LAN 4 は、それぞれ拡張スイッチ・モジュール 3 と 4 に進み、ドーター・カードが取り付けられている場合のみ使用されます。1 つ以上のプロセッサー・ブレードにドーター・カードが取り付けられている場合を除いて、スイッチ・モジュール 3 と 4 は必要ありません。さらに、これらのスイッチ・モジュールには、プロセッサー・ブレードによって生成される LAN インターフェースとの互換性が必要です。1 つの BladeCenter HS20 プロセッサー・ブレードにファイバー・チャネ

ル・ドーター・カードが取り付けられている場合、スイッチ・モジュール 3 および 4 もファイバー・チャネル・ベースでなければなりません。また、残りの BladeCenter HS20 プロセッサー・ブレードに取り付けられているすべてのドーター・カードは、ファイバー・チャネルでなければなりません。

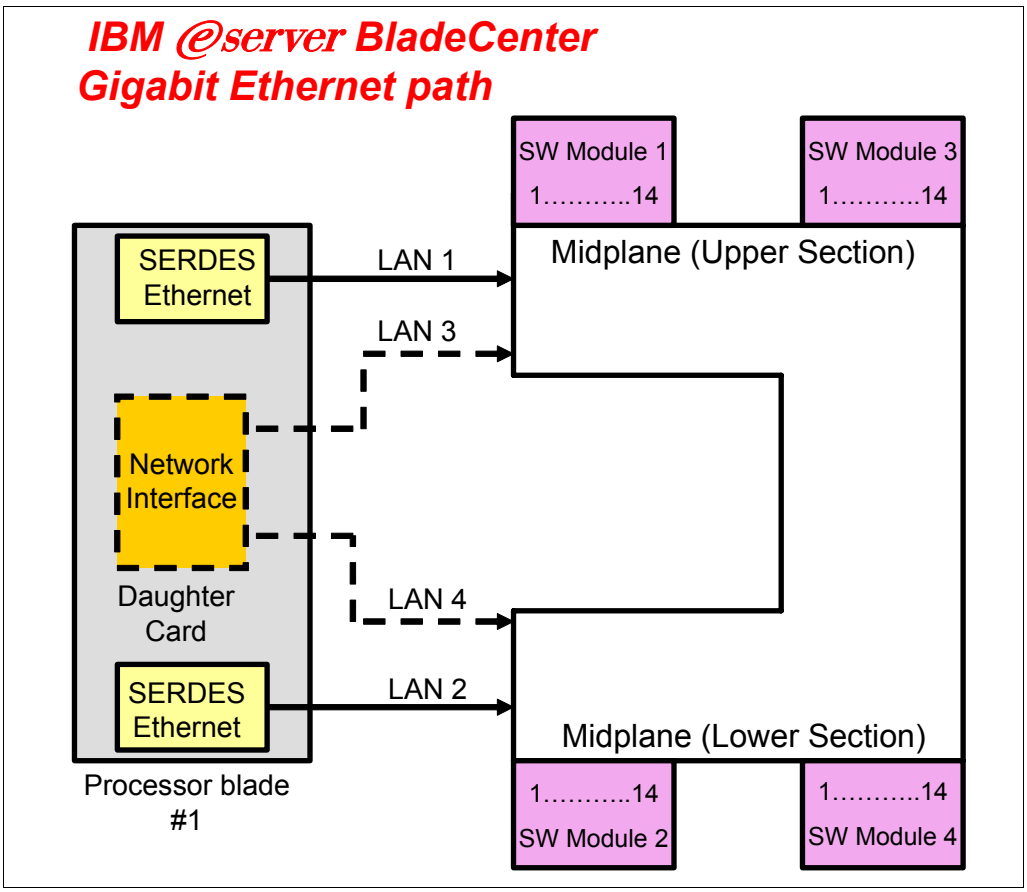

図 2-4　Gigabit Ethernet パス

2.3 IBM eServer HS20 アーキテクチャー

ここでは、BladeCenter HS20 のアーキテクチャー設計について説明します。これは、標準的なデュアル・プロセッサー・サーバー用のブレード設計の 1 例に過ぎません。

BladeCenter HS20 は、ServerWorks Grand Champion LE（または 4.0 Low End）チップ・セットを使用します。（10 ページの図 2-5 の HS20 アーキテクチャーを参照してください。）

Champion Memory and I/O Controller（CMIC）は、次のものをサポートする、I/O バス用のメモリー・コントローラーおよびインターフェースです。

- Xeon プロセッサー 2 個
- DDR-SDRAM メモリー・チャネル 2 個

CMIC には、2 つの Champion I/O Bridge（CIOB-X2）チップとの 2 つのモジュール間バス（IMB2）があります。また CMIC は Thin IMB バスを通じて Champion South Bridge（CSB5）にも接続されます。

CSB5 は、次のものとのインターフェースを提供します。

- 8 MB のメモリーを装備する ATI Rage XL ビデオ・コントローラーとの接続に使用される 1 つの PCI バス
- 4 MB EEPROM（POST/BIOS コードを保持）および SIO（SuperI/O）チップとの接続に使用される 2 つの LPC（ロー・ピン・カウント）バス
- 内部ストレージをサポートする 2 つの IDE チャネル
- FDD/CDROM およびキーボード / マウスとの冗長接続用の 4 つの USB バス

このシステムは、I2C バスに接続されている H8S2148 IBM 内蔵システム管理プロセッサーを使用します。

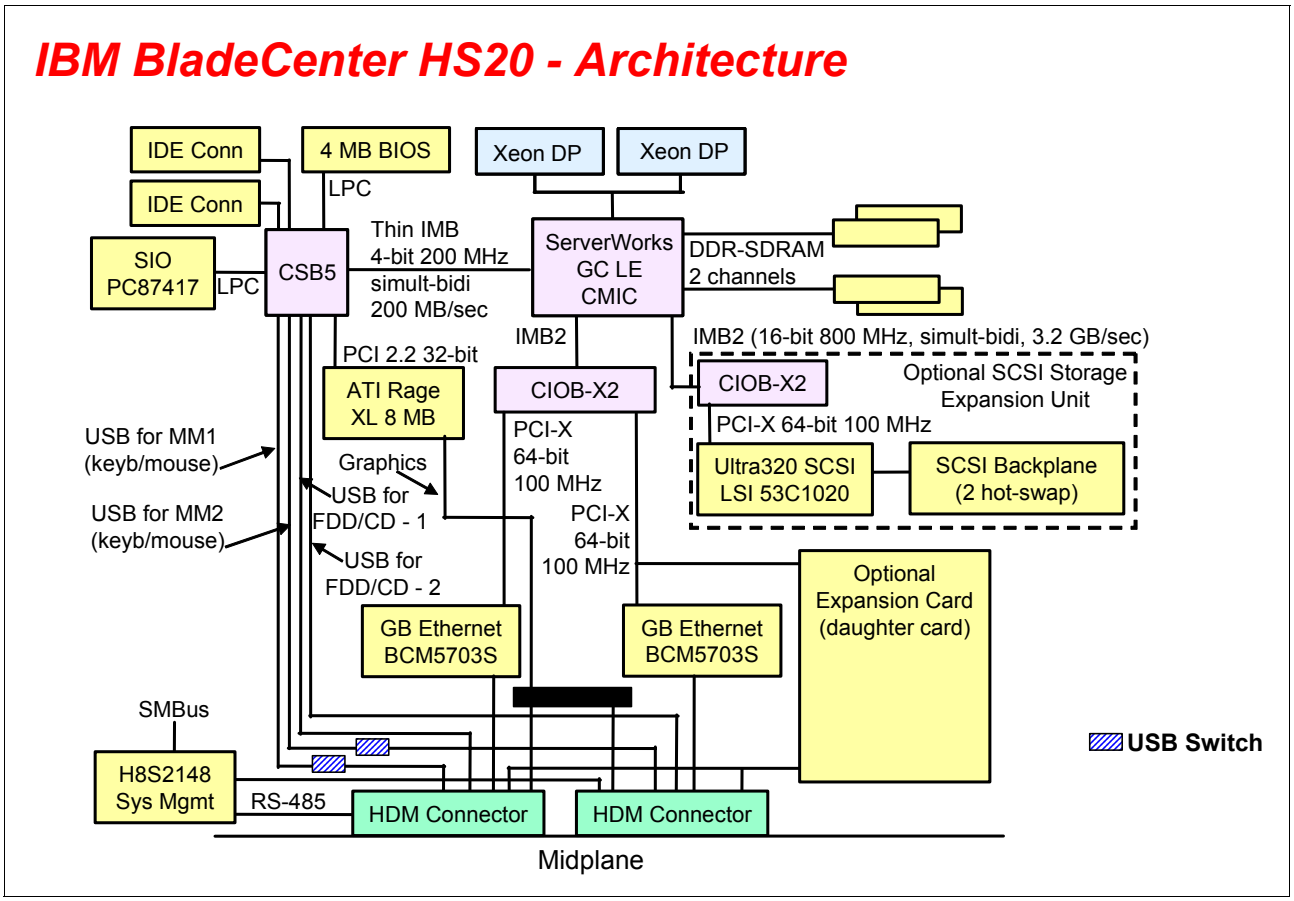

図 2-5　HS20 アーキテクチャー

2.4 スタンドアロン構成ツール

BladeCenter ハードウェアの構成には、すべてのメインストリーム OS プラットフォームで使用可能な、Web ブラウザーや Telnet クライアントなどの標準ソフトウェアを使用します。これには、管理モジュールとイーサネット・スイッチ・モジュールの両方に組み込まれている Web および ANSI インターフェースを活用します。Web インターフェースから総合的なツールにアクセスできます。このツールには各種の構成サブメニューが含まれています。そのサブメニュー（「Switch Tasks」）の 1 つを使用すると、お客様はイーサネット・スイッチ・モジュールをセットアップできます。基本設定（例えば、イーサネット・スイッチ・モジュールの IP アドレスや外部ポートの使用可能化など）は、I2C バスを利用して構成されます。拡張メニューでは、モジュールの微調整が可能です。これには、Web ブラウザーで別のウィン

ドウを開くか、ANSI インターフェースとの接続を可能にする Java™ アプレットを実行します（これには、管理システムに Java 2 V1.4 がインストールされている必要があります）。これを行うには、BladeCenter バックプレーンを通じて MM とイーサネット・スイッチ・モジュールを接続する 10/100 Mb 内部リンクを利用します。（MM の内部ネットワーク・インターフェースのデフォルト静的 IP アドレスは 192.168.70.126 であることに注意してください。）これらのより完全なツールには、Web ブラウザーまたは Telnet クライアントにイーサネット・スイッチ・モジュール自体の IP を指定してアクセスすることもできます。（背面のベイ 1 に接続されるモジュールのデフォルトは 192.168.70.127 ですが、DHCP ベースのアドレッシングを構成可能です。）この後者の機能には、管理システムがイーサネット・スイッチ・モジュールの外部ポート（実動 LAN 上）を通じて接続されている必要があります。したがって、セキュリティーに関する懸念が生じる可能性があります。そのため、お客様には、MM インターフェースの「Switch Tasks」で、外部ポートを通じた構成制御を使用不可にする機能が用意されています。図 2-6 を参照してください。

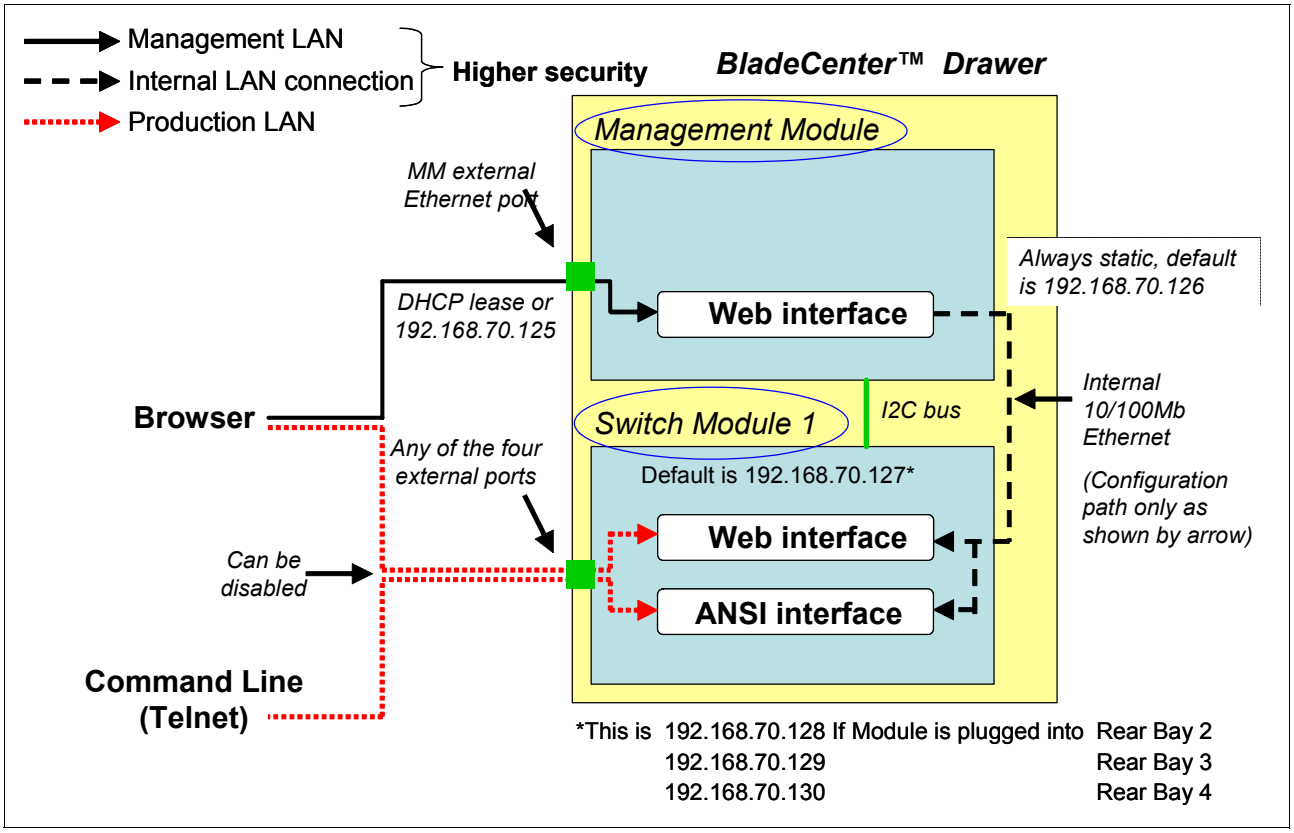

図 2-6　スタンドアロン構成ツール

2.5 IBM TotalStorage

本書では、Topspin InfiniBand Switch Module における IBM TotalStorage Storage Area Network（SAN）テクノロジーの使用について説明します。今日、Storage Area Network は広く受け入れられています。SAN ファブリックによって接続されている、異なるベンダー製のコンポーネント間のインターオペラビリティーの問題が、注目を集め、おおむね解決されていますが、異なるベンダー製の各種装置に保管されているデータを管理する問題は、業界の大きな課題として残っています。

2.5.1 サポートされるホスト接続ストレージ製品

ここでは、Topspin InfiniBand Switch Module および BladeCenter 構成のサポートに使用される Topspin Switch Module 360 または 90 がサポートするホスト接続ストレージ製品をリストしています。

DS4300

IBM TotalStorage DS4300（旧称 FAStT600）は、ミッドレベルのディスク・システムであり、3 台の EXP700 を使用する場合は 8 TB を超えるファイバー・チャネル・ディスクまで拡張し、7 台の EXP700 を使用する場合は Turbo 機能を使用して 16 TB を超えるファイバー・チャネル・ディスクまで拡張できます。また、最新のストレージ・ネットワーキング・テクノロジーを使用して、エンドツーエンド 2 Gbps ファイバー・チャネル・ソリューションを提供します。モデル 4300（Turbo 機能付き）は、IBM DS4000 ミッドレンジ・ディスク・システム・ファミリーの製品であり、同じ共通ストレージ管理ソフトウェアとハイパフォーマンス・ハードウェア設計を使用し、ハイエンド・モデルに見られるエンタープライズ機能に似た機能を、非常に低いコストでお客様に提供します。

新しい DS4000 Storage Manager では、ストレージ・パーティションごとに最大 256 個の論理ボリューム（LUN）、2 TB を超えるアレイ・グループの定義、および SCSI-3 Persistent Reservation が使用可能です。DS4000 内の完全な論理ボリューム・コピー用の新機能である VolumeCopy を備えた FlashCopy® は、DS4300 Turbo で利用できます。

DS4300 および DS4300（Turbo 機能付き）を EXP100 と結合すると、最大 28 TB の RAID 保護ストレージ・ソリューションを構成して、アクセス制限、データ参照、ニアライン・ストレージに対して急速に高まるアプリケーションのニーズに応じて経済的かつ拡張が容易なストレージを提供することができます。

IBM TotalStorage DS4300 の詳細については、次の Web サイトをご覧ください。

http://www-1.ibm.com/servers/storage/disk/ds4000/ds4300/index.html

DS4400

IBM TotalStorage DS4400（旧称 FAStT700）は、2 Gbps ファイバー・チャネル・テクノロジーを使用して優れたパフォーマンスを提供します。DS4400 は、拡張機能および柔軟な機能を使用して投資を保護するように設計されています。e-business アプリケーションによるストレージ要件の増大をサポートするために、36 GB から 32 TB 以上まで拡張します。DS4400 は、ビジネスの継続性をサポートするための拡張複製サービスを提供します。DS4400 は、無制限のパフォーマンスを必要とする企業向けに有効なディスク・システムです。

IBM TotalStorage DS4400 の詳細については、次の Web サイトをご覧ください。

http://www-1.ibm.com/servers/storage/disk/ds4000/ds4400/index.html

DS4500

IBM TotalStorage DS4500（旧称 FAStT900）は、データ集中のコンピューティング環境における要求の厳しいアプリケーション用に画期的なディスク・パフォーマンスと卓越した信頼性を提供します。DS4500 は、拡張機能および柔軟な機能を使用して投資を保護するように設計されています。DS4500 は、今日のオンデマンド・ビジネス・ニーズ用に設計され、EXP700 を使用して最大 32 TB のファイバー・チャネル・ディスク・ストレージ容量を提供します。DS4500 は、ビジネスの継続性と災害時回復をサポートするための拡張複製サービスを提供します。DS4500 は、無制限のパフォーマンスを必要とする企業向けに有効なディスク・システムです。

DS4500 は EXP100 と結合すると、最大 56 TB の RAID 保護ストレージ・ソリューションを構成して、アクセス制限、データ参照、ニアライン・ストレージに対して急速に高まるアプ

リケーションのニーズに応じて経済的かつ拡張が容易なストレージを提供することができます。

IBM TotalStorage DS4500 の詳細については、次の Web サイトをご覧ください。

http://www-1.ibm.com/servers/storage/disk/ds4000/ds4500/index.html

IBM TotalStorage Product Guide は次の Web サイトでご覧いただけます。

ftp://ftp.software.ibm.com/common/ssi/rep_sp/n/TSB00364USEN/TSB00364USEN.PDF

3

InfiniBand テクノロジー

InfiniBand は、インターネット・インフラストラクチャーの I/O 接続をサポートするために設計された、新しい強力なアーキテクチャーです。InfiniBand は、I/O インターコネクト標準を拡張して、サーバーにおける次世代 I/O インターコネクト標準を作成する手段として、あらゆる主要 OEM サーバー・ベンダーによってサポートされます。初めて、大容量の業界標準 I/O インターコネクトが、従来の筐体内（in-the-box）バスの役割を拡張します。InfiniBand は、筐体内（in-the-box）バックプレーン・ソリューション、外部インターコネクト、および「筐体外帯域幅（Bandwidth Out of the Box）」のいずれも提供するという点で固有であるので、以前は従来のネットワーク・インターコネクトのみに確保されていた方法で接続を提供します。こうした I/O とシステム・エリア・ネットワーキングの統合には、以前は別々であったこれらの 2 つのドメインのニーズをサポートする新しいアーキテクチャーが必要です。この大幅な I/O 変換の基礎になるのは、インターネットにおける RAS（信頼性 • 可用性 • 保守性）要件をサポートする InfiniBand の機能です。この章は、Mellanox Technologies 社による寄稿です。

この章では、既存の PCI バスおよびその他の専有スイッチ・ファブリックや I/O ソリューションと比較して、RAS をサポートする InfiniBand の優れた能力を実証する機能について説明します。さらに、InfiniBand アーキテクチャーが包括的なシリコン、ソフトウェア、およびシステム・ソリューションをどのようにサポートするかについて概要を記載します。InfiniBand 1.1 仕様の主な項目の概要を述べ、このアーキテクチャーの包括的な性質を明らかにします。1.1 仕様の範囲は、業界標準の電気インターフェースや機械的なコネクターから、適切に定義されたソフトウェアと管理インターフェースまでにわたります。

この章は、4 つのセクションで構成されています。概要のセクションでは、InfiniBand の概要を記述し、主要なサーバー・ベンダーがすべて、この新しい標準の採用を決定した理由を示します。次のセクションでは、既存のテクノロジーによって現在対処されている各種市場に InfiniBand が与える影響を説明します。3 番目のセクションでは、スイッチ・ファブリックとバス・アーキテクチャーとを全体的に比較してから、InfiniBand と PCI やその他の専有ソリューションとを比較して詳しく説明します。最後のセクションでは、アーキテクチャーについて詳しく説明して、InfiniBand の最も重要な機能について高水準の検討を行います。

3.1 概要

「アムダールの法則」は、コンピューターサイエンスの基本原理の 1 つです。この法則は、効率的なシステムが CPU パフォーマンス、メモリー帯域幅、および I/O パフォーマンスとの間のバランスを取る必要があることを示しています。これに対立するのが「ムーアの法則」です。これは、半導体はほぼ 18 カ月ごとにパフォーマンスを倍増すると正確に予測しています。I/O インターコネクトは、半導体のスケーリング機能よりも厳しい機械的および電気的な制限が適用されるので、これらの 2 つの法則により、最終的に不均衡が生じ、システムのパフォーマンスが制限されます。これは、システムのパフォーマンスを維持するために I/O インターコネクトが数年ごとに根本的に変化する必要があることを示唆しています。実際に、I/O インターコネクトの頻繁な変更を妨げる、もう 1 つの実際的な法則があります。つまり、「壊れていないものを修理するな」です。

バス・アーキテクチャーには大量の慣性があります。これは、バス・アーキテクチャーが、半導体装置のバス・インターフェース・アーキテクチャーだけでなく、コンピューター・システムとネットワーク・インターフェース・カードの機械的接続を指示するからです。このため、バス・アーキテクチャーが成功すると、通常、10 年以上も支配的な地位を占めています。PCI バスは、1990 年代始めに標準 PC アーキテクチャーに導入され、1 回大幅なアップグレード（32 ビット /33 MHz から 64 ビット /66 Mhz へ）をしただけで、優位を維持しています。PCI-X イニシアチブはこれをもう一歩進めて 133 MHz にアップグレードしたので、PCI アーキテクチャーはもう数年寿命が延びたように思われます。しかし、パーソナル・コンピューター（PC）とサーバーに必要なものには相違があります。

PC は、PCI 64/66 の帯域幅能力を圧迫することはありません。PCI スロットは、ホーム・ユーザーやビジネス・ユーザーが PC の能力をアップグレードするために、ネットワーキング・カード、ビデオ・デコード・カード、拡張サウンド・カード、またはその他のカードを購入する手段を提示します。他方、今日のサーバーには、多くの場合、単一のシステム内にクラスタリング、ネットワーキング（Gigabit Ethernet）、およびストレージ（ファイバー・チャネル）カードが組み込まれています。これは、PCI-X の 1-GB 帯域幅の限界を圧迫しています。InfiniBand アーキテクチャーを配置すると、PCI-X の帯域幅制限がさらに厳しくなります。InfiniBand アーキテクチャーは、今日の市場で PCI HCA（ホスト・チャネル・アダプター）として配置される 4X リンクを定義しました。これらの HCA が、従来達成されたものより多くの帯域幅を提供する場合であっても、PCI-X はボトルネックになります。これは、単一の InfiniBand 4X リンクの総計帯域幅が 20 Gbps または 2.5 GBps であるからです。ここで、PCI Express などの新しい *ローカル* I/O テクノロジーが、InfiniBand に対して重要な補完的役割を果たします。

インターネットの広がりと 24x7（1 日 24 時間 週 7 日）アップタイムの需要により、システムのパフォーマンスと信頼性の要件が、今日の PCI インターコネクト・アーキテクチャーではサポートできなくなるレベルまで進んでいます。データ・ストレージ、Web サーバー、アプリケーション・サーバー、およびデータベース・サーバー、ならびにエンタープライズ・コンピューティングにより、さらに高いパフォーマンスを提供する、常時使用可能なフェイルセーフ・システムの需要が高まっています。業界の傾向としては、ストレージがサーバーから、分離されたストレージ・ネットワークに移され、フォールト・トレラント・ストレージ・システム全体にデータが分散されています。こうした需要が、単に帯域幅を増やすという要件を超え、PCI ベースのシステムは、共用バス・アーキテクチャーの限界に達しています。CPU 周波数がギガヘルツ（GHz）の限界を超え、ネットワーク帯域幅が 1 Gbps を超えると、現在の装置でサポートし、拡張するためにより高い帯域幅を提供する、新しい I/O インターコネクトが必要になります。

InfiniBand は、ポートあたり双方向 2.5 Gbps または 10 Gbps の基本速度で動作する、スイッチ・ベースのシリアル I/O インターコネクト・アーキテクチャーです。共用バス・アーキテクチャーとは異なり、InfiniBand は、プリント基板（PCB）上の装置を接続する ロー・ピン・カウント・シリアル・アーキテクチャーであり、「筐体外帯域幅（Bandwidth Out of the Box）」を使用可能にして、通常の対より線カッパー・ワイヤー上で最大 17 m の距離に及び

ます。一般的なファイバー・ケーブルでは、数 km あるいはそれ以上の距離に及ぶことができます。さらに、InfiniBand は QoS（Quality of Service）と RAS の両方を提供します。これらの RAS 機能は、当初から InfiniBand アーキテクチャーの設計に取り入れられ、インターネットの心臓部にある次世代の計算サーバーとストレージ・システム用の共通 I/O インフラストラクチャーとしての機能を果たすのに不可欠です。その結果、InfiniBand は、インターネット・インフラストラクチャーのシステムとインターコネクトを根本的に変更します。本書では、この変革を可能にする、InfiniBand 特有の機能について説明します。

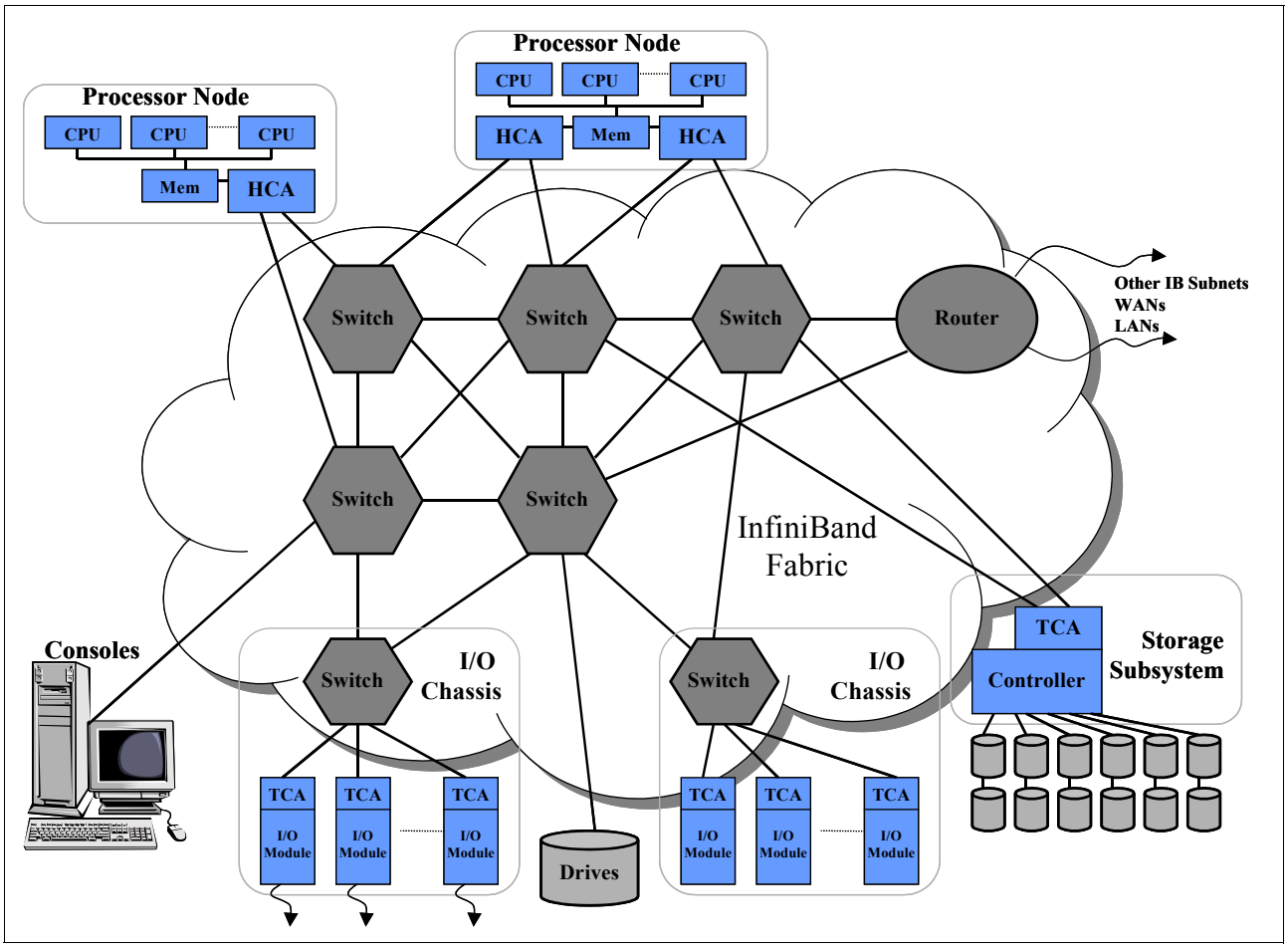

図 3-1　InfiniBand システム・ファブリック

InfiniBand システム・ファブリックは、すべての主要サーバー・ベンダーを始めとして、業界のトップ企業によってサポートされています。InfiniBand アーキテクチャーは、本書に記述しているすべての利点を提供しますが、現行の 10 Gbps リンクのフルパフォーマンス帯域幅を実現するには、PCI の制限を解決する必要があります。この点で、現在開発中のインターコネクト・テクノロジーが InfiniBand を支援します。

3.2 市場

アプリケーション・クラスタリング、Storage Area Network、層間通信、およびプロセッサー間通信などの重要な市場では、高帯域幅、QoS、および RAS 機能が必要です。また、多くの組み込みシステム（ルーター、ストレージ・システム、およびインテリジェント・スイッチを含む）は、内部 I/O アーキテクチャー用に、多くの場合 Compact PCI 形式で PCI バスを利用しています。このようなシステムは、Gigabit Ethernet や ATM などの高速ネットワーキング・インターコネクトに対応できません。したがって、多くの企業は独自仕様の I/O イン

ターコネクト・アーキテクチャーを開発しています。イーサネット・ローカル・エリア・ネットワーク（LAN）、ファイバー・チャネル Storage Area Network、および多数の広域ネットワーク（WAN）インターコネクトの開発経験に基づいて、今日の市場のニーズを超え、幅広いシステム用の密接なインターコネクトを提供するように InfiniBand はネットワーク化されています。これは、RAS、QoS、および拡張容易性などの非常に重要な項目を直接サポートすることによって実現されます。

3.2.1 アプリケーション・クラスタリング

今日のインターネットは、ストリーミング・メディア、企業間ソリューション、e-commerce、および対話式ポータル・サイトなどのアプリケーションをサポートする、グローバル・インフラストラクチャーに進化しています。これらのアプリケーションはそれぞれ、増え続けるデータ量や信頼性の要求をサポートする必要があります。一方、サービス・プロバイダーは、これらのアプリケーションのサポートを非常に負担に感じるようになってきています。さまざまな QoS やセキュリティーのレベルで課金を提供すると同時に、ますます輻輳している通信回線を通じて効率よくトラフィックを処理しなければなりません。アプリケーション・サービス・プロバイダー（ASP）は、e-commerce、e-marketing などの e-business アクティビティーを Web ベース・アプリケーションを専門とする企業に外注することをサポートするために生まれました。これらの ASP は、爆発的なインターネットの成長に対応するために短期間で大幅に拡張する機能を提供する、信頼性の高いサービスを提供できなければなりません。こうした要件をサポートするための適切なメカニズムとして、クラスターが発展してきました。クラスターとは、ロード・バランス・スイッチに接続し、並列稼動して特定のアプリケーションにサービスを提供するサーバーのグループです。

InfiniBand は、機能が豊富な管理されたアーキテクチャーとネットワーク・インターコネクトを統合することによって、アプリケーション・クラスター接続を単純化します。InfiniBand のスイッチ・アーキテクチャーは、ネイティブ・クラスター接続を提供することにより、筐体内部と外部で拡張容易性と信頼性をサポートします。装置を追加可能であり、ファブリックにスイッチを追加して複数のパスを利用できます。InfiniBand に組み込まれている QoS メカニズムを使用して、装置間の高優先順位トランザクションを低優先順位項目より前に処理できます。

3.2.2 プロセッサー間通信

プロセッサー間通信により、単一のアプリケーションで複数のサーバーが連携して動作することができます。信頼性の高い処理を確保するには、高帯域幅、低遅延の信頼できる接続がサーバー間に必要です。アプリケーションが必要とするプロセッサー帯域幅が増えるにつれて、拡張容易性が非常に重要になります。InfiniBand の交換（スイッチ）の性質は、システム間に複数のパスを可能にすることによって、プロセッサー間通信システムに信頼性の高い接続を提供します。拡張容易性は、単一のユニット（サブネット・マネージャー）によって管理される完全にホット・スワップ可能な接続を使用してサポートされます。マルチキャスト・サポートを使用すると、複数の宛先に対して単一のトランザクションを行うことができます。これには、サブネット上のすべてのシステムへの送信、またはこれらのシステムのサブセットのみへの送信が含まれます。InfiniBand によって定義されるより高い帯域幅接続（4X、12X）は、2 次 I/O インターコネクトを必要とすることなく、プロセッサー間通信クラスター用のバックボーン機能を提供します。

3.2.3 Storage Area Network

Storage Area Network は、複数のサーバーから非常に大量のデータにアクセスできるようにするために、管理対象スイッチを通じて接続されている複合ストレージ・システムのグループです。今日、Storage Area Network は、ファイバー・チャネル・ホスト・バス・アダプター（HBA）を通じて接続されているファイバー・チャネル・スイッチ、ハブ、およびサーバーを使用して構築されます。Storage Area Network は、インターネット・データ・センターが必要とする大規模な情報データベースとの信頼性の高い接続を提供するために使用されます。

Storage Area Network は、個々のサーバーがアクセスできるデータを制限することができ、それによって重要な区分化メカニズム（場合によっては、ゾーニングまたは隔離と呼ばれます）を提供します。

InfiniBand のファブリック・トポロジーにより、ストレージとサーバー間の通信を単純化することができます。ファイバー・チャネル・ネットワークを除去すると、コストのかかる HBA なしにサーバーを Storage Area Network に直接接続できます。リモート DMA（RDMA）サポート、同時対等通信、エンドツーエンド・フロー制御などの機能を使用することにより、InfiniBand は、コストの高い複雑な HBA を必要とすることなく、ファイバー・チャネルの欠点を克服します。帯域幅の比較については、後述します。

3.3 I/O アーキテクチャー：ファブリックとバスの比較

共用バス・アーキテクチャーは、多数の欠点がありますが、現在最も一般的な I/O インターコネクトです。クラスターとネットワークには、高速のフォールト・トレラント・インターコネクトを備えたシステムが必要です。これは、バス・アーキテクチャーでは適切にサポートできません。したがって、すべてのバス・アーキテクチャーには、拡張が容易なネットワーク・トポロジーを使用可能にするネットワーク・インターフェース・モジュールが必要です。システムに対応するために、I/O アーキテクチャーは、拡張機能を備えた高速接続を提供する必要があります。表 3-1 は、スイッチ・ファブリック・アーキテクチャーと共用バス・アーキテクチャーとの単純な機能の比較を示しています。

表 3-1　ファブリック・アーキテクチャーと共用バス・アーキテクチャーの比較

機能	ファブリック	バス
トポロジー	スイッチ	共用バス
ピン数	少ない	多い
エンドポイント数	多い	少ない
最大信号長	KM	インチ
信頼性	あり	なし
拡張容易性	あり	なし
フォールト・トレラント	あり	なし

3.3.1 共用バス・アーキテクチャー

バス・アーキテクチャーでは、すべての通信が同一帯域幅を共用します。バスに追加されるポート数が多いほど、各周辺装置が使用可能な帯域幅が少なくなります。また、電気、機械、および電源に関する厳しい項目もあります。パラレル・バスでは、接続ごとに多くのピンが必要であり（64 ビット PCI には 90 個のピンが必要）、ボードのレイアウトが非常に難しくなり、貴重なプリント基板（PCB）のスペースを消費します。高いバス周波数では、各信号の距離が、PCB ボード上の短い線に制限されます。複数のカード・スロットを備えたスロット・ベースのシステムでは、終端が制御されず、正しく設計されていないと問題が発生する可能性があります。

3.3.2 従来型の共用バス・アーキテクチャー

バス設計にはロードの制限があるので、バスごとに数個の装置しか使用できません。この制限を克服するには、ブリッジ装置を追加して、新しいロード制限のある別のバスをブリッジの背後に提供します。この方法ではシステムに接続できる装置数を増やすことができますが、システムの他の部分の装置にアクセスするときは、データは引き続き中央バスを流れま

す。システムにブリッジを追加するごとに、待ち時間と輻輳が増えます。仕様で許可される最悪のケースの装置数を想定して、フルロードされた状態で動作するようにバスを設計する必要があります。これにより、バスの周波数は根本的に制限されます。バスの主な問題の 1 つは、「筐体外（out of the box）」システム・インターコネクトをサポートできないことです。システムが対話するには、イーサネット（サーバー間通信）やファイバー・チャネル（ストレージ・ネットワーキング）などの別個のインターコネクトが必要です。

3.3.3 スイッチ・ファブリック・アーキテクチャー

スイッチ・ファブリックは、フォールト・トレランスと拡張容易性を確保するために設計された、Point-to-Point スイッチ・ベース・インターコネクトです。Point-to-Point スイッチ・ファブリックとは、すべてのリンクでリンクの各端に正確に 1 つの装置が接続されていることを意味します。したがって、ロードと終端の特性が適切に制御され、（バス・アーキテクチャーとは異なり）1 つの装置が許可されるだけで、最悪のケースが通常のケースと同じであるので、ファブリックでは I/O パフォーマンスが大幅に上昇します。

スイッチ・ファブリック・アーキテクチャーは、拡張容易性を提供します。これを実現するには、スイッチをファブリックに追加し、スイッチを通じてより多くのエンド・ノードを接続します。共用バス・アーキテクチャーとは異なり、ネットワークにスイッチが追加されると、システムの総計帯域幅が増えます。装置間の複数のパスが、総計帯域幅を高く保ち、フェイルセーフの冗長接続を提供します。

3.3.4 InfiniBand を補完する新規インターコネクト

PCI Express などの新しいインターコネクトは、新しいレベルのプロセッサー帯域幅へのアクセスを提供し、InfiniBand が　この帯域幅を外部に拡張できるようにするので、実際に InfiniBand の重要なイネーブラーです。PCI Express などのテクノロジーが開発されています。これは、4X InfiniBand リンクに必要な 20 Gbps、さらに 12X InfiniBand リンクに必要な 60 Gbps までもサポートできるシステム・ロジックとの接続点を InfiniBand に提供します。これらのテクノロジーにより InfiniBand は適切に補完されます。

3.3.5 筐体外帯域幅（Bandwidth Out of the Box）

InfiniBand アーキテクチャーの基本的な特徴は、「筐体外帯域幅（Bandwidth Out of the Box）」という概念です。InfiniBand は帯域幅を占有する機能があります。これは、従来、サーバー内部に限定されていましたが、これをファブリック全体に拡張します。InfiniBand は、ファブリック内のどこでも必要とされる場所に正確にデータを配信することによって、10 Gbps のパフォーマンスを有効に利用できるようにします。従来、CPU からデータが遠ざかると、帯域幅は減少します。「*筐体外部（Outside the box）*」とは、プロセッサーから I/O 、クラスタリングまたはプロセッサー間通信のサーバー間、ストレージ、さらにデータ・センターの端までの帯域幅を意味します。最新のプロセッサーには、25 Gbps で他のプロセッサーやメモリーと通信できるフロント・サイド・バスがありますが、現在使用可能な PCI-X システムは、筐体外部で使用可能な帯域幅を 8 Gbps のみに制限します。データ・センター内の実際の帯域幅はさらに制限され、プロセッサー間通信帯域幅は 1 または 2 Gbps に、ファイバー・チャネルまたはストレージ通信は最高で 2 Gbps に、また通常のイーサネットを介したシステム間の通信は 1 Gbps に制限されます。これは、プロセッサーからデータ・センターの端までの間に 1 桁の帯域幅が失われることを示しています。

前述のように、新しいインターコネクト（すなわち、PCI Express）は、30 Gbps を優に超えて、60 Gbps までにもローカル I/O 帯域幅を増やすことができます。新しいプロセッサーとシステム・チップ・セットがこれらのインターコネクトを組み込むと、現在の PCI の制限が克服されます。HCA がこれらのインターコネクトに接続すると、InfiniBand の帯域幅は制限から解放され、これにより、クラスタリング、通信、およびストレージがすべて、ネイティブ InfiniBand 速度で接続できるようになります。1X（2.5 Gbps）および 4X（10 Gbps）リンク

は 2001 年から配置され、2003 年には、12X すなわち 30 Gbps リンクの配置が記録されています。

21 ページの図 3-2 は、データ・センター内の帯域幅を年次順に表示して、InfiniBand がどのように「筐体外帯域幅（Bandwidth Out of the Box）」を解放するかを示しています。1998 年頃、Intel® Penium® II は、ワールド・クラスのパフォーマンスを提供しましたが、計算サーバー・アーキテクチャーの全体的な設計により、プロセッサーの帯域幅は「筐体内部（inside the box）」に制限されました。プロセッサーからデータが遠ざかるほど、帯域幅は低くなり、1 桁を超える帯域幅が失われ、端では 100 Mbps になります。1999 年頃、Pentium III がプロセッサーのパフォーマンスを改善しましたが、この相関関係は同じままです。帯域幅は距離と共に失われ、データ・センターは端ではなお 100 Mbps でしか通信できません。2000 年と 2001 年には、Pentium 4 とその他のすべてのデータ・センター・サブシステムが帯域幅を改善しましたが、相関関係は同じままです。プロセッサーからデータ・センターの端までの間に、1 桁を超える帯域幅が失われます。InfiniBand アーキテクチャーはこの相関関係を変えます。InfiniBand アーキテクチャーは、LAN/WAN やストレージ接続を含めて、プロセッサーからデータ・センターの端まで 20 Gbps の帯域幅（総計ボー・レート）を提供します。InfiniBand は「筐体外帯域幅（Bandwidth Out of the Box）」を可能にし、プロセッサー・レベルの帯域幅がデータ・センターの端までずっと続きます。また 2003 年には、InfiniBand アーキテクチャーが、60 Gbps まで拡張する余裕の 12X を提供しました。

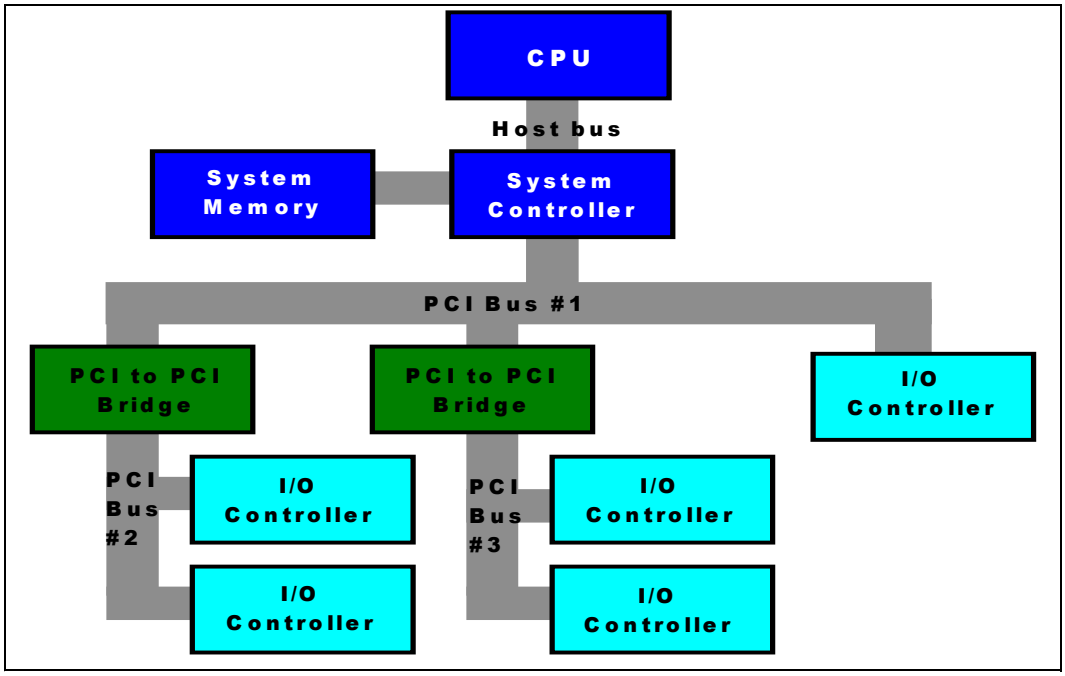

図 3-2　従来型の共用バス・アーキテクチャー

InfiniBand は帯域幅を提供しますが、システム・メモリー間の RDMA 転送を使用して、必要な場所にデータを提供することにも注目してください。InfiniBand は、ハードウェア内に信頼性の高い正常なトランスポート接続を実装します。したがって、ホスト CPU の支援なしで、低遅延で非常に効率よくデータが配信されます。これは、イーサネットと比べると非常に大きな利点です。イーサネットでは、待ち時間がはるかに長く、TCP スタックの実行に相当な CPU サイクルを消費します。

3.4 InfiniBand の技術的な概要

InfiniBand は、次世代のシステム要件に合わせて拡張する機能を備えた今日のシステム用に開発された、スイッチ・ベースの Point-to-Point インターコネクト・アーキテクチャーです。

コンポーネント間インターコネクトとして、および「筐体外（out of the box）」のシャーシ間インターコネクトとして、PCB 上で動作します。個々のリンクはそれぞれ、4 線式 2.5 Gbps 双方向接続に基づいています。このアーキテクチャーは、初期化と装置間通信を管理するために、ソフトウェア層と共に、階層化ハードウェア・プロトコル（物理層、リンク層、ネットワーク層、トランスポート層）を定義します。各リンクは、信頼性を確保するために複数のトランスポート・サービスをサポートし、優先順位が付けられた複数の仮想通信チャネルをサポートすることができます。

サブネット内の通信を管理するために、このアーキテクチャーは、各 InfiniBand エレメントの構成と管理を行う通信管理スキームを定義します。堅固な接続ファブリックを確保するために、エラー報告、リンク・フェイルオーバー、シャーシ管理、およびその他のサービスに対して管理スキームが定義されます。

InfiniBand 機能セットには、次のものがあります。

- 階層化プロトコル : 物理層、リンク層、ネットワーク層、トランスポート層、上位層
- パケット・ベース通信
- Quality of Service
- 3 つのリンク速度
- 1X - 2.5 Gbps、4 線式
- 4X - 10 Gbps、16 線式
- 12X -30 Gbps、48 線式
- PCB、カッパー、およびファイバー・ケーブル・インターコネクト
- サブネット管理プロトコル
- リモート DMA サポート
- マルチキャストおよびユニキャスト・サポート
- 信頼性の高いトランスポート方式 : メッセージ・キューイング
- 通信フロー制御 : リンク・レベルおよびエンドツーエンド

3.5 InfiniBand の層

InfiniBand アーキテクチャーは、複数の層に分割され、各層は互いに独立して作動します。図 3-3 に示されているように、InfiniBand には物理層、リンク層、ネットワーク層、トランスポート層、上位層があります。

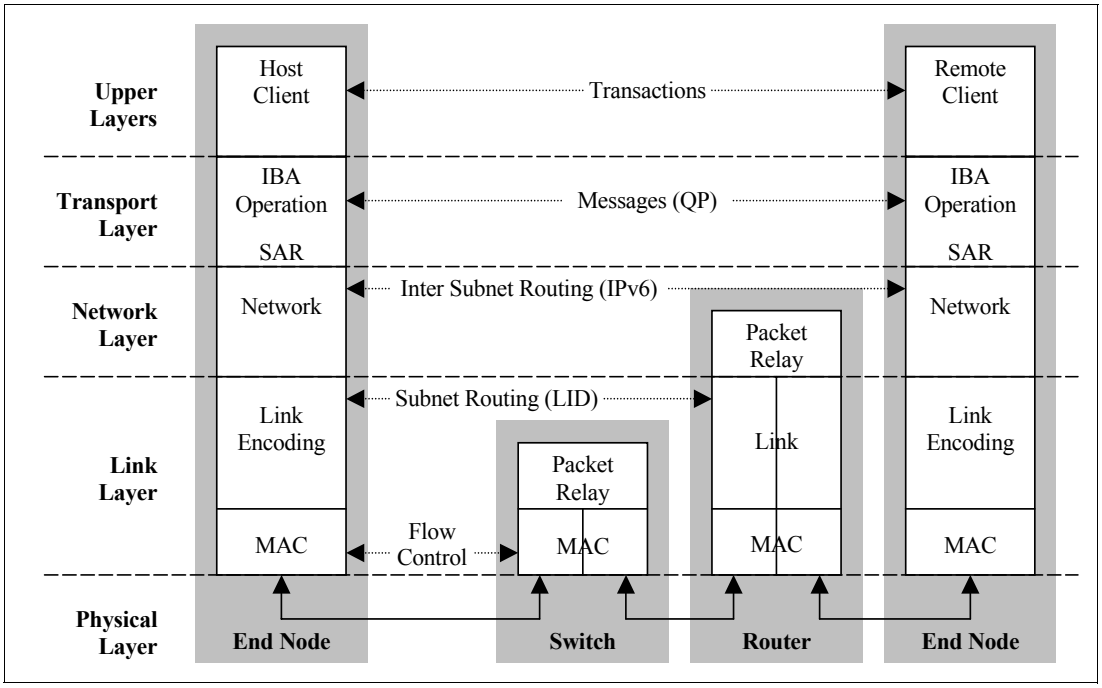

図 3-3 InfiniBand の層

3.5.1 物理層

InfiniBand は、システムの電気的特性および機械的特性を定義する包括的なアーキテクチャーです。これには、ファイバーおよびカッパー・メディア用のケーブルとコンセント、バックプレーン・コネクター、およびホット・スワップ特性が含まれます。

InfiniBand アーキテクチャー仕様 v1.0、サンプル・コネクター : 機械的特性

InfiniBand は、物理層で 1X、4X、および 12X の 3 つのリンク速度を定義します。個々のリンクはそれぞれ、4 線式シリアル差分接続（各方向に 2 線）であり、2.5 Gbps で全二重接続を提供します。これらのリンクは、24 ページの図 3-4 に示されています。

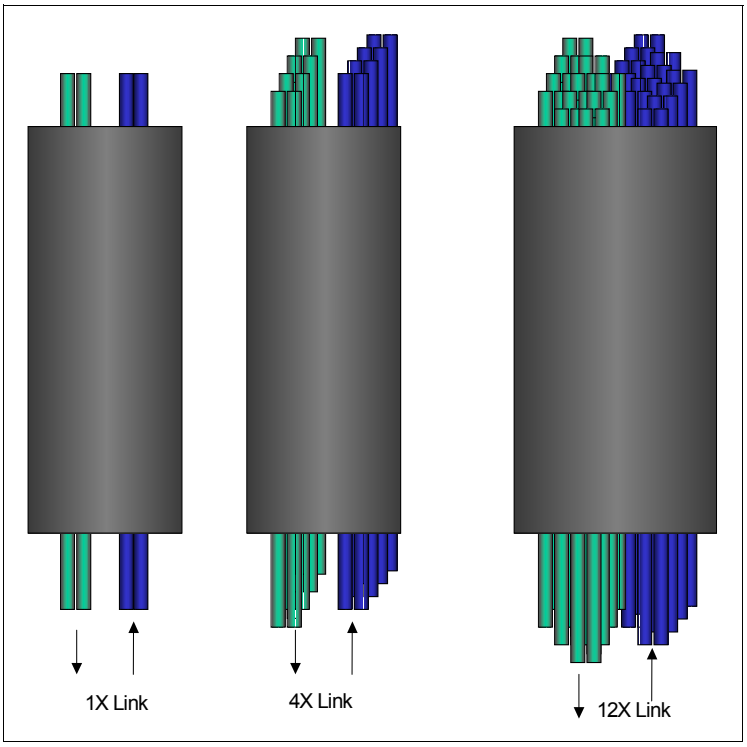

図 3-4 InfiniBand の物理リンク

InfiniBand の物理リンク

これらのリンクのデータ転送速度とピン・カウントが 表 3-2 に表示されています。

表 3-2 InfiniBand リンク速度

InfiniBand リンク	シグナル・カウント	データ信号速度	データ転送速度	全二重データ転送速度
1X	4	2.5 Gbps	2.0 Gbps	4.0 Gbps
4X	16	10 Gbps	8 Gbps	16.0 Gbps
12X	48	30 Gbps	24 Gbps	48.0 Gbps

注 : InfiniBand 1X リンクの帯域幅は 2.5 Gbps です。実際のロー・データ帯域幅は 2.0 Gbps です（データは 8b/10b エンコードされます）。リンクは双方向であるので、バスに対する総計帯域幅は 4 Gbps です。大部分の製品は、総計システム I/O 帯域幅が加算されるマルチポート設計です。

InfiniBand は、「筐体外（out of the box）」通信用の複数のコネクターを定義します。ラック・マウント・システム用のバックプレーン・コネクターだけでなく、ファイバー・ケーブル・コネクターとカッパー・ケーブル・コネクターの両方が定義されます。

3.5.2 リンク層

リンク層は、トランスポート層と共に InfiniBand アーキテクチャーの中心です。リンク層は、ローカル・サブネット内のパケット・レイアウト、Point-to-Point リンク操作、およびスイッチングを含みます。

パケット

リンク層には、管理パケットとデータ・パケットの 2 つのタイプのパケットがあります。管理パケットは、リンクの構成と保守に使用されます。仮想レーン・サポートなどの装置情報は、管理パケットで決定されます。データ・パケットは最大 4 KB のトランザクション・ペイロードを搬送します。

スイッチング

サブネット内では、パケット・フォワーディングとパケット・スイッチングがリンク層で処理されます。サブネット内のすべての装置には、サブネット・マネージャーによって割り当てられる 16 ビット・ローカル ID（LID）があります。サブネット内で送信されるすべてのパケットは、アドレッシングに LID を使用します。リンク・レベルのスイッチングは、パケットのローカル経路ヘッダー（LRH）内の宛先 LID によって指定される装置にパケットを転送します。LRH はすべてのパケットに存在します。

QoS

QoS は、仮想レーン（VL）を通じて InfiniBand によってサポートされます。これらの VL は、単一の物理リンクを共用する別々の論理通信リンクです。各リンクは、最大 15 個の標準 VL と 1 つの管理レーン（VL 15）をサポートできます。VL15 の優先順位が最高であり、VL0 の優先順位が最低です。管理パケットは排他的に VL15 を使用します。各装置は、少なくとも VL0 と VL15 をサポートする必要があり、その他の VL はオプションです。

パケットがサブネットを通過するときに、その QoS レベルを保証するために、サービス・レベル（SL）が定義されます。パスに沿った各リンクは異なる VL を持つことができ、SL は、各リンクに適切な通信優先順位を提供します。各スイッチ / ルーターには、SL-to-VL マッピング・テーブルがあります。このテーブルは、各リンクでサポートされる VL 数に合わせて適切な優先順位を保つためにサブネット・マネージャーによって設定されます。したがって、InfiniBand アーキテクチャーは、スイッチ、ルーターを通じて長期間、エンドツーエンド QoS を確保することができます。

クレジット・ベースのフロー制御

リンク・レベルのフロー制御は、2 つの Point-to-Point リンク間のデータ・フローの管理に使用されます。フロー制御は VL 単位で処理されるので、別々の仮想ファブリックが同じ物理メディアを利用する通信を保持することができます。リンクの各受信側は、リンク上の送信側装置にクレジットを提供して、データの損失なく受信可能なデータ量を指定します。各装置間で渡されるクレジットは、受信側が受け入れることができるデータ・パケット数を更新するために、専用リンク・パケットによって管理されます。受信バッファー・スペースが使用可能であることを示すクレジットを受信側が公示しない限り、データは送信されません。

データ保全性

リンク・レベルでは、データ保全性を確保するパケット当たり 2 つの CRC、すなわち可変 CRC（VCRC）と不変 CRC（ICRC）があります。16 ビット VCRC には、パケット内のすべてのフィールドが含まれ、各ホップで再計算されます。32 ビット ICRC は、ホップ間で変更されないフィールドのみを含みます。VCRC は、2 つのホップ間でリンク・レベルのデータ保全性を提供し、ICRC は、エンドツーエンドのデータ保全性を提供します。単一の CRC のみを定義する、イーサネットなどのプロトコルでは、エラーはデバイス内で生成され、その後、そのデバイスが CRC を再計算します。データが破壊された場合であっても、ネクスト・ホップの検査により、有効な CRC が明らかになります。ビット・エラーが生成されたときに、そのエラーが必ず検出されるように、InfiniBand には ICRC が組み込まれています。

3.5.3 ネットワーク層

ネットワーク層は、サブネット間のパケットのルーティングを処理します。（サブネット内では、ネットワーク層は必要ありません。）サブネット間で送信されるパケットには、グ

ローバル経路ヘッダー（GRH）が含まれています。GRH には、パケットの送信元と宛先の 128 ビット IPv6 アドレスが入っています。パケットは、各装置の 64 ビット・グローバル固有 ID（GUID）に基づいて、ルーターを通じてサブネット間で転送されます。ルーターは、各サブネット内の適切なローカル・アドレスを使用して LRH を変更します。したがって、パス内の最後のルーターは、LRH 内の LID を宛先ポートの LID で置き換えます。ネットワーク層内では、単一のサブネット内で使用される場合（InfiniBand システム・エリア・ネットワークに起こりそうなシナリオ）、InfiniBand パケットには、ネットワーク層の情報とヘッダーのオーバーヘッドは必要ありません。

3.5.4 トランスポート層

トランスポート層は、適切なパケット配信、パーティショニング、チャネル多重化、およびトランスポート・サービス（高信頼性接続、高信頼性データグラム、低信頼性接続、低信頼性データグラム、ロー・データグラム）を担当します。また、トランスポート層は、送信時のトランザクション・データのセグメンテーション、および受信時の再アセンブリーも処理します。パスの最大伝送単位（MTU）に基づいて、トランスポート層は、適切なサイズの複数のパケットにデータを分割します。受信側は、宛先キュー・ペアとパケット・シーケンス番号を含む基本トランスポート・ヘッダー（BTH）に基づいて、パケットを再アセンブルします。受信側は、パケットの受信を確認し、送信側はこれらの受信確認を受け取り、完了キューを更新して操作の状況を反映します。InfiniBand アーキテクチャーは、トランスポート層を大幅に改善します。すなわち、すべての機能がハードウェアに実装されます。

InfiniBand は、データの信頼性を確保するために複数のトランスポート・サービスを指定します。表 3-3 では、サポートされる各サービスについて説明します。所定のキュー・ペアには、1 つのトランスポート・レベルが使用されます。

表 3-3 トランスポート・サービス

サービス・クラス	説明
高信頼性接続	確認応答あり - コネクション型通信
高信頼性データグラム	肯定応答 - 多重化
低信頼性接続	確認応答なし – コネクション型通信
低信頼性データグラム	確認応答なし - コネクションレス型通信
ロー・データグラム	確認応答なし - コネクションレス型通信

3.5.5 InfiniBand のエレメント

InfiniBand アーキテクチャーは、システム通信用の複数の装置（チャネル・アダプター、スイッチ、ルーター、およびサブネット・マネージャー）を定義します。サブネット内には、エンド・ノードごとに 1 つ以上のチャネル・アダプター、およびリンクをセットアップし、保守するサブネット・マネージャーが必要です。すべてのチャネル・アダプターとスイッチには、サブネット・マネージャーとの通信を処理するのに必要なサブネット管理エージェント（SMA）が含まれていなければなりません。

3.6 InfiniBand アーキテクチャー

ここでは、InfiniBand アーキテクチャーを構成する 2 つの項目について説明します。

3.6.1 チャネル・アダプター

チャネル・アダプターは、InfiniBand とその他の装置とを接続します。チャネル・アダプターには 2 つのタイプがあります。すなわち、ホスト・チャネル・アダプター（HCA）とターゲット・チャネル・アダプター（TCA）です。

HCA は、ホスト装置とのインターフェースを提供し、InfiniBand によって定義されるすべてのソフトウェア verb をサポートします。Verb は、クライアント・ソフトウェアと HCA の機能との間に必要なインターフェースを定義する抽象的な表現です。Verb は、オペレーティング・システムのアプリケーション・プログラミング・インターフェース（API）を指定するのではなく、使用可能な API を開発するために OS ベンダーの操作を定義します。

TCA は、各装置の特定の操作に必要な HCA 機能のサブセットを使用して、InfiniBand から I/O 装置への接続を提供します。

3.6.2 スイッチ

スイッチは、InfiniBand ファブリックの基本的なコンポーネントです。スイッチは、複数の InfiniBand ポートを含み、レイヤー 2 LRH 内に含まれている LID に基づいて、1 つのポートから別のポートにパケットを転送します。管理パケットを除いて、スイッチはパケットの消費も生成も行いません。スイッチは、チャネル・アダプターのように、サブネット管理パケットに応答するために SMA を実装する必要があります。スイッチは、ユニキャスト・パケット（単一のロケーションへ）か、マルチキャスト・パケット（複数の装置にアドレス指定）のどちらかを転送するように構成できます。

3.7 InfiniBand コンポーネント

ここでは、InfiniBand のコンポーネントについて説明します。

3.7.1 ルーター

InfiniBand ルーターは、パケットの消費も生成も行わずに、サブネット間でパケットを転送します。スイッチとは異なり、ルーターはグローバル経路ヘッダーを読み取って、その IPv6 ネットワーク層アドレスに基づいてパケットを転送します。ルーターは、次のサブネット上の適切な LID を使用して各パケットを再作成します。

3.7.2 サブネット・マネージャー

サブネット・マネージャーは、ローカル・サブネットを構成し、その継続動作を保証します。スイッチとルーターのすべてのセットアップを管理し、またリンクの停止または新規リンクの立ち上げ時にサブネットの再構成を行うために、サブネット内には少なくとも 1 つのサブネット・マネージャーが存在する必要があります。サブネット・マネージャーは、サブネット上の任意の装置内に存在することができます。サブネット・マネージャーは、（各 InfiniBand コンポーネントが必要とする）各専用 SMA を通じてサブネット上の装置と通信します。

アクティブであるのが 1 つのサブネット・マネージャーのみである限り、サブネット内に複数のサブネット・マネージャーが存在することが可能です。アクティブでないサブネット・マネージャー（スタンバイ・サブネット・マネージャー）は、アクティブ・サブネット・マネージャーの転送情報のコピーを保持し、アクティブ・サブネット・マネージャーが作動可能であることを確認します。アクティブ・サブネット・マネージャーが作動しない場合、スタンバイ・サブネット・マネージャーが、ファブリックが停止しないことを保証する責任を引き受けます。

3.7.3 管理インフラストラクチャー

InfiniBand アーキテクチャーは、すべてのサブネットの起動、維持、およびサブネット内の装置に関連した一般的なサービス機能の処理に関して、2 つのシステム管理方法を定義しています。各方法には、管理トラフィックをそれ以外のすべてのトラフィックと区別するために、サブネット上のすべての装置によってサポートされる専用キュー・ペア（QP）があります。

サブネット管理

第 1 の方法のサブネット管理は、サブネット・マネージャーによって処理されます。構成と保守を処理するために、サブネット内には少なくとも 1 つのサブネット・マネージャーが必要です。これらの責務には、LID 割り当て、SL-to-VL マッピング、リンクの立ち上げと解除、およびリンクのフェイルオーバーが含まれます。

すべてのサブネット管理は、QP0 を使用し、サブネット内の最高の優先順位を確保するために、優先順位が高い仮想レーン（VL15）で排他的に処理されます。サブネット管理パケット（SMP）のみが、QP0 および VL15 で許可されるパケットです。この VL は、低信頼性データグラム・トランスポート・サービスを使用し、リンク上の他の VL と同じフロー制御制約には従いません。サブネット管理情報は、リンク上の他のすべてのトラフィックより先に、サブネットを通じて渡されます。

サブネット・マネージャーは、すべての構成要件を引き受け、それらをバックグラウンドで処理することによって、クライアント・ソフトウェアの責務を単純化します。

汎用サービス

InfiniBand によって定義される 2 番目の方法は、General Services Interface（GSI）です。GSI は、シャーシ管理、アウト・オブ・バンド I/O 操作などの機能を始めとして、サブネット・マネージャーに関連しない機能を処理します。これらの機能には、サブネット管理と同じ高優先順位は必要ないので、GSI 管理パケット（GMP）は、優先順位が高い仮想レーン VL15 を使用しません。すべての GSI コマンドは QP1 を使用し、他のデータ・リンクのフロー制御要件に従う必要があります。

3.8 DAPL (Direct Access Programming Library) に対する InfiniBand サポート

DAPL (Direct Access Programming Library) は、ハードウェアから独立し、現行のネットワーク・インターコネクトとの互換性がある、分散メッセージング・テクノロジーです。アーキテクチャーが提供する API は、クラスター・アプリケーション内のピア間の高速低遅延通信を行うのに利用できます。

InfiniBand は、DAPL アーキテクチャーを念頭に置いて開発されました。InfiniBand は、実行キューの使用により、ソフトウェア・クライアントからトラフィック制御をオフロードします。これらのキュー（作業キューと呼ばれます）は、クライアントによって開始された後、管理が InfiniBand にゆだねられます。装置間の通信チャネルごとに、1 つの作業キュー・ペア（WQP - 送信キューと受信キュー）が各エンドに割り当てられます。クライアントは、トランザクションを作業キューに入れます（作業キュー・エントリー、WQE）。このトランザクションは、送信キューからチャネル・アダプターによって処理され、リモート装置に送信されます。リモート装置が応答すると、チャネル・アダプターは、完了キューまたはイベントを通じて、クライアントに状況を戻します。

クライアントは、複数の WQE を通知することができ、チャネル・アダプターのハードウェアは各通信要求を処理します。次に、チャネル・アダプターは完了キュー・エントリー

（CQE）を生成して、適切な優先順位順に各 WQE の状況を提供します。これにより、クライアントは、トランザクションの処理中に他の処理を続行できます。

3.9 ブレード・コンピューティングの潜在能力の実現

ブレード・ベース・サーバー・コンピューティングの TCO（総所有コスト）の利点を完全に実現するために、ブレード・テクノロジーは、少なくとも次の中心機能を提供する必要があります。すなわち、拡張容易性、フォールト・トレランス、ホット・スワップ、QoS、クラスタリング、I/O 接続のサポート（メモリーとメッセージのセマンティック）、信頼性、冗長性、フェイルオーバー用のアクティブ・スタンバイ、インターコネクトの管理性、およびエラー検出です。IT 管理者が新規に展開するサーバー・プラットフォームにこれらの属性を必要とする理由を理解するのは、非常に簡単です。本書で概要を説明したように、これらの属性はすべて、InfiniBand アーキテクチャー内に本来備わっているものであり、ブレード・コンピューティングが約束するすべての潜在能力を真に引き出します。ホワイト・ペーパー「Realizing the Full Potential of Server, Switch & I/O Blades with InfiniBand Architecture」（文書番号 2009WP）では、これらの属性と TCO の利点を詳しく説明しています。このホワイト・ペーパーは次の Web サイトでご覧ください。

http://www.mellanox.com/technology/shared/Blade_WP_120.pdf

3.10 要約

業界のリーダー達の総力により、InfiniBand は、テクノロジーのデモンストレーションから、最初の製品実動段階に移行することに成功しました。InfiniBand 仕様は成熟し、複数のベンダーがシリコンおよびシステム・ソリューションを出荷し、InfiniBand シリコン・ベンダー間のインターオペラビリティーが IBTA の優先事項になり、実証されてきました。InfiniBand アーキテクチャーの利点は明らかであり、RAS（信頼性 • 可用性 • 保守性）のサポート、筐体内（in-the-box）で機能すると共に筐体外帯域幅（Bandwidth Out of the Box）を可能にするファブリック、および将来に備えた拡張容易性が含まれます。

IBTA は、InfiniBand テクノロジーと、InfiniBand テクノロジーがサーバー、通信、およびストレージのインターコネクトとして作成するファブリックを通じて、データ・センターを改善し、単純化するというビジョンを抱いています。

すべてが密接にクラスター化されているサーバーで構成されるデータ・センターを想像してください。これらのサーバーには、ストレージに接続し DAPL InfiniBand ポートと通信するプロセッサーとメモリーのみがあります。これにより、RDMA クラスタリングを使用するプロセッサーのパフォーマンスが大幅に上昇し、プロセッサーとメモリーの密度が向上し（大部分の周辺装置はサーバー・ラックの外に出されているので）、（InfiniBand）I/O 帯域幅が大幅に増えます。特に、これらの改善点はすべて、RAS 用に設計されたアーキテクチャーに基づいています。では、PCI-X と PCI Express を使用し、InfiniBand サーバー・ブレードを通じて、これらの改善点により既存のサーバーを柔軟にアップグレードできる状態を想像してみてください。InfiniBand の急速な採用は続き、インターネットをより頻繁により高い帯域で使用する企業や消費者が増えるにつれて、InfiniBand が受け入れられつつあるでしょう。

4

Topspin InfiniBand Switch Module とホスト・チャネル・アダプター・カード

この章では、Topspin InfiniBand Switch Module とホスト・チャネル・アダプター・カードの機能について説明します。

4.1 Topspin InfiniBand Switch Module

IBM @server BladeCenter 用の Topspin InfiniBand Switch Module は、ご使用の BladeCenter シャーシ内のホストに InfiniBand スイッチング機能を追加します。ご使用の BladeCenter シャーシに 1 つまたは 2 つの Topspin InfiniBand Switch Module を追加し、BladeCenter ホストに HCA 拡張カードを追加すると、ホストはシャーシ内で InfiniBand を介して互いに通信できます。外部 InfiniBand ファブリックに Topspin InfiniBand Switch Module を接続すると、BladeCenter ホストは、InfiniBand ネットワークに接続するすべてのノードと通信できます。図 4-1 に Topspin InfiniBand Switch Module を示しています。

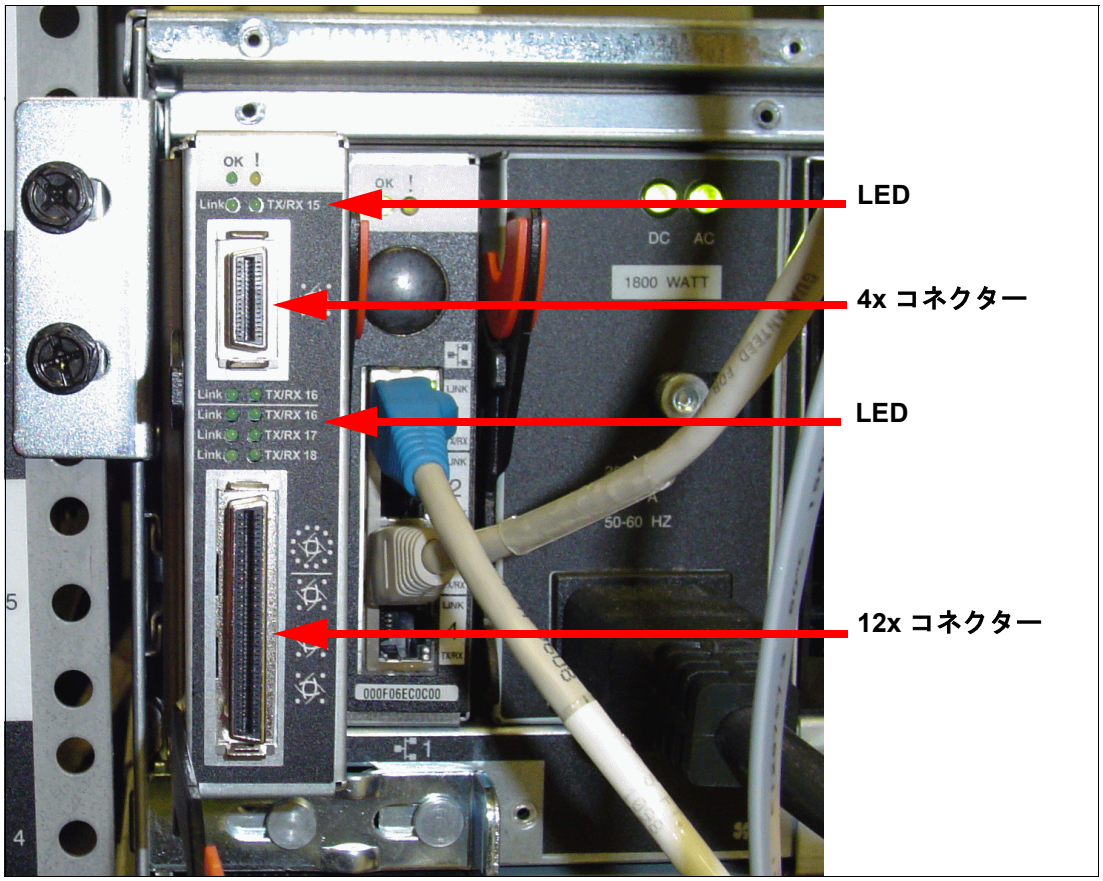

図 4-1 IBM eServer BladeCenter 用の Topspin InfiniBand Switch Module

Topspin InfiniBand Switch Module の基本的な目的は、サーバー・ブレード間および外部環境との InfiniBand 接続を提供することです。Topspin InfiniBand Switch Module は、次の基本機能を提供するのに必要です。

- 2.5 Gbps でのサーバー・ブレードに対する内部 1x ポート（BladeCenter では 14 枚のブレード、BladeCenter T では 8 枚のブレード）。
- 標準 InfiniBand 4x カッパー・ケーブル・コネクターを使用した、10 Gbps での装置または追加スイッチ接続用の外部 4x ポート。
- すべての InfiniBand ポートで完全な非ブロッキング機能を十分に提供できる処理能力。
- 2 つの内部全二重 100 Mbps イーサネット・リンク。それぞれ、Topspin InfiniBand Switch Module のセットアップと管理用の 2 つの管理モジュールに接続されます。
- 12 V 常時電力での稼動に必要な電力変換と電源調整。
- 管理モジュールから制御を提供し、管理モジュールに情報を提供するための I2C 回路。
- シャーシ内のスイッチ・ベイ位置を識別し、VPD を介してそれを提供する方法。
- 電源オン時に実行され、管理モジュールが読み取れるように増分の進行状況を保管する、診断機能。
- ローカル初期化と管理用の組み込み CPU サブシステム。
- 管理モジュールを通じてプログラミングできる、不揮発性構成およびコード・ストレージ。
- イーサネット管理ポートに対する SNMP のサポート。
- 標準 InfiniBand 管理データグラム（MAD）、管理情報ベース（MIB）、およびスイッチ・モジュール・モニター用の SNMP のサポート。
- 接続装置が対応する最高速度（1x/4x）にネゴシエーション可能な、自動検知機能を持つ外部ポート。
- InfiniBand スイッチ間でより高い帯域リンクを提供するために、1 つのスイッチから複数の外部ポートを結合する機能。
- 高信頼性、低信頼性、接続、および非接続のトランスポート・データ・トラフィック・タイプのサポート。
- 管理機能用の PowerPC® プロセッサーを内蔵。
- ネットワーク全体に分散知能を提供するサブネット・マネージャー。
- 最大 512 個のスイッチの相互接続により、中規模から大規模な InfiniBand ファブリックの作成を可能にする柔軟なトポロジー。新しいスイッチまたはリンクがファブリックに追加されたり、ユーザーのニーズが変化すると、このトポロジーは動的に変化することができます。
- 常にすべての InfiniBand ポートとの送信および受信のフル帯域幅。
- 負荷が大きいシステム内のスイッチ・パフォーマンス特性の最適化に使用できる、複数の高性能キューイング、メッセージング、およびバッファー・プール管理スキーム。
- カット・スルー・フレーム・ルーティングの使用によるポート間待ち時間の最小化。
- パーティション・キーのインプリメンテーションによりサポートされるサーバー・トラフィック分離、および許可ノード間のトラフィックの適切なルーティング。

Topspin InfiniBand Switch Module はフィールド・アップグレードに対応しており、現場で最新のファームウェアをロードすることが可能です。アップグレードに失敗した後であっても、ファームウェアのアップグレードを実行できます。CLI と GUI のインターフェースはどちらもアップグレードをサポートします。詳細については、「Command Line Interface Reference Guide」または「*InfiniBand User Guide*」を参照してください。

注 : Topspin InfiniBand Switch Module は、BladeCenter シャーシ内で相互に接続されません。サブネット・マネージャー・フェイルオーバーなどの機能を使用可能にするためにモジュールを接続するには、外部コネクターを介して InfiniBand ケーブルでモジュールを接続してください。シャーシ内の 2 つの Topspin InfiniBand Switch Module を接続する前に、モジュールでサブネット・マネージャーの優先順位を構成します。手順を追った説明については、121 ページの 8.4、『IP over InfiniBand の構成と 2 つの BladeCenter シャーシの接続』を参照してください。

4.1.1 LED

Topspin InfiniBand Switch Module は、モジュール状況 LED とポート状況 LED を備えています。図 4-2 は各種 LED を識別しています。

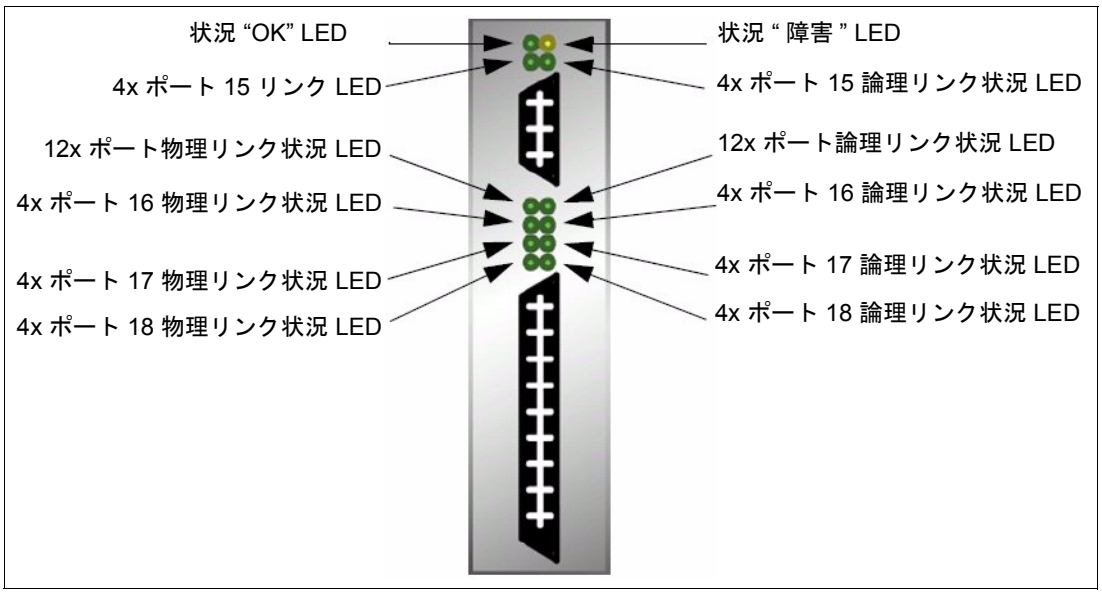

図 4-2 Topspin InfiniBand Switch Module LED

注 : このリリースの時点では、Topspin InfiniBand Switch Module は 12x ポート物理リンク LED と論理リンク LED をサポートしていません。

Topspin InfiniBand Switch Module 状況 LED

モジュール状況 LED では、Topspin InfiniBand Switch Module の稼働状況が一目で分かります。表 4-1 では、モジュール状況 LED の状態をリストし、説明しています。

表 4-1 モジュール状況 LED の説明

状態	指示
両方の LED がオフ	システムの電源が入っていないか、LED の誤動作。
黄色が点灯したまま、緑色がオフ	モジュール・エラーが検出されました。オペレーターの介入が必要です。
緑色が点灯したまま、黄色がオフ	エラーが検出されることなく、モジュールが稼働中です。

ポート状況 LED

Topspin InfiniBand Switch Module のポート LED は、接続と伝送の状況を示します（表 4-2）。

表 4-2 Topspin InfiniBand Switch Module のポート物理リンク状況 LED

状態	指示
オフ	インターフェースを介した論理接続がありません。
緑色に点灯	ドライバーがインストールされ、実行されている状態で、物理リンクが確立されていることを示しています。

リンク LED は接続の状況を示します（表 4-3）。トラフィック LED は伝送の状況を示します。

表 4-3 Topspin InfiniBand Switch Module のポート論理リンク状況 LED

状態	指示
オフ	インターフェース上で実行されるトラフィックはありません。
緑色に明滅	インターフェース上でトラフィックが正常に実行されています。

4.1.2 外部 InfiniBand ポート

3 つの 4x ポートをサポートするために、Topspin InfiniBand Switch Module の 12x InfiniBand コネクターは、オクトパス・ケーブルを使用した 3 つの 4x コネクターになります。このケーブルは、1 つの 12x コネクター（Topspin InfiniBand Switch Module に接続する）として開始した後、3 つの 4x コネクターに分岐し、追加の InfiniBand ハードウェアに接続できます。物理コネクターには、上から下の順に 1 から 4 の番号が付けられます（Topspin InfiniBand Switch Module の独立した 4x コネクターが一番上になります）。

しかし、すべてのユーザー・インターフェースでは、ポート番号は 15 から始まります。ポート 1 から 14 は内部で機能し、ポート 15 から 18 が Topspin InfiniBand Switch Module を外部装置に接続します。コネクター 1 はポート 15 にマップし、コネクター 2 はポート 16 にマップし、以下同様です。

BladeCenter 用の Topspin InfiniBand Switch Module を管理する方法については、50 ページの 6.1、『管理』を参照してください。

4.2 Topspin InfiniBand ホスト・チャネル・アダプター拡張カード

IBM @server BladeCenter 用の Topspin InfiniBand ホスト・チャネル・アダプター拡張カードは、BladeCenter 格納装置内のプロセッサー・ブレードへの InfiniBand I/O 機能を提供します。Topspin InfiniBand ホスト・チャネル・アダプター拡張カード（HCA）は、InfiniBand 対応の筐体内クラスター（cluster-in-a-box）を作成するために、CPU ブレードに 2 つの InfiniBand ポートを追加します。図 1-1 は、HCA 拡張カードを表示しています。

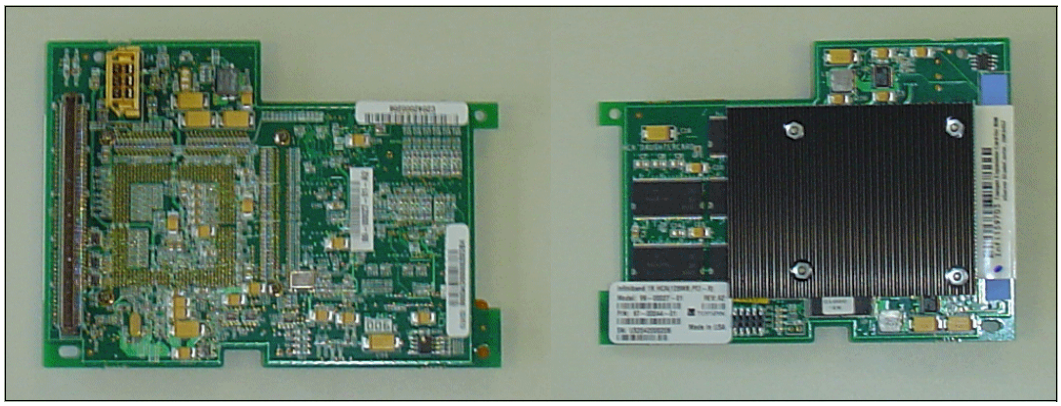

図 4-3　Topspin InfiniBand ホスト・チャネル・アダプター拡張カード（前面と背面）

Topspin InfiniBand ホスト・チャネル・アダプター拡張カードは、IBM @server® BlaeCenter 用の Topspin InfiniBand Switch Module を通じて相互に通信します。Topspin InfiniBand Switch Module の詳細については、32 ページの 4.1、『Topspin InfiniBand Switch Module』および「*Topspin InfiniBand Switch Module for IBM @server BladeCenter User Guide*」を参照してください。

Topspin InfiniBand ホスト・チャネル・アダプター拡張カードの機能は次のとおりです。

- デュアル 1x InfiniBand ブリッジとの 133 MHz PCI-X インターフェース。
- 2 つの 1x InfiniBand インターフェース。インターフェース回線速度は、リンク当たり 2.5 Gbps（理論上の最大値）です。
- 128 MB テーブル・メモリー（133 MHz DDR SDRAM）。
- システムの重要プロダクト・データ（VPD）を保持する I2C Serial EEPROM。
- IBM 専用のブレード・拡張カード・フォーム・ファクター。
- 既存の Topspin HCA ドライバーに対する同一操作とインターフェース。
- フラッシュ・リカバリー用のユーザー構成可能ジャンパー。
- 信頼性の高い稼働に適した強制空冷。
- 特定のプロトコル用のポート間フェイルオーバー（42 ページの 4.6、『InfiniBand プロトコル』を参照）。

4.3 Topspin InfiniBand ホスト・チャネル・アダプター拡張カードとスイッチ・モジュール

BladeCenter 内で、Topspin InfiniBand Switch Module は、Topspin InfiniBand ホスト・チャネル・アダプター拡張カードとの間のトラフィックを管理します。各 HCA ポートは、BladeCenter バックプレーンを経由して、特定の Topspin InfiniBand Switch Module ベイに接続します。各 ib0 ポートはベイ 3 の Topspin InfiniBand Switch Module に接続し、各 ib1 ポートはベイ 4 に接続します。Topspin InfiniBand Switch Module と HCA 拡張カードを使用すると、非冗長シングル・スイッチ・トポロジーまたは冗長デュアル・スイッチ・トポロジーを作成できます。

シングル・スイッチ・トポロジー

1 つの BladeCenter モジュール・ベイのみに Topspin InfiniBand Switch Module を取り付ける場合、2 等分帯域幅トポロジーを作成することになります。しかし、このトポロジーは、Topspin InfiniBand ホスト・チャネル・アダプター拡張カードから Topspin InfiniBand Switch Module までの冗長リンクを提供しません。単一障害点を避けるために、デュアル・スイッチ・トポロジーを構成することを強くお勧めします。

注 : ブレード・サーバーの特定のオペレーティング・システムでは、ベイ 3 に単一の Topspin InfiniBand Switch Module が存在する方が適しているものがあります。その他のオペレーティング・システムではベイ 4 が適しています。この選択は、オペレーティング・システムが ib0 としてどの HCA ポートを識別するかに応じて決まります。シングル・スイッチ・トポロジーでトラフィックが実行されない場合、モジュールをもう一方のベイに入れてください。

デュアル・スイッチ・トポロジー

BladeCenter で InfiniBand 冗長性を使用可能にするには、使用可能な各ベイに 1 つの Topspin InfiniBand Switch Module を取り付ける必要があります。Topspin InfiniBand ホスト・チャネル・アダプター拡張カードは、シングル・スイッチ・ベイとの冗長リンクをサポートしません。BladeCenter 格納装置に 2 つ目の Topspin InfiniBand Switch Module を追加する場合、各 HCA 拡張カードの各ポートには、1 つの Topspin InfiniBand Switch Module が接続されます。

注 : Topspin InfiniBand Switch Module は、BladeCenter シャーシ内で相互に接続しません。サブネット・マネージャー・フェイルオーバーなどの機能を使用可能にするには、外部コネクターを経由して IBM ケーブルでモジュールを接続してください。シャーシ内の 2 つのモジュールを接続する前に、モジュールでサブネット・マネージャーの優先順位を構成する必要があります。手順を追った説明については、121 ページの 8.4、『IP over InfiniBand の構成と 2 つの BladeCenter シャーシの接続』を参照してください。

4.4 Topspin サーバー・スイッチ

BladeCenter と Topspin InfiniBand Switch Module を外部スイッチに接続する場合、次の Topspin サーバー・スイッチのいずれかに接続します。

- **Topspin 90 サーバー・スイッチ** : このエントリー・レベルのスイッチを使用すると、IT 管理者は、分散データベースおよびハイパフォーマンス・グリッドまたはユーティリティー・コンピューティング用の費用対効果の高い 10 Gbps サーバー・クラスターを作成できます。その内蔵拡張スロットを使用すると、イーサネットまたはファイバー・チャネル・ゲートウェイを組み込むように Topspin 90 を容易に拡張できます。
- **Topspin 120 サーバー・スイッチ** : ハイパフォーマンス・コンピューティング・クラスターを構築するための価格、パフォーマンス、およびパッケージの組み合わせを提供します。完全に非ブロッキングであり、サブネット管理が組み込まれている Topspin 120 は、単一のコンパクトな 1U シャーシ内で 24 ポートの 10 Gbps 接続か、8 ポートの 30 Gbps 接続のどちらかを提供します。さらに、Topspin 90 または Topspin 360 のどちらか、およびそのイーサネットとファイバー・チャネル・ゲートウェイと組み合わせると、お客様は、ハイパフォーマンス・サーバー・クラスターとその LAN および SAN との間でシームレスな接続を提供できます。
- **Topspin 270 サーバー・スイッチ** : 信頼性 • 可用性 • 保守性（RAS）を最大化するように設計された、業界唯一の InfiniBand スイッチを提供します。完全な冗長性のあるホット・プラグ可能なコンポーネントと中断のないフェイルオーバーを備えた Topspin 270 は、拡張が容易で可用性が高いクラスターを構築するための、ディレクター・クラスの 96 ポート 4x または 32 ポート 12X スイッチを提供します。Topspin 90 または Topspin 360 のどちらか、およびそのイーサネットとファイバー・チャネル・ゲートウェイと組み合わせると、お客様は、ハイパフォーマンス・サーバー・クラスターとその LAN および SAN との間でシームレスな接続を提供できます。
- **Topspin 360 サーバー・スイッチ** : グリッドまたはユーティリティー・コンピューティングを配置するための前例のないレベルの可用性、拡張容易性、および管理容易性を提供します。スイッチ・モジュールは、数テラビットの内部帯域幅を提供し、モジュラー拡張モジュールは、単一の格納装置内で最大 72 ポートのファイバー・チャネル、イーサ

ネット、および InfiniBand を使用可能にして、成長に合わせた段階的な投資（pay-as-you-grow）の拡張容易性を提供します。

Topspin サーバー・スイッチの詳細については、次の Web サイトをご覧ください。

http://www.Topspin.com/solutions/index.html

4.5 Topspin VFrame ソフトウェア

VFrame は Topspin の仮想化ソフトウェア・スイートです。BladeCenter は、ポリシーおよびプロビジョニング・インテリジェンスをファブリックに組み込む、Topspin の VFrame ソフトウェアと互換性があります。この機能により、ファブリックは、データ・センター・マネージャーが定義するポリシーに基づいて、適切な組み合わせのリソースを適切な時点で相互接続して、自動的に仮想サーバーを作成することができます。

条件が変わると、必要に応じて仮想サーバーからリソースを追加および除去するようにスイッチをプログラムできます。

4.5.1 VFrame/Tivoli® Intlligent Orchestrator の統合

IBM Tivoli Intelligent Orchestrator によって管理される環境では、VFrame は InfiniBand リソースのリソース・プロバイダーの役目をすることができます。IBM と Topspin の共同チームは、Tivoli Intelligent Orchestrator と VFrame ソフトウェアが通信できるようにする 1 組の *リスナー* を作成し、テストしました（38 ページの図 4-4）。

注： Tivoli Intelligent Orchestrator の詳細については、次の Web サイトをご覧ください。

http://www-306.ibm.com/software/tivoli/products/intell-orch/

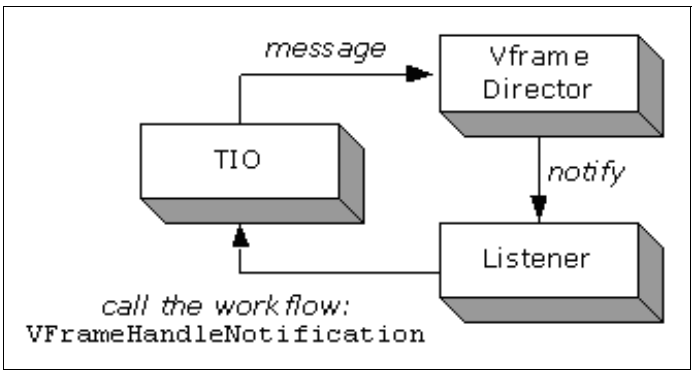

図 4-4　リスナー

リスナーをテストするために、IBM ブレード・サーバー、Topspin 360 InfiniBand スイッチ、および IBM TotalStorage DS4300 を使用して、マルチ・ファブリック I/O 環境がセットアップされました（図 4-5）。

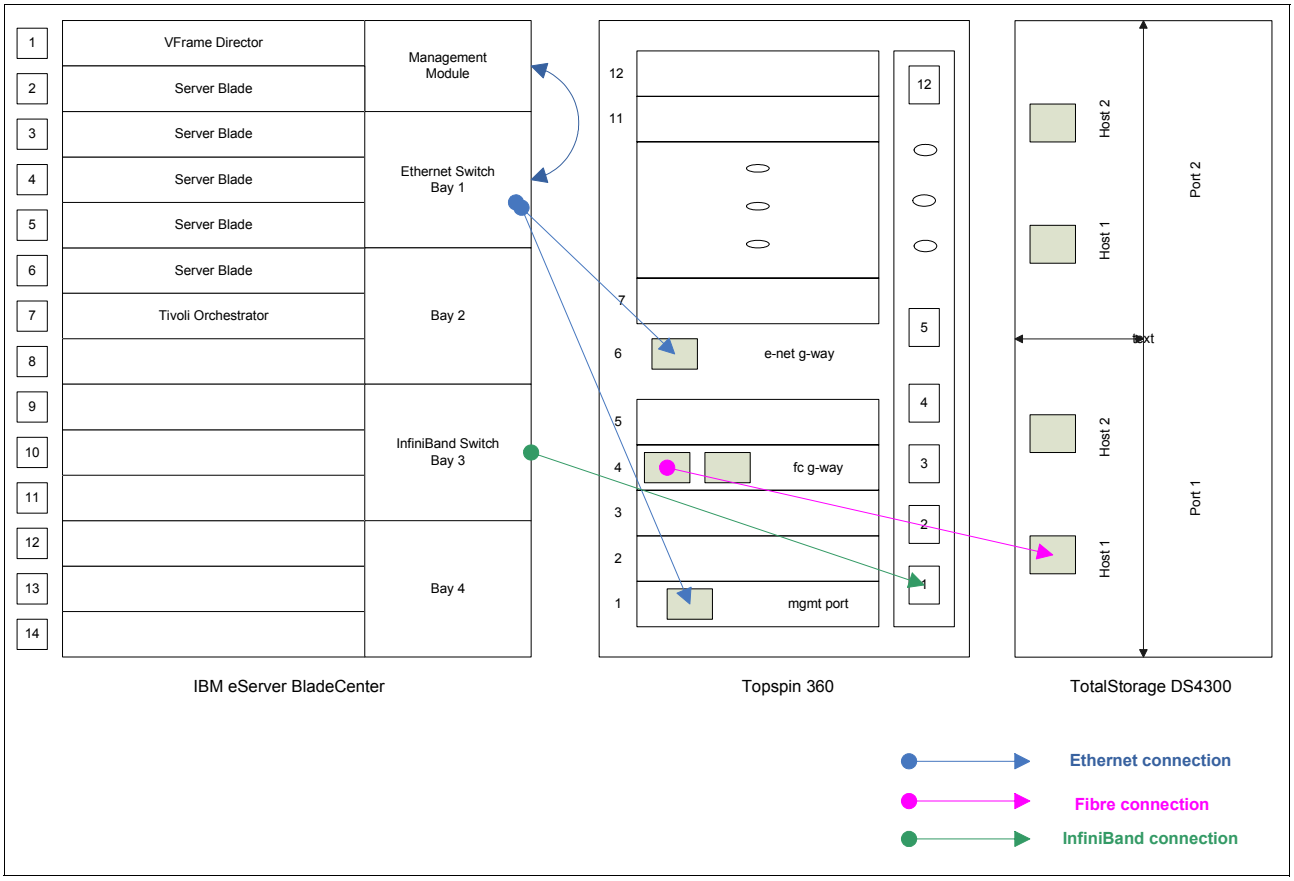

図 4-5　VFrame-Tivoli Intelligent Orchestrator のセットアップ

図 4-5 では、VFrame サーバー（VFrame Director）は、InfiniBand ファブリック上のすべてのリソースを提供します。

Tivoli Intelligent Orchestrator サーバーは、使用可能なすべてのリソース、アプリケーション、およびサービスを管理します。リソースの割り振りの作成と解除を行うコマンドを VFrame に提供します。

複数のディスクレス・ブレードが、使用可能なリソースとして使用されました。

Topspin 360 スイッチは、管理対象のブレードにマルチ・ファブリック I/O（ストレージとイーサネット接続）を提供します。

IBM TotalStorage DS4300 は、リソース・サーバー・ブレードによって使用されるイメージを保持します。

Tivoli Intelligent Orchestrator では、VFrame が提供するリソースのプールから（図 4-7）、新しい仮想サーバーを要求できます（図 4-6）。

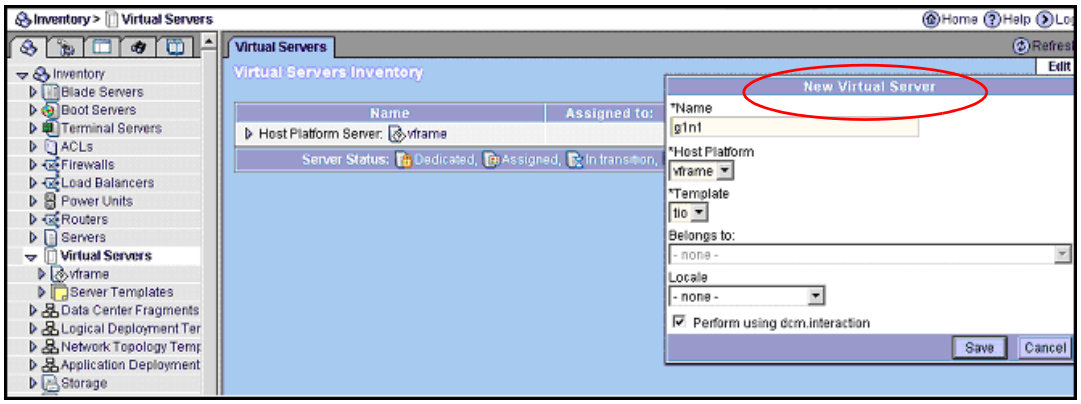

図 4-6　新規仮想サーバーの作成

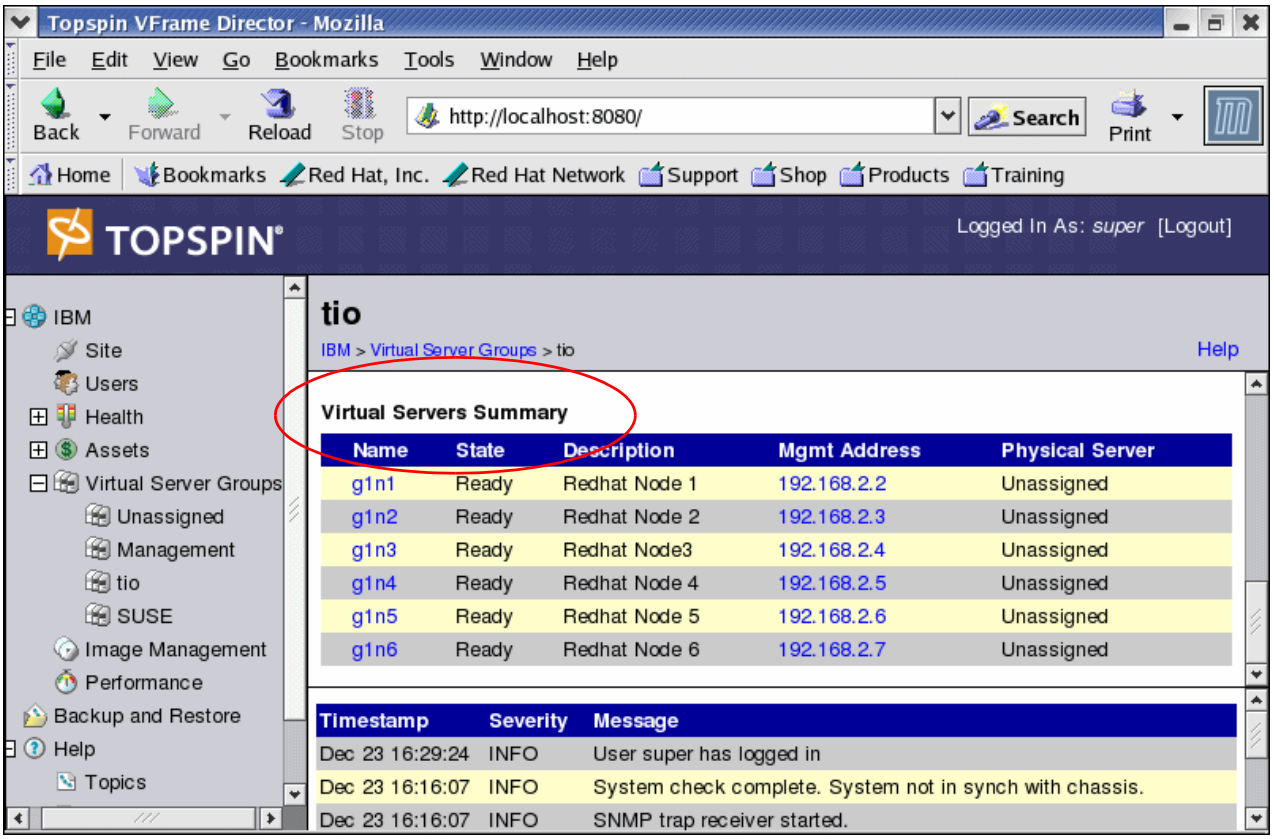

図 4-7　Tivoli Intelligent Orchestrator 仮想サーバーの要約

VFrame は、割り振りが解除されている物理サーバーの 1 つに仮想サーバーを割り当てます (41 ページの図 4-8)。

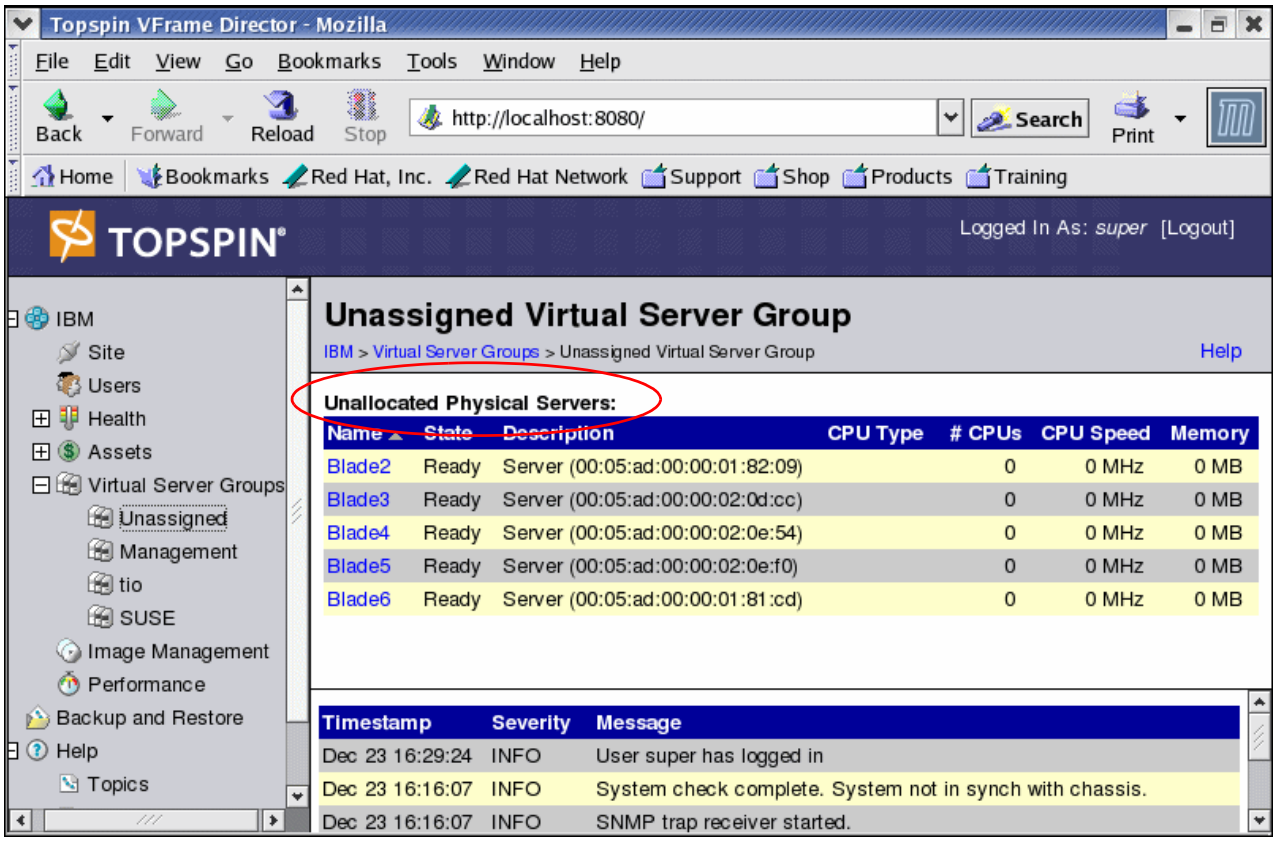

図 4-8　割り当て解除されたサーバー

次に、Tivoli Intelligent Orchestrator に通知を戻します（図 4-9）。

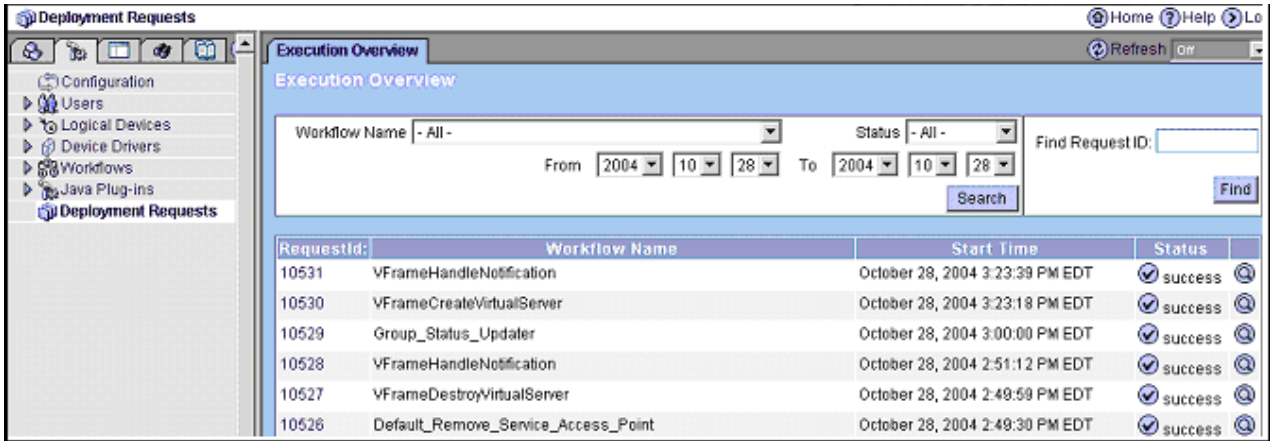

図 4-9　配置の結果

これで、サーバーは、Tivoli Intelligent Orchestrator によって管理される仮想サーバー・インベントリー内で使用可能になりました（42 ページの図 4-10）。

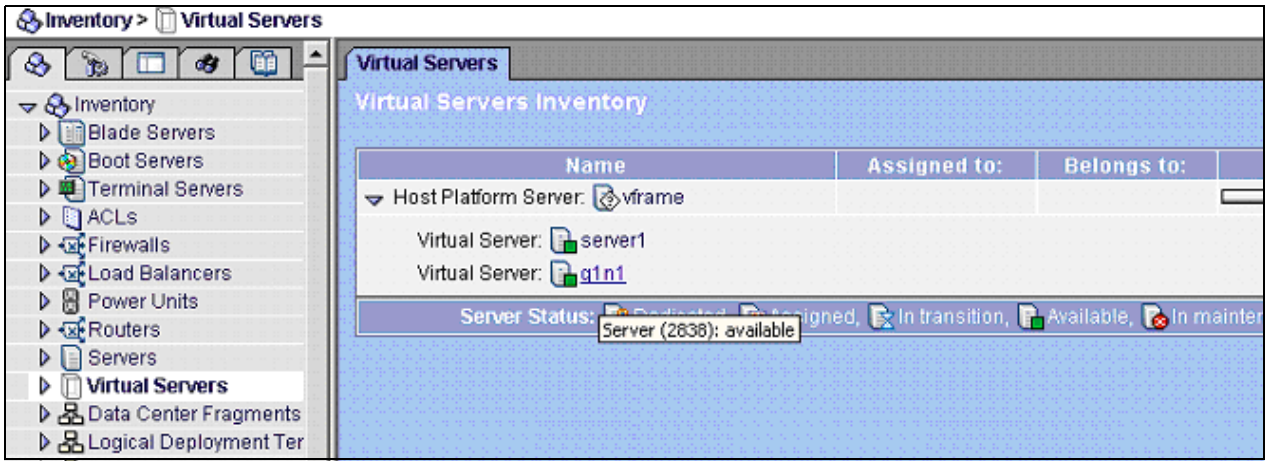

図 4-10　Tivoli Intelligent Orchestrator 仮想サーバー・インベントリー

4.6 InfiniBand プロトコル

InfiniBand は、複数の上位層プロトコル（ULP）をサポートします。このプロトコルを使用すると、異なる要件や目的を持つ様々なタイプのソフトウェアで InfiniBand を活用できるようになります。

4.6.1 Internet Protocol over InfiniBand（IPoIB）

IPoIB は、InfiniBand で提供される最下位レベルの既存ネットワーク・インターフェースです。IPoIB は、さまざまな業界標準アプリケーション、ミドルウェア、およびオペレーティング・システムをサポートします。IP 上に階層化される任意の通信プロトコル（つまり、TCP/IP や UDP/IP などの IP ベース・プロトコルを使用して通信するすべてのもの）は、IPoIB インターフェースを使用して InfiniBand 上でトランスポートできます。

4.6.2 Sockets Direct Protocol（SDP）

SDP は、プロトコル・スタックのやや上位にある既存インターフェースを提供します。TCP ソケット・インターフェースに書き込まれるアプリケーションの場合、InfiniBand 上で通信するために、すべての TCP および IP プロトコル層を実装する必要があるわけではありません。SDP プロトコルは、*非同期ソケット・インターフェース*を提供します。このインターフェースを通じて、アプリケーションとミドルウェアは、TCP または IP プロトコル層が下部にないことを理解する必要なく、通信することができます。

同期ソケットと非同期ソケットの相違点はわずかなものですが、SDP にとっては非常に重要です。これには、通信を要求するためのアプリケーション・ソフトウェアのプログラム方法、および要求の実行方法（ソケット・インターフェースの呼び出し）の詳細が含まれます。アプリケーションからの同期ソケット呼び出しでは、呼び出しから戻った後、要求された通信が完了している必要があり、関係するすべてのデータ・バッファーは必要なくなり、即時に再利用できます。非同期ソケットは、要求された作業をエンキューすることができ（一般にエンキューします）、要求が完了せず（多くの場合、開始もしていない）、データ・バッファーが引き続き使用中に、即時に戻ることができます。

非同期ソケットの使用は、Windows® オペレーティング・システムでは標準です。Windows プログラム用のソケット・ベース通信は常に非同期であるので、SDP 上で操作するための変更は必要ありません。従来、UNIX® および Linux® アプリケーションは、同期ソケットを使用してきたので、SDP プロトコルへの透過的なマップは行いません。この不一致に対応す

るために、アプリケーションとのソケット・インターフェース（同期）と SDP インターフェース（非同期）との間に同期 / 非同期変換プロセスを提供できます。この変換には、呼び出しからの戻りの遅延、データ・バッファーのコピー、または既存の完全な互換性を実現するために一部のパフォーマンスと効率を引き換えにするその他の方法を伴う場合があります。

さらに、現在の UNIX および Linux 環境では同期ソケットが開発され、活用されています。同期ソケットが標準になると、ソフトウェアはこのインターフェースをもっと活用するようになり、SDP または同種のプロトコルを使用してより効率の高いトランスポートにより容易に移行するでしょう。すでに多くのアプリケーションが代替オペレーティング・システムやプラットフォームにクロスコンパイルされているので、非同期ソケットを利用するためのコンパイル・オプションがすでに存在する可能性があります。

4.6.3 Storage RDMA Protocol（SRP）

これは、InfiniBand 用の業界標準ストレージ・プロトコルです。SRP は、InfiniBand 上での標準 SCSI プロトコル・トランスポートを提供します。アプリケーションまたはファイル・システム（多くの場合、オペレーティング・システムの一部）によって生成される SCSI 通信は、SRP インターフェースに直送され、このインターフェースが完全なレガシー・サポートを提供します。SCSI がファイバー・チャネル（FCP プロトコル）と TCP/IP（iSCSI プロトコル）上でトランスポートされるのと同じように、SCSI プロトコルと、SCSI で使用されるデータ構造を変更することなく、SRP over InfiniBand によりトランスポートされます。

SRP と FC、iSCSI との相違点は、デバイスの検出や一覧参照、マルチパス・サポート、パス・フェイルオーバーなどのネットワーク関連機能に関連するものです。SRP の上にネットワークの特徴をマップするには、2 つの技術ソリューションがあります。ファイバー・チャネルとそのネットワークやネーミング構成のエミュレーションと、iSCSI とそのネットワークやネーミング構成のエミュレーションです。

現在の InfiniBand ストレージ関連の実装、特に Topspin InfiniBand-FibreChannel ブリッジは、ファイバー・チャネル・エミュレーションを実装しています。これにより、既存のファイバー・チャネル SAN と新しい InfiniBand 接続ホスト・システムとの間に、最もシームレスなインターオペラビリティーが実現されます。ホスト・アプリケーションは、通常の SCSI 接続を確認し、Topspin ドライバーを通じて、通常のファイバー・チャネル接続と同じように SCSI 接続を処理することができます。ストレージ・ベンダー固有のロード・バランシングとフェイルオーバー・ドライバーは、異なるファブリック上で通信していることを認識することなく、SRP および混在 SRP-FCP SAN 上で動作します。

混在 SAN 環境がもっと複雑になり、ホストとストレージの両方が InfiniBand ファブリックと非 InfiniBand （ファイバー・チャネルまたは iSCSI）ファブリックの両方で広がると、ファイバー・チャネル・エミュレーション・モードでそれらをすべて一緒にブリッジすることは、もっと困難になります。

Voltaire 社は、IBM やその他のベンダーの協力を得て、iSCSI に似たネットワークの標準化をリードしてきました。iSCSI が開発されたときに、ディスカバリー、ネーミング、許可、および大部分のファブリック認識機能用の明確で相互運用可能なメカニズムを提供するために、非常に豊富な機能セットが設計されました。iSCSI とそのサービス（iSCSI ネーム・サービス（iSNS）など）を利用することによって、SRP は、InfiniBand ファブリック上の装置が、名前をファブリック間でマップする方法を理解する共通のネーム・スペースを使用して、ファイバー・チャネルまたは iSCSI ファブリック上の装置を認識できるように、標準のネーミング方式を選択します。基礎として iSCSI 類似のメカニズムを使用することにより、このような複数の機能が使用可能になり、単純化されます。これは、ファイバー・チャネル接続のエミュレーションを妨げるものではなく、はるかに機能性の高いモードを使用可能にします。このモードでは、アプリケーションは、ファブリック間の違いを理解し明らかにする必要なく、混合 SAN ファブリック環境を認識することができます。

さらに iSCSI に似た SRP インプリメンテーションが、2005 年に使用可能になると予想されています。

4.6.4 ユーザー・レベルの Device Access Programming Layer（uDAPL）

ユーザー・スペース・アプリケーション用の DAPL（Direct Access Programming Layer）プロトコル（uDAPL）、およびカーネル・モード用の DAPL（kDAPL）は、高効率、低遅延のサーバー間通信用の比較的新しい業界標準です。uDAPL は、InfiniBand で使用可能な最低遅延、最高帯域幅、および最高効率の標準プロトコルの 1 つです。MPI は、現在、uDAPL よりやや優れていますが、さらにパフォーマンスが調整されれば、uDAPL が首位の位置を占める可能性があります。uDAPL は、低遅延の商用アプリケーションにおける新しい開発用に適したインターフェースです。データベース・ベンダー（RAC 10i を使用する Oracle など）は、スケールアウト・データベース・クラスタリングに uDAPL を利用します。

uDAPL は、他のプロトコル・インプリメンテーションの基礎サポートとしても使用されています。特に、Scali 社は、InfiniBand 上の uDAPL の上に MPI スタックを階層化することによって、InfiniBand に対する MPI サポートを提供してきました。Scali は、uDAPL を小規模で単純な共通層と見なし、その層の上に、MPI は有効なパフォーマンスと効率の利点を提供できます。

kDAPL は、カーネル・モード・アプリケーション、または OS 自体から使用するための類似インターフェースです。主な用途は、ファイル・システム・インターフェースであり、NFS/RDMA、iSER、およびその他のストレージ・インターフェースは、直接 kDAPL を通過できます。

IT API は、商用アプリケーション用の共通 API の標準化に取り組む業界標準のイニシアチブです（IBM も所属しています）。現時点では、uDAPL がこの機能を提供しています。この取り組みは、ある時点で別のインターフェースに変更される可能性があります。しかし、uDAPL の背景にある推進力により、おそらく*新しい*インターフェースは、現在のインターフェースに基づくものになるでしょう。

4.6.5 Message Passing Interface（MPI）

MPI は、科学的および技術的なハイパフォーマンス・クラスタリング（HPC）用の標準プロトコルです。MPI は、イーサネット、Myrinet、Quadrics、および InfiniBand を含む、多くの異なるクラスター・ファブリックで利用できます。MPI は、InfiniBand で使用可能なすべての標準プロトコルの中で最低の遅延、最高の帯域幅、および最高の効率を提供します。

大部分の HPC Linux 環境では MPI プロトコルが利用されています。この市場の大部分は、世界中の国立研究所や大学の研究機関で行われた研究に先導されています。オハイオ州立大学（OSU の Panda 教授）で開発された MPI スタック（MPICH）は、Topspin InfiniBand ソリューションで最も一般的に推奨される MPI 実装方法です。

5

Topspin InfiniBand Switch Module アーキテクチャー

この章では、Topspin InfiniBand Switch Module アーキテクチャーについて説明します。

5.1 InfiniBand アーキテクチャー

図 5-1 は、IBM @server BladeCenter プラットフォームにおける InfiniBand アーキテクチャーの上位トポロジー・ビューを示しています。ミッドプレーンは、各サーバー・ブレード上の InfiniBand 拡張カードと Topspin InfiniBand Switch Module との間の InfiniBand 接続を行います。また、ミッドプレーンは、管理モジュールと Topspin InfiniBand Switch Module との間の ENET 接続も提供します。

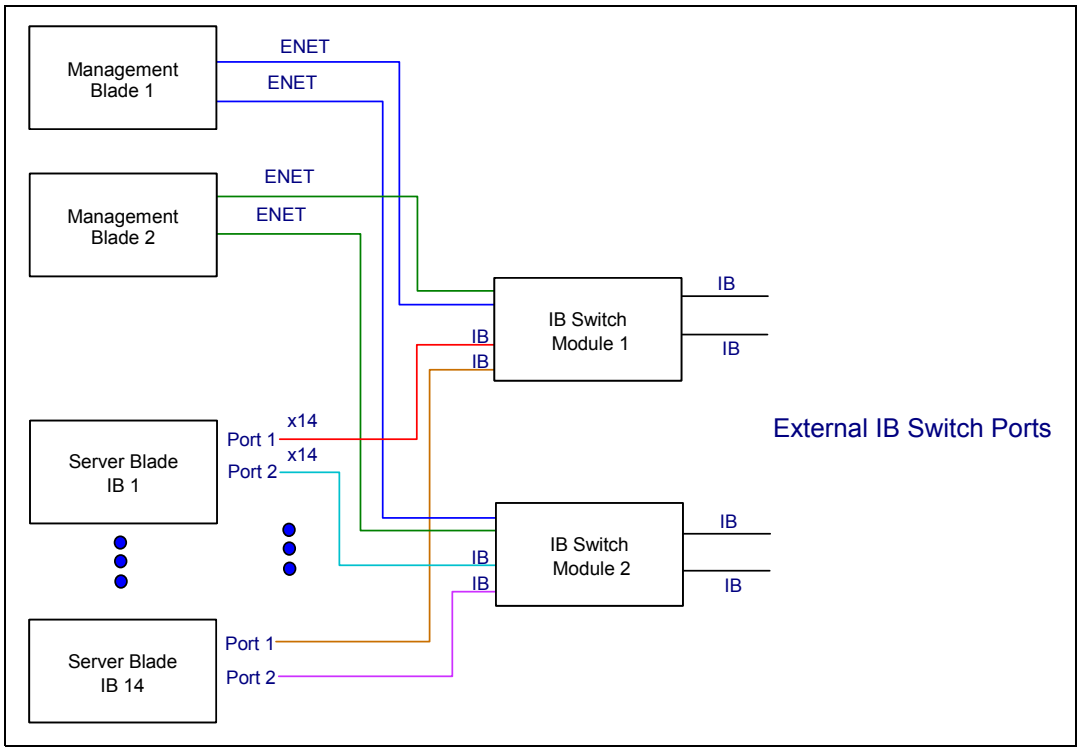

図 5-1　上位の BladeCenter InfiniBand トポロジー

管理モジュールでは、BladeCenter シャーシを一箇所で制御できます。各管理モジュールには、4 つの各スイッチ・ベイとの ENET 接続があります。管理モジュールは、ENET リンクを使用して、コマンド・ライン・インターフェース（CLI）、Web インターフェース、およびその他の IP ベース・インターフェース（Telnet や SNMP など）にアクセスできるようにします。これらの接続は、取り付けられているスイッチの構成とモニターに使用されます。

さらに、管理モジュールには、特定のスイッチ・モジュールが取り付けまたは取り外しされたときを認識するための存在検出ビットがあります。また、スイッチ・モジュールから下位レベル情報と状況を収集し、大部分の基本電源制御およびスイッチ初期化を行うのに使用される I2C インターフェースもあります。例えば、管理モジュールは I2C インターフェースを使用して、始動時に各スイッチに IP アドレス情報を提供します。

5.1.1 スイッチ・インターフェース

Topspin InfiniBand Switch Module には、次の 4 つのシグナル・インターフェースがあります。

1. 4 つの 4x 外部 InfiniBand ポートと 14 個の 1x 内部 InfiniBand ポート
2. 交換内部 I2C インターフェース
3. 2 つの内部 100 Mbps 全二重 ENET ポート
4. 1 つの RS232 シリアル・ポート・インターフェース（開発のみ）

外部 InfiniBand 4x ポート

Topspin InfiniBand Switch Module は、4 つの外部 4x InfiniBand ポートを提供します。

内部 InfiniBand 1x ポート

スイッチを BladeCenter シャーシに組み込むと、通常はサーバー・システムに必要なケーブルを省くことができます。したがって、内部 InfiniBand 接続は、論理レベルで、同じように動作し、可能なすべての InfiniBand 接続をシャーシから得る手段を提供します。内部 InfiniBand インターフェースは、14 個のすべてのサーバー・ブレード・インターフェース用の 1x（2.5 Gbps）ポートです。これらの内部 InfiniBand シグナルの物理インターフェースは、カッパー SerDes インターフェースです。

内部 I2C インターフェース

I2C アーキテクチャーは必要条件です。（このアーキテクチャーは、すべての BladeCenter スイッチ製品に共通です。）スイッチと管理モジュールとの間の I2C インターフェースは、スイッチ・モジュールから状況およびその他のシステム管理情報を収集します。

内部 100 Mb イーサネット・インターフェース

Topspin InfiniBand Switch Module には 2 つの内部 ENET ポートがあります。2 つのイーサネット・ポートは、BladeCenter 管理モジュールに接続し、2 つの物理ポートまたは NIC とは対照的に 1 つの論理ポート（例えば、1 つの IP アドレス）を表します。これらの管理イーサネット・インターフェースは、Telnet、FTP、または HTTP トランザクションを介して通常の操作で Topspin InfiniBand Switch Module の構成を設定し、管理するために管理モジュールで使用されます。これらのポートには、固定 100 Mbps 全二重 ENET 構成があります。物理インターフェースは、カッパー*無磁気*接続です。スイッチ・モジュールは、デフォルト設定を使用してイーサネット・インターフェースを初期化します。

デュアル管理モジュールを使用するオプションにより、シャーシ内の単一障害点が排除されます。該当するマスター管理モジュールをイーサネット管理ポートに接続するには、Topspin InfiniBand Switch Module 上に選択された回路が必要です。MM_SELECT_A および MM_SELECT_B シグナルが、その選択を行うのに使用されます。この BladeCenter 管理インターフェースのアーキテクチャーの実装については、36 ページの 4.3、『Topspin InfiniBand ホスト・チャネル・アダプター拡張カードとスイッチ・モジュール』を参照してください。

RS232 シリアル・ポート（開発のみ）

Topspin InfiniBand Switch Module は、オプションとして、ケーブルまたはボード・コネクターを介して、なんらかの形の標準シリアル・インターフェース接続を提供できます。このインターフェースは、スイッチ・ファームウェアの Telnet インターフェースを通じて、スイッチ・モジュールの開発制御と構成を提供するためのものです。このインターフェースが提供される場合でも、お客様または現場担当者からは見ることも、使用することもできないようになっています。

6

Topspin InfiniBand Switch Module ユーザー・オリエンテーション

この章では、BladeCenter 用の Topspin InfiniBand Switch Module の構成と管理に使用したさまざまな管理ツールについて説明します。50 ページの 6.1、『管理』の節では、Element Manager と Chassis Manager 、およびラボ環境の構築と構成に役立てるためのこれらの Manager の使用方法の例を示します。これらの管理ツールの使用について詳しくは、「InfiniBand User Guide, Release 2.1.0」をご覧ください。

6.1 管理

次のいずれかのインターフェースを使用して Topspin InfiniBand Switch Module を管理できます。

- Element Manager の Java ベース GUI
- Chassis Manager の Web ベース GUI
- TopspinOS コマンド・ライン・インターフェース（CLI）
- Topspin の管理情報ベース（MIB）を使用する Simple Network Management Protocol（SNMP）バージョン 1、2、および 3
- API（SNMP を経由）

Topspin InfiniBand Switch Module および関連した Topspin 装置の管理には、Topspin Element Manager の使用をお勧めします。以下の節では、Element Manager と Chassis Manager について説明します。

上記の管理ツールの使用について詳しくは、「*Topspin InfiniBand User Guide*」を参照してください。

6.2 Element Manager: Topspin InfiniBand Switch Module

ここでは、まず Element Manager を十分に理解します。Element Manager を使用すると、トラブルシューティング・アクティビティーを構成、モニター、および実行するために、Topspin 装置を視覚的に管理できます。

Element Manager は、次のプラットフォームで実行されます。

- Windows NT™/200/XP
- Solaris
- Linux

Element Manager ソフトウェアがインストールされた後、いくつかの手順を実行して Element Manager にアクセスできます。

ラボでは、Element Manager を Windows プラットフォームにインストールしました。

1. デスクトップで Element Manager アイコンをクリックするか、「**スタート**」→「**プログラム**」→「**Topspin Element Manager**」→「**TopspinEM**」の順にクリックする。これで、図 6-1 のようなウィンドウが開きます。

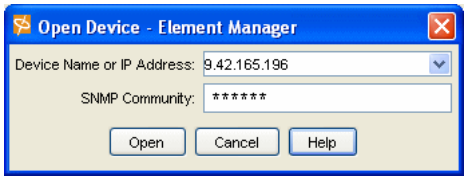

図 6-1　9.24.165.196 はスイッチ・モジュールです

2. 「Device Name or IP Address」フィールドに、管理ポートの IP アドレスまたはネットワーク名を入力し、「SNMP Community」ストリングを入力する。この例では、Topspin InfiniBand Switch Module（アウト・オブ・バンド管理ポート）の IP アドレスを入力しました。

注 : 図 6-1 では、アウト・オブ・バンド管理ポートまたはインバンド管理ポートを入力できます。

「**Open**」をクリックして、51 ページの図 6-2 のようなウィンドウを表示します。

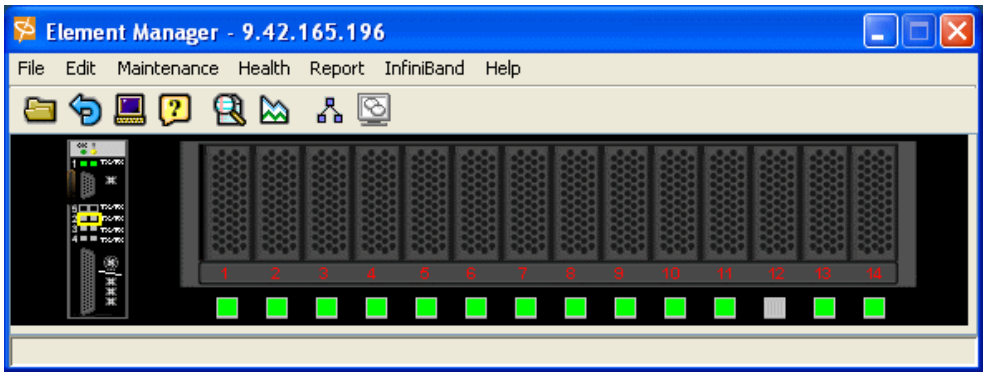

図 6-2 Element Manager スイッチ・モジュールとビュー

このビューには、Topspin InfiniBand Switch Module との Element Manager インターフェースが表示されています。左側に実際のスイッチ、右側にブレードが表示され、メニュー・バーとクイック起動アイコンがあります。図 6-3 では、各アイコンを拡大し、説明しています。

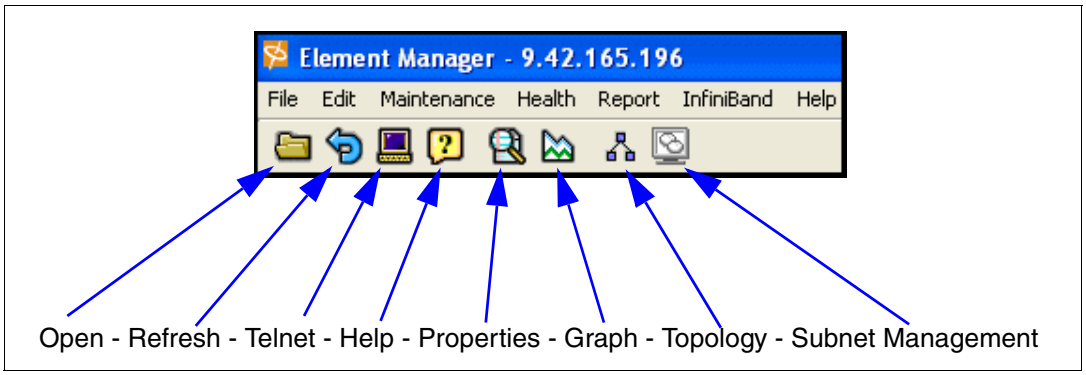

図 6-3 グラフィカル・アイコン

52 ページの図 6-4 は、Topspin 360 サーバー・スイッチとの Element Manager インターフェースを示しています。Topspin 360 サーバー・スイッチにアクセスすると表示される、Storage Manager のグラフィカル・アイコンに注目してください。

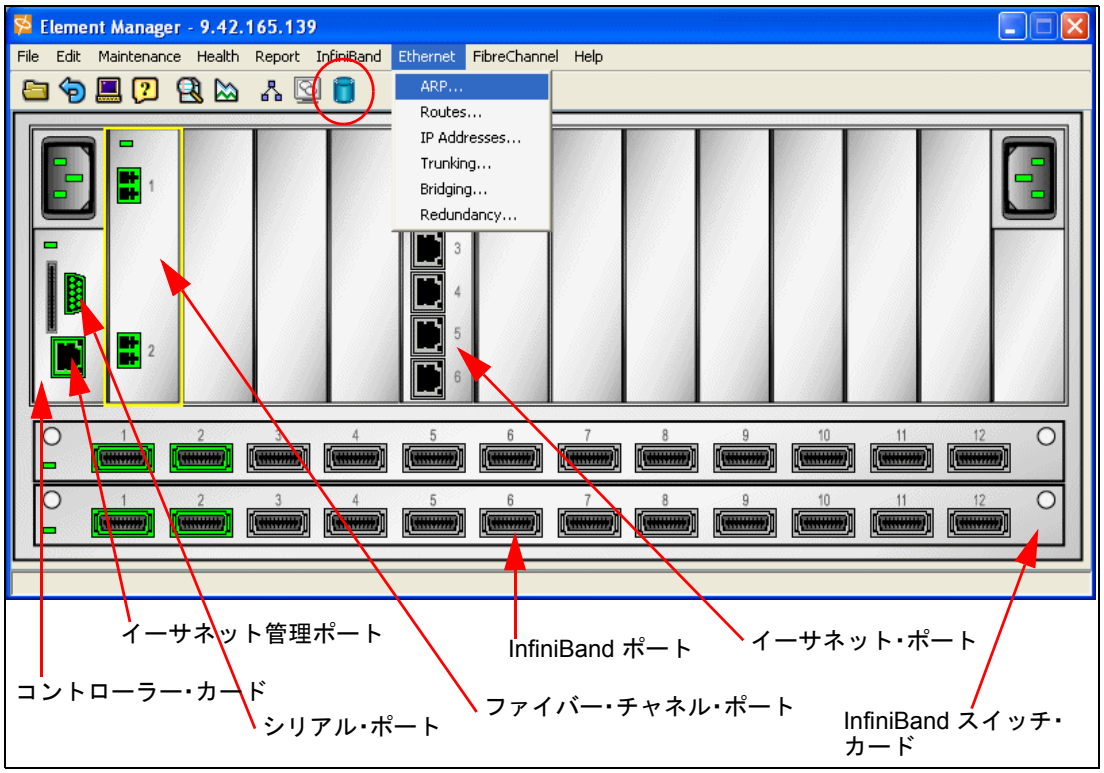

図 6-4 Element Manager: Topspin 360 サーバー・スイッチ・ビュー

以下の節の図は、Topspin シャーシ（サーバー・スイッチ）と Topspin InfiniBand Switch Module を参照しています。

6.2.1 Element Manager: 「File」メニュー

「File」メニューには、Topspin 装置との接続の確立、Element Manager ウィンドウの最新表示、および CLI バッチ・コマンドを使用して Topspin シャーシの一部を構成するための管理ポートとの Telnet 接続などの、一般的な管理機能の実行に使用される項目が含まれています。

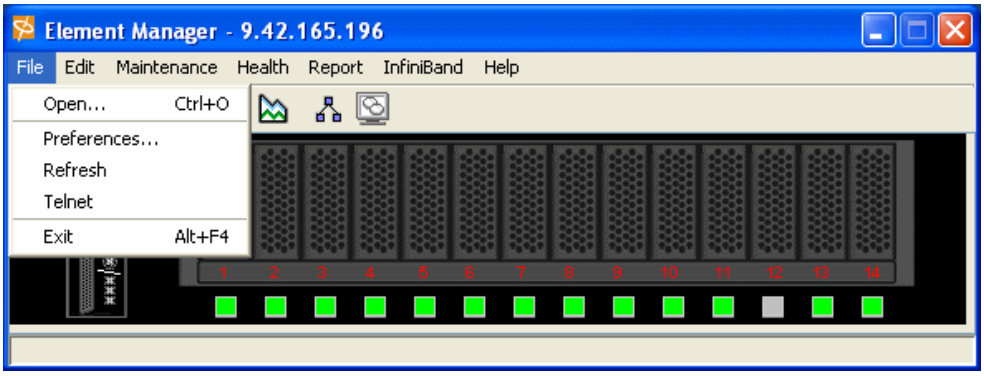

図 6-5 Element Manager: 「File」メニュー

BladeCenter 内の Topspin InfiniBand スイッチ・モジュールなどの装置をオープンして Element Manager セッションを開始するか、Topspin 外部スイッチにアクセスするには、「**File**」→「**Open**」をクリックします（50 ページの図 6-1）。次に進むには、以下の情報を入力します。

- DNS 名または IP アドレスを使用して装置名にアクセスします。
- SNMP コミュニティー : 構成データの表示および変更用の特権を判別する、割り当て済みのユーザー・コミュニティー・ストリングを入力します。

Polling、SNMP、および Misc のアクティビティーを実行するには、「**Preferences**」（52 ページの図 6-5 に表示）を選択してください。

- 「Polling」タブ（図 6-6）は、一般的な状況およびホット・スワップ・インターフェース・ボードの状態の変更がないかどうか、Topspin シャーシを検査する間隔を指定します。

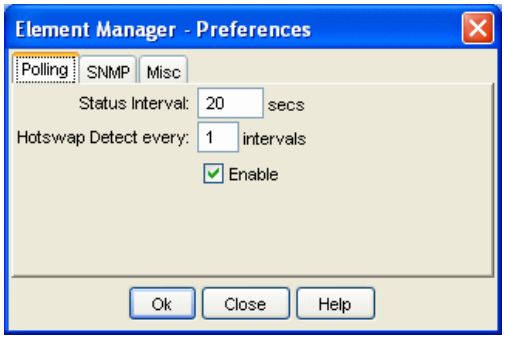

図 6-6 Element Manager: Preferences: 「Polling」タブ

- 「SNMP」タブ（図 6-7）では、一般的な SNMP セッション・パラメーターを設定し、Topspin シャーシに接続できない場合に SNMP をデバッグする手段を提供します。

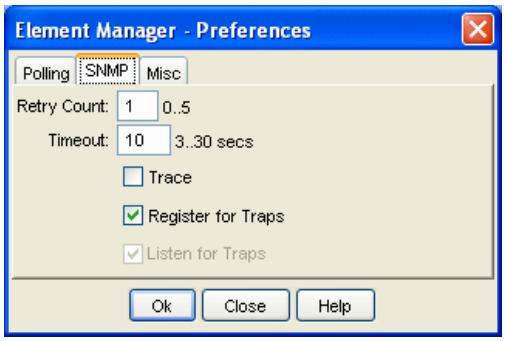

図 6-7 Element Manager: Preferences: 「SNMP」タブ

- 「Misc」タブ（図 6-8）は、次のことを行うために一般的な構成パラメーターを設定します。
  - トラップ・キャプチャーを調整する。
  - テーブル行の削除を確認する。
  - ログイン信用証明情報を保管する。

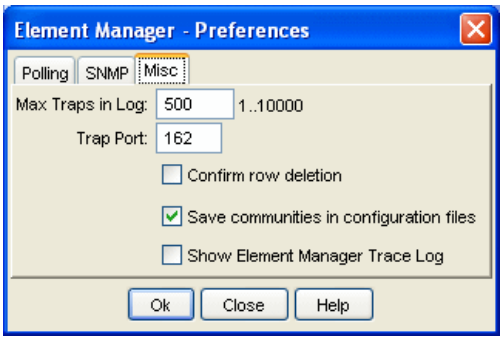

図 6-8 Element Manager: Preferences: 「Misc.」タブ

6.2.2 Element Manager: 「Edit」メニュー

「Edit」メニューの項目は、物理的な Topspin シャーシのインターフェース・ポート、カード、および部分の構成に使用されます。

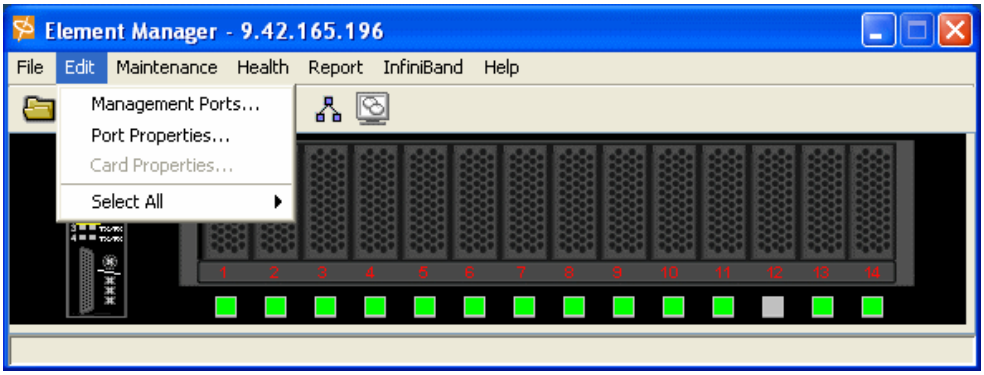

図 6-9 Element Manager: 「Edit」メニュー

プルダウン・リストには、次の項目があります。

- 「Card Properties」と「Port Properties」（図 6-9）は、1 つ以上のカードまたはポートが選択されるまでぼかし表示されます。
- 「Select All」メニュー項目は、特定タイプのすべてのポートまたはカードを選択するためのリストを開きます。
- 「Management Ports」は図 6-10 のようなウィンドウを表示します。

「Serial Port」タブは、シリアル・コンソール・ポートの構成を表示します。このタブが開くのは、選択されたオブジェクトがコントローラー・カード上のシリアル・コンソール・ポートである場合です。現在シリアル・コンソール・ポートに接続されているホスト上の端末ウィンドウに割り当てられている接続属性を表示します。

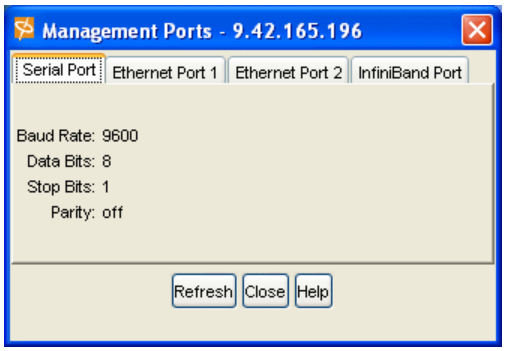

図 6-10　Management Ports:「Serial Port」タブ

「Ethernet Port」タブは、選択された管理ポートの構成を表示します。このタブが開くのは、選択されたオブジェクトが管理イーサネット・ポートである場合です。図 6-11 では、イーサネット・ポート 1 がアクティブ（up）であり、イーサネット・ポート 2 が非アクティブ（down）です。

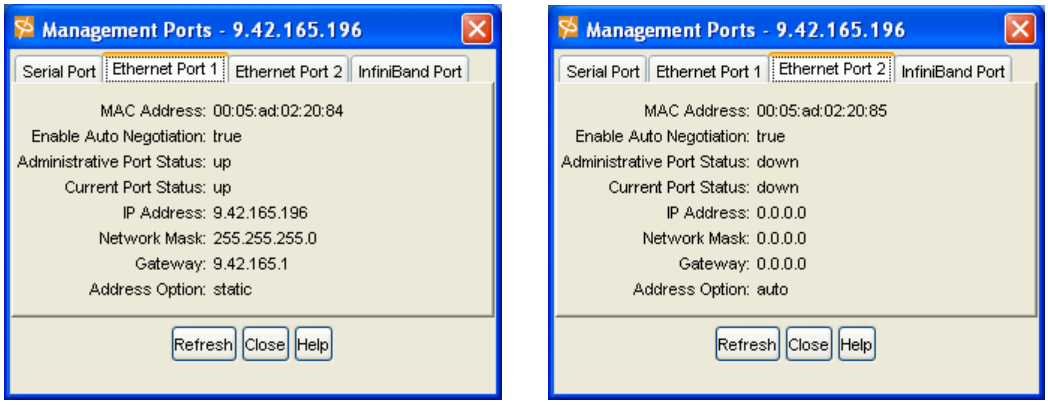

図 6-11　Management Ports:「Ethernet Port 1」と「Ethernet Port 2」タブ

「InfiniBand Port」タブ（図 6-12）は、構成されている InfiniBand 管理ポートがある場合にその構成を表示します。

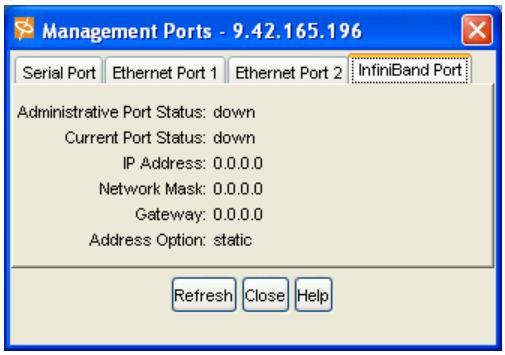

図 6-12　Management Ports:「InfiniBand Port」タブ

「InfiniBand Port」タブは、論理オブジェクトであるので、直接開くことはできません。管理 InfiniBand ポートのプロパティー・タブを開くには、コンソール・ポートか、管理イーサネット・ポートのどちらかをダブルクリックします。（これは、実際には別の装置で実行されます。52 ページの図 6-4 を参照してください。）開いたウィンドウで「**InfiniBand Port**」タブを選択します（56 ページの図 6-13）。

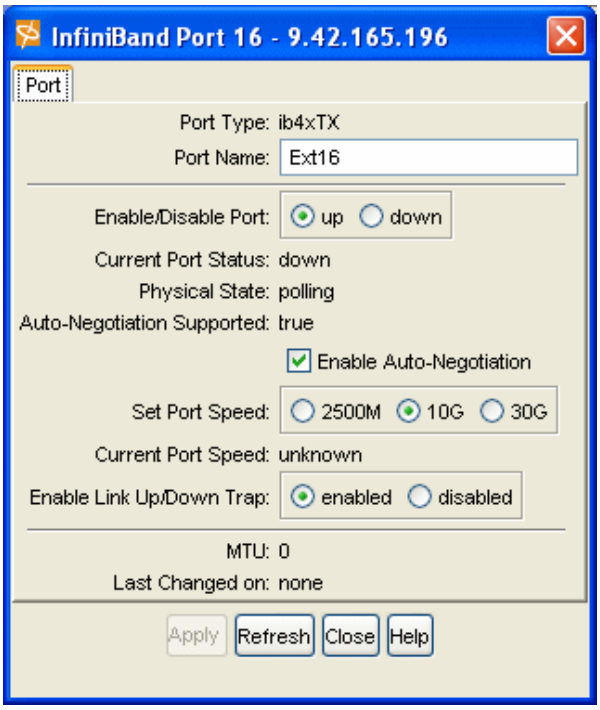

図 6-13 InfiniBand Port のプロパティー

「InfiniBand Port」タブは、選択された InfiniBand スイッチ・ポートの構成を表示します。

6.2.3 Element Manager: 「Maintenance」メニュー

「Maintenance」メニュー（図 6-14）は、一般的なシャーシ情報、システム時刻、および構成データの保守に使用されます。

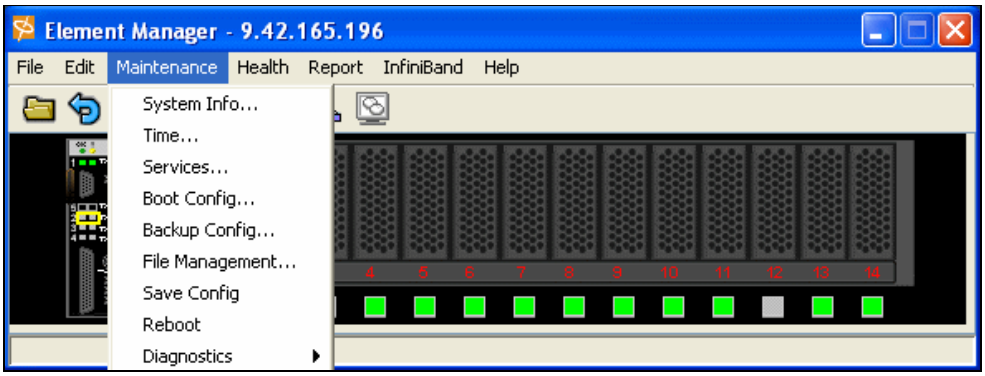

図 6-14 Element Manager: 「Maintenance」メニュー

「Maintenance」メニューは次の目的に使用されます。

- シャーシとソフトウェアのバージョンを記述する。
- 企業と連絡先の情報を提供する。
- システム・クロックを設定する。
- システムのブート時に使用するイメージと構成ファイルを指定する。
- 現行の構成を保管する。
- イメージ・ファイル、構成ファイル、およびログ・ファイルを管理する。
- システム・ファームウェアを再初期化する。

「System Info」タブ（図 6-15）は、次のことを行います。

- Topspin InfiniBand Switch Module または Topspin シャーシについての一般的な情報を表示する（例えば、実行しているソフトウェアのバージョン、前回のリブートからの間隔）。
- 前回の変更が行われた時間を示す。
- ユーザー連絡先情報を設定する。

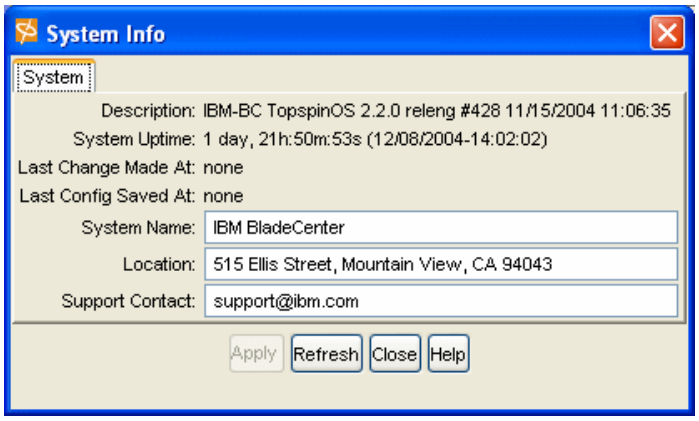

図 6-15　Element Manager: Maintenance: System Info

「Date and Time」ウィンドウ（図 6-16）は、Topspin シャーシと、Network Time Protocol（NTP）を実行している 1 つ以上のサーバーのクロックとを同期するのに使用されます。これにより、クロックの正確さが保証され、その結果、Topspin シャーシで生成されるすべての時刻ベースの統計の正確さが保証されます。

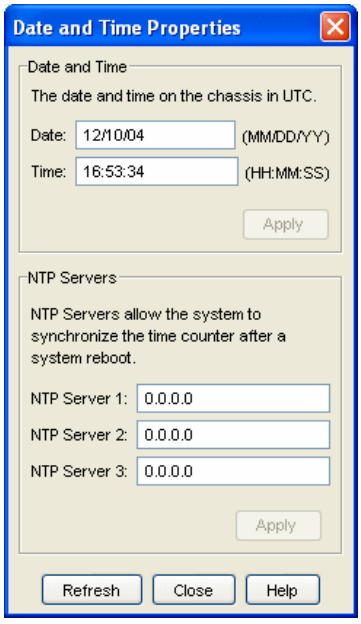

図 6-16　Element Manager: Maintenance: Date and Time Properties

Topspin シャーシにはシステム・クロックが搭載されているので、リブートごとにクロックを再設定する必要はありません。時刻は、時刻値を入力して手動で設定するか、NTP タイム・サーバーを使用して自動的に設定できます。Topspin シャーシは、サーバー・フィールドにリストされているサーバーに対する時刻要求を開始する NTP クライアントの役目をします。初めは、ブート後に Topspin シャーシは iburst キーワードを使用して、シャーシ時刻を設定してから、デフォルトの NTP 間隔で時刻の更新がないかどうかサーバーをポーリン

グします。最大 3 つのサーバーを指定できます。NTP は TCP/IP ネットワーク・アドレスを使用して、アクセスするサーバーを識別します。

「DNS」タブ（図 6-17）は、DNS 名を IP アドレスに解決するために、Topspin コントローラーが使用する DNS サーバーを構成するのに使用されます。

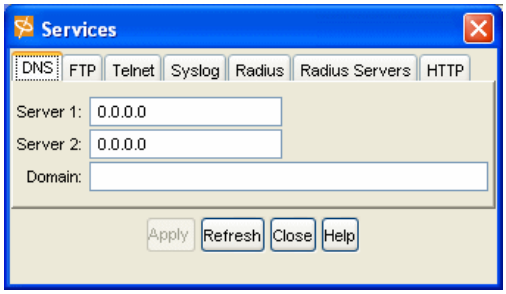

図 6-17 Element Manager: Maintenance: Services - 「DNS」タブ

「FTP」タブ（図 6-18）は、Topspin コントローラーで FTP サーバーを使用可能および使用不可にするのに使用されます。デフォルトでは FTP サーバーは使用不可になっています。

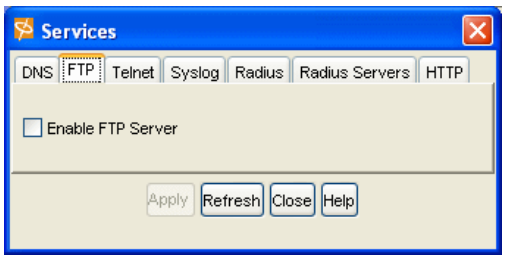

図 6-18 Element Manager: Maintenance: Services - 「FTP」タブ

「Telnet」タブ（図 6-19）は、Topspin コントローラーで Telnet サーバーを使用可能および使用不可にするのに使用されます。スイッチへの非セキュア・アクセスを制限するために、Telnet サーバーを使用不可にすることができます。デフォルトでは Telnet サーバーは使用可能になっています。

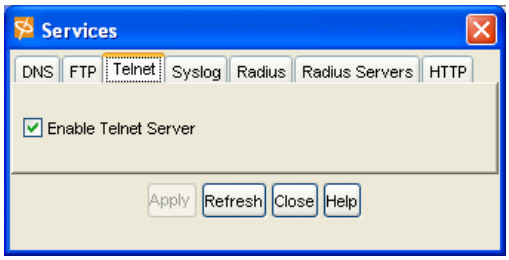

図 6-19 Element Manager: Maintenance: Services - 「Telnet」タブ

「Syslog」タブ（59 ページの図 6-20）は、Topspin シャーシからのログ・メッセージの宛先 Syslog サーバーを指定するのに使用されます。ネットワーク上に Syslog サーバーがある場合、その IP アドレスをこのタブで指定して、Topspin シャーシにそのサーバーにログ・メッ

セージを転送させることができます。外部 Syslog サーバーが指定される場合であっても、Topspin シャーシは、ログ・メッセージのコピーを引き続き保管します。

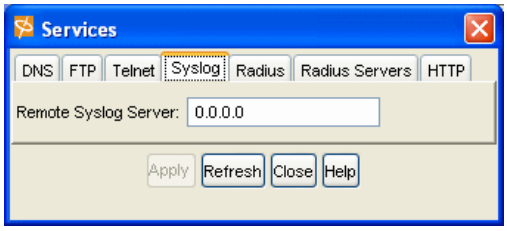

図 6-20　Element Manager: Maintenance: Services -「Syslog」タブ

Topspin シャーシは、ローカル定義ユーザーに対する、および外部 Radius サーバーによる CLI ユーザー認証をサポートします。「Radius」タブ（図 6-21）は、CLI ユーザー・ログオンの認証方法の構成に使用されます。

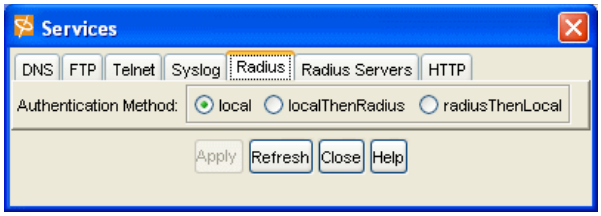

図 6-21　Element Manager: Maintenance: Services -「Radius」タブ

「Radius Servers」タブ（図 6-22）では、シャーシがユーザー・ログオンの認証に使用できる Radius サーバーを指定できます。ログオンの認証に Radius を使用すると、ユーザー定義を集中化することができます。現在、最大 1 つの Radius サーバーを指定できます。

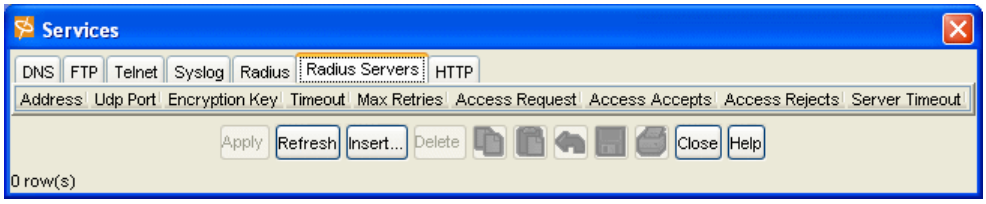

図 6-22　Element Manager: Maintenance: Services -「Radius Servers」タブ

図 6-23 に表示されているウィンドウは、HTTP 用の Web 設定を表示しています。

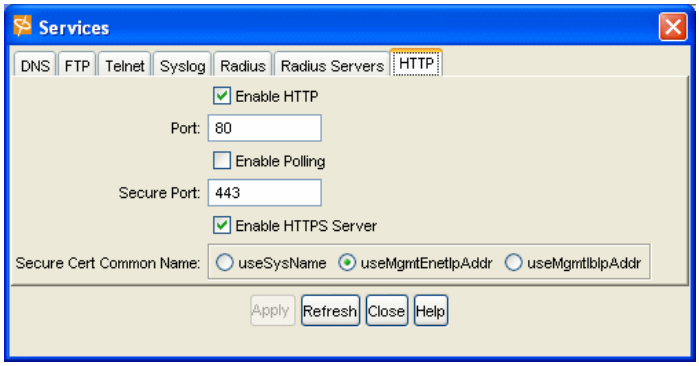

図 6-23　Element Manager: Maintenance: Services -「HTTP」タブ

「Boot Config」メニュー（図 6-24）項目は、Topspin シャーシの初期化と構成に使用されるファイルを選択するためのウィンドウを開きます。「Boot Configuration」ウィンドウには、Topspin シャーシの初期化に使用されるシステム・イメージがリストされます。このウィンドウでは代替システム・イメージを選択できます。また、リブート後にシャーシの構成に使用されるデフォルト構成ファイル startup-config を上書きすることもできます。

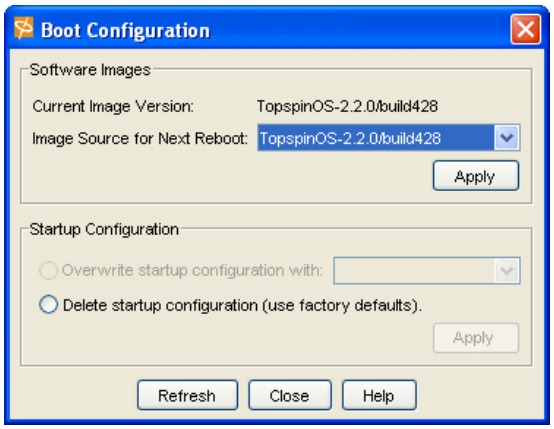

図 6-24　Element Manager: Maintenance: Boot Configuration

「Backup Config」メニュー項目は、現行の Element Manager セッション時に行われた構成を保管するウィンドウを開きます（図 6-25）。実行時に、保管された箇所までの構成を複写する、CLI コマンドの ASCII テキスト・ファイルを作成します。

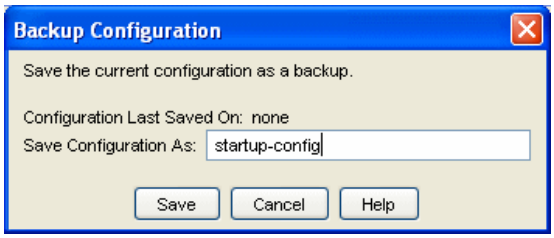

図 6-25　Element Manager: Maintenance: Backup Configuration

システムはリブート時に、config ファイル・システム内の startup-config という名前のファイルを探して使用します。電源投入またはリブート時のシャーシの構成には、常に startup-config が使用されます。このファイルに構成を保管するか、別のファイル名を指定することができます。

「File Management」メニュー項目は、「File Management」ウィンドウを開きます。このウィンドウは、Topspin InfiniBand Switch Module（図 6-26 に表示）または Topspin シャーシ上のイメージ・ファイル、構成ファイル、およびログ・ファイルを表示し、管理します。Topspin シャーシと外部 FTP サーバーとの間で、イメージ・ファイル、構成ファイル、およびログ・ファイルをインポート、インストール、およびエクスポートすることができます。バックアップと表示のために Topspin シャーシからファイル（特にログ・ファイル）をエクスポートします。次回のリブート時に Topspin ファームウェアの更新またはシャーシの構成を行うには、イメージ・ファイルと構成ファイルをシャーシにインポートします。

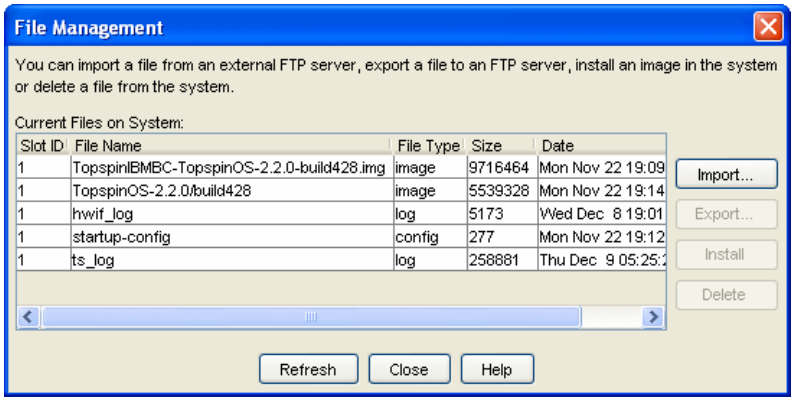

図 6-26 Element Manager: Maintenance: File Management

このウィンドウは、ファイルを内部メモリーから除去するのにも使用されます。コントローラーには最大 2 つのシステム・イメージをインストールし、2 つのイメージ・ファイルを持つことができます。

「Chassis」タブ（図 6-27）は、各種モジュールでシステム全体の診断テストを開始するのに使用されます。

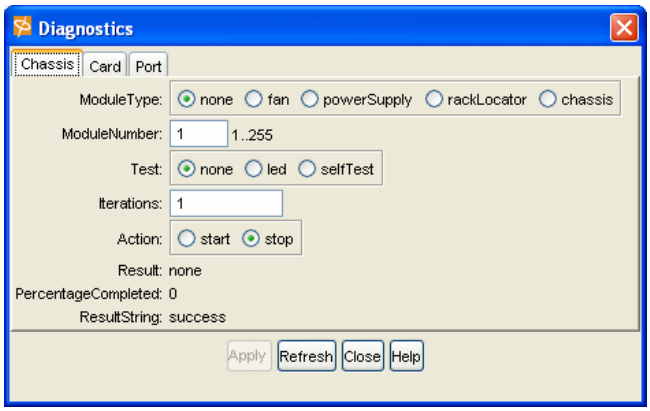

図 6-27 Element Manager: Maintenance: Diagnostics

「Card」タブは、シャーシ内のすべての診断カード・テストのテーブルを表示します。このタブを使用すると、ユーザーは診断テストの作成、削除、および編集を行うことができます。

6.2.4 Element Manager: 「Health」メニュー

「Health」メニュー（図 6-28）は、次の情報を提供します。

- Topspin InfiniBand Switch Module または Topspin シャーシ・モジュールの状況
- シャーシ状態の遷移
- 遷移ログの保管
- トラップ・レシーバー
- ログ・ファイル

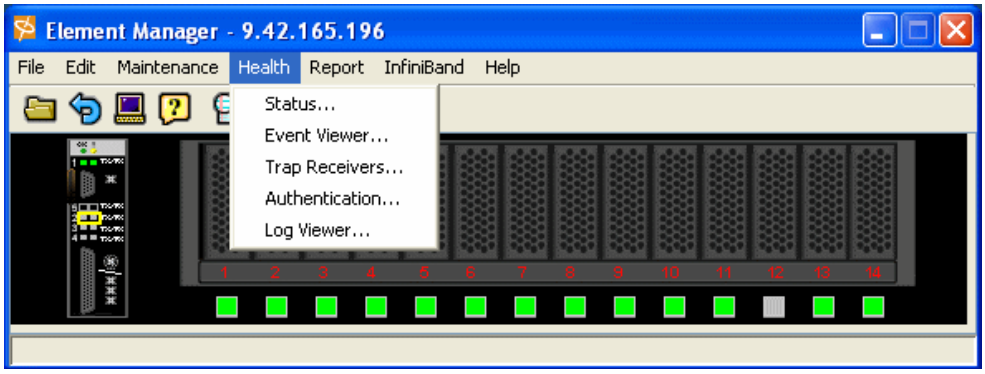

図 6-28 Element Manager: 「Health」メニュー

「Sensors」タブ（図 6-29）は、Topspin シャーシまたは Topspin InfiniBand Switch Module に取り付けられている温度センサーを識別します。また、シャーシ内の位置およびその位置での現行温度も報告されます。ファンおよびコンピューターセンターの空調の冷却効率を確認するために、温度のモニターが必要です。

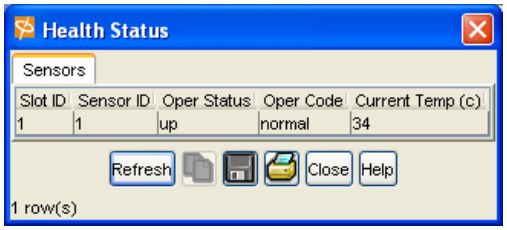

図 6-29 Element Manager: Health: Health Status

「Event Viewer」メニュー項目は、シャーシで発生する重要イベントを識別するウィンドウを開きます（図 6-30）。これらのイベントには、カード・アップ、カード・ダウン、リンクアップ、リンクダウンなどが含まれます。

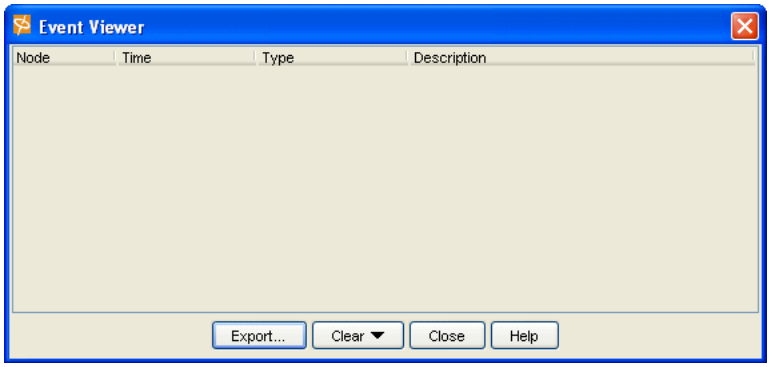

図 6-30 Element Manager: Health: Event Viewer

トラップ・メッセージを受信し、それらを Event Viewer で表示するためのトラップ・レシーバーとして、Element Manager を構成する必要があります。

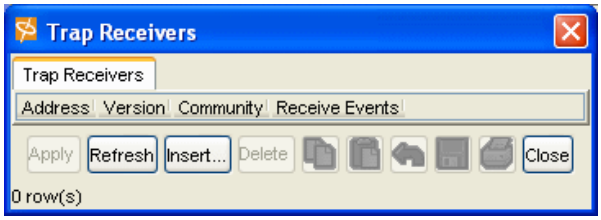

図 6-31　Element Manager: Health: Trap Receivers

図 6-32 では、発生する障害または違反を認証できます。

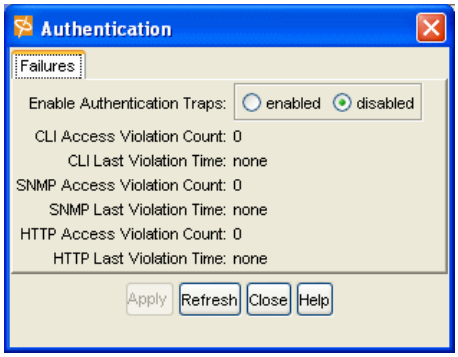

図 6-32　Element Manager: Health: Authentication

「Log Viewer」（図 6-33）は、システムの動作を分析し、構成の変更を監査するための Topspin ログ・ファイルを表示します。zip ログ・ファイルと unzip ログ・ファイルの両方を表示できます。zip ファイルを開くと unzip されます。8 MB のログ・ファイルを表示できます。

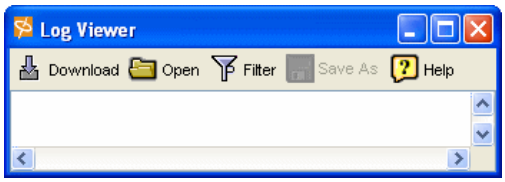

図 6-33　Element Manager: Health: Log Viewer

ログ・ファイルの表示には、通常のテキスト・エディターをダウンロードし、unzip し、使用することができますが、Topspin Log Viewer は、表示されるデータを制御するためのフィルターを備えているので、Topspin Log Viewer の使用をお勧めします。表示されるログ・データは、次のものに基づいてフィルター処理できます。

- 日付の範囲
- スロット番号
- メッセージのタイプと重大度
- メジャー・カテゴリー
- テキスト・ストリング検索
- ログ・データが発生したソフトウェア・モジュール

6.2.5 Element Manager: 「Report」メニュー

「Report」メニュー（図 6-34）は、複数の方法のどちらかでポートおよびイーサネット・カード・データをテキストとグラフィックで表示する手段を提供します。

- テキスト形式。ネットワーク・パフォーマンス統計が表形式で表示されます。
- グラフィック形式。ネットワーク・パフォーマンス統計が図表形式で表示されます。1 組のデータを選択し、棒グラフ、線グラフ、円グラフなどで比較データとして表示できます。

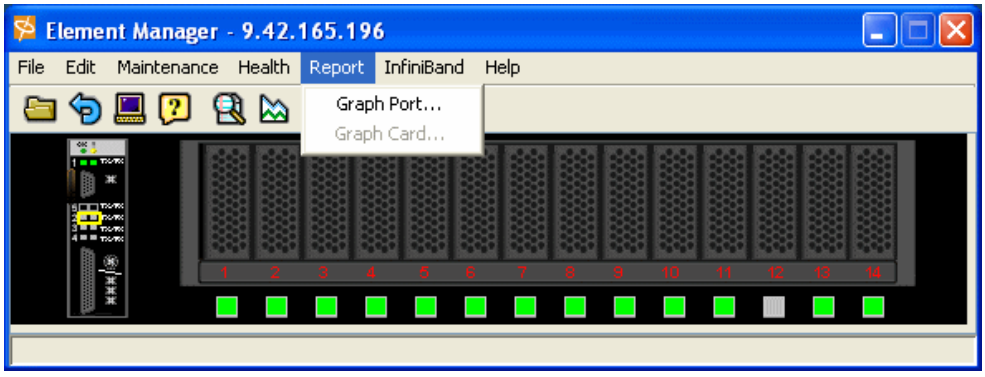

図 6-34 Element Manager: 「Report」メニュー

表示される統計データのフォーマットは、選択されたポート数またはカード数によって決まります。「Graph Port」または「Graph Card」を選択する前に、1 つのポートまたはカードを選択すると、すべてのカウンターが同時に表示されます。あらかじめ複数のポートまたはカードが選択されている場合は、選択されたポートまたはカードのすべてについて 1 つのカウンターのみが表示されます。それ以外のカウンターを表示するには、グラフ・ウィンドウの下部にあるプルダウン・リストから個別に選択してください。

「Graph Port」メニュー項目は、ネットワーク・パフォーマンス統計を表示するウィンドウを開きます（図 6-35）。統計データの選択、および各種フォームでの表示については、「Topspin InfiniBand User Guide」の『Analyzing Network Data』をご覧ください。

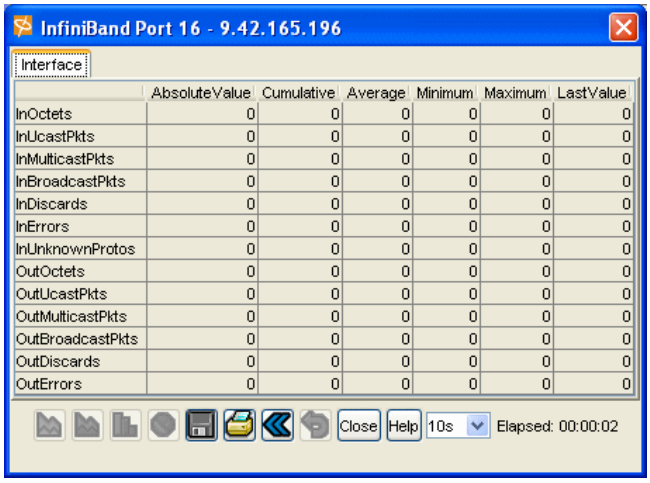

図 6-35 Graph Port

各ウィンドウのタブは、グラフ表示されるインターフェース・ポートのタイプによって決まります。すべてのインターフェース・ポートは、「Interface」タブを表示します。それ以外の

タブは、インターフェース固有です。すなわち、イーサネット・ポートには「Ethernet and IP」タブ、ファイバー・チャネル・ポートには「Fibre Channel」タブです。

6.2.6 Element Manager: 「InfiniBand」メニュー

「InfiniBand」メニュー（図 6-36）は、InfiniBand ファブリックを構成するスイッチ、ルーター、およびチャネル・アダプターを表示し、管理するためのウィンドウを表示します。

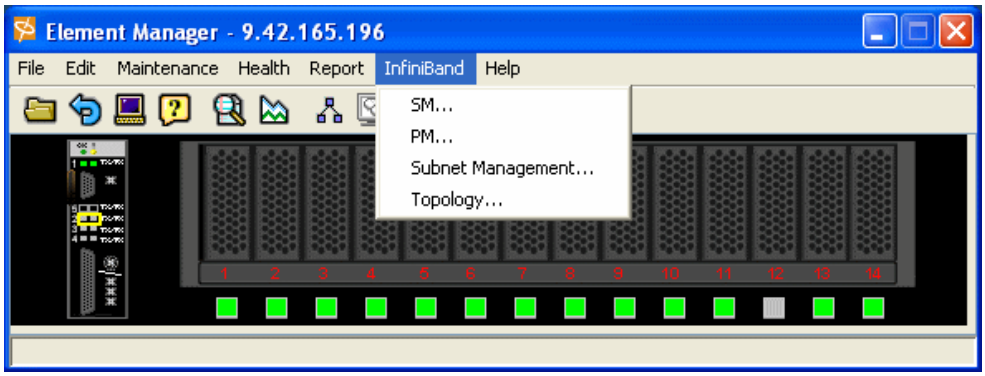

図 6-36 Element Manager: 「InfiniBand」メニュー

Topspin は、Topspin InfiniBand ファブリック上のスイッチ、ルーター、およびチャネル・アダプターを管理するための専有 Subnet Manager を提供します。サード・パーティー製のサブネット・マネージャーを使用できますが、Topspin が提供する Subnet Manager（図 6-37）は、Topspin シャーシおよび Topspin InfiniBand Switch Module の内部アーキテクチャーとよく適合するので、Topspin 提供の Subnet Manager の使用をお勧めします。Topspin Subnet Manager は、厳密に均質な管理を保証します。

Subnet Manager は次のことを行います。

- サブネット・トポロジーを検出し、指定されたスイープ間隔で動的に更新する。
- チャネル・アダプター・ポートごとにローカル ID（LID）、グループ ID（GID）サブネット接頭部、およびパーティション・キー（P_Keys）を構成する。
- サブネット上のスイッチごとに LID、サブネット接頭部、および転送データベースを構成する。
- サブネットのエンド・ノードおよびサービス・データベースを保守して、サービス・ディレクトリーと共に GUID to LID/GID 解決サービスを提供する。

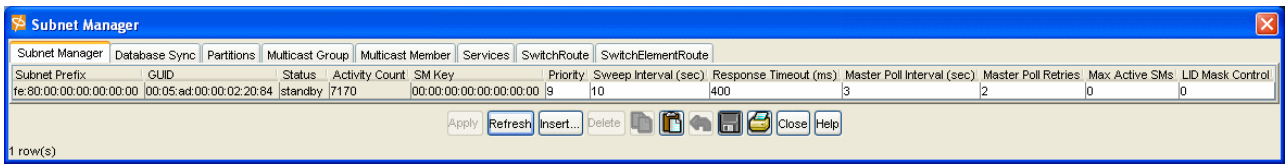

図 6-37 Element Manager: InfiniBand: Subnet Manager

「Subnet Management」ウィンドウ（図 6-38）は、サブネット管理に関連した設定値を構成するための使いやすいインターフェースを提供します。左側にナビゲーション・ツリーがあり、右側にコンテンツ・ペインがあります。

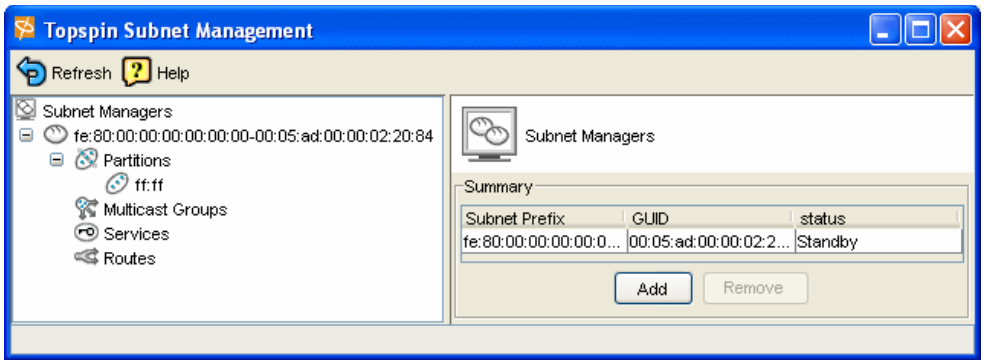

図 6-38　Element Manager: Report: Subnet Management

コンテンツ・ペインに表示される情報は、ナビゲーション・ツリーの選択内容によって決まります。例えば、パーティションを管理するには、「**Partitions**」ツリー・ノードを展開して、パーティションを表示してから、特定のパーティションを選択します。右側のペインに、選択されたパーティションの構成が表示されます。

Element Manager によって管理される Topspin 装置（図 6-39）またはスイッチ・モジュールを表示するには、メニューから「**InfiniBand**」→「**SM**」を選択します。

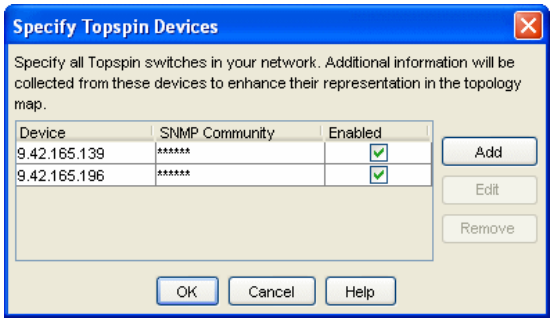

図 6-39　Specify Topspin Devices

Performance Manager 内のアクティビティーを表示するには、メニューから「**InfiniBand**」→「**PM**」を選択します（図 6-40）。

図 6-40　Element Manager: InfiniBand: Performance Manager

メニューから「**InfiniBand**」→「**Topology**」を選択すると、外部装置である Topspin 360 スイッチ・モジュールと通信する Topspin InfiniBand Switch Module に割り当てられている BladeCenter サーバーと HCA のグローバル・ビュー（図 6-41）が表示されます。

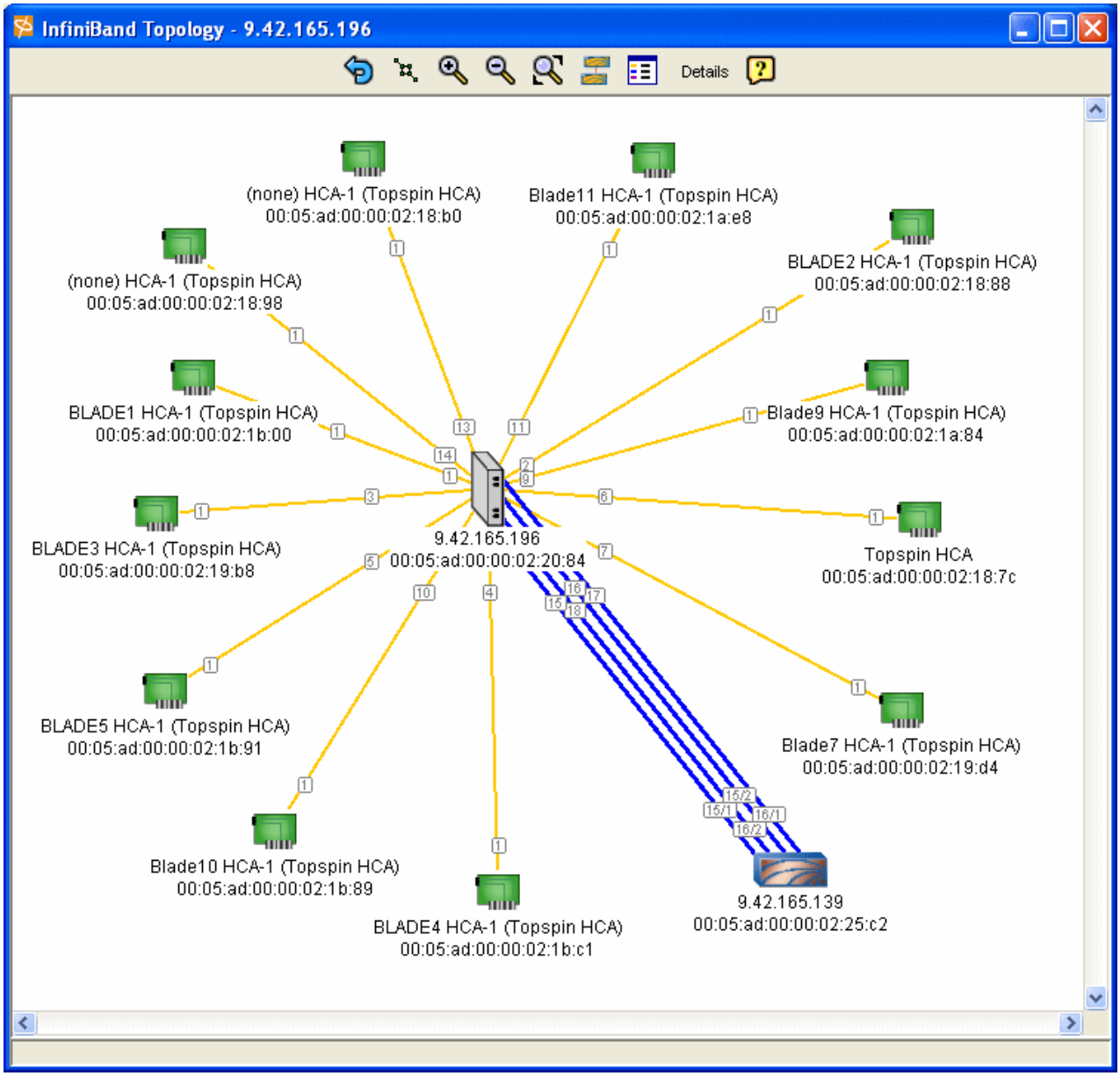

図 6-41　Element Manager: Report: InfiniBand Topology

6.2.7 Element Manager: 「Help」メニュー

「Help」メニュー（図 6-42）では、Element Manager の使用に関連した資料およびオンライン・サポート情報を表示できます。

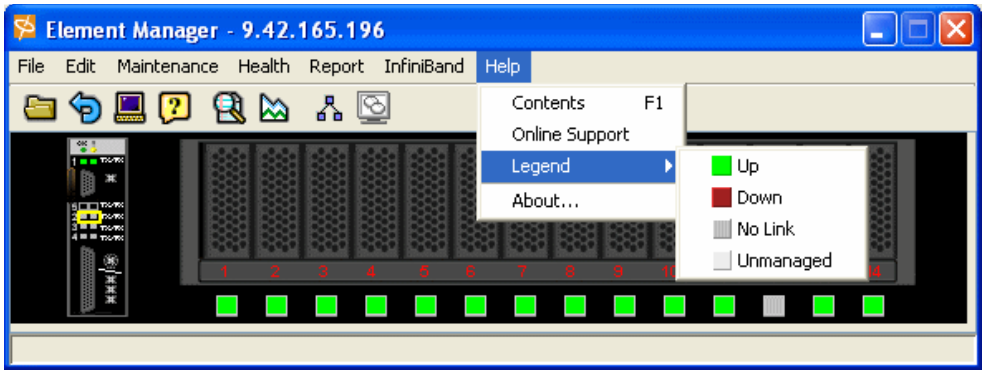

図 6-42 Element Manager - Help

6.3 Element Manager: Topspin 360 ビュー

Element Manager から Topspin 360 を表示すると、接続されている InfiniBand システムの現行構成が表示されます。

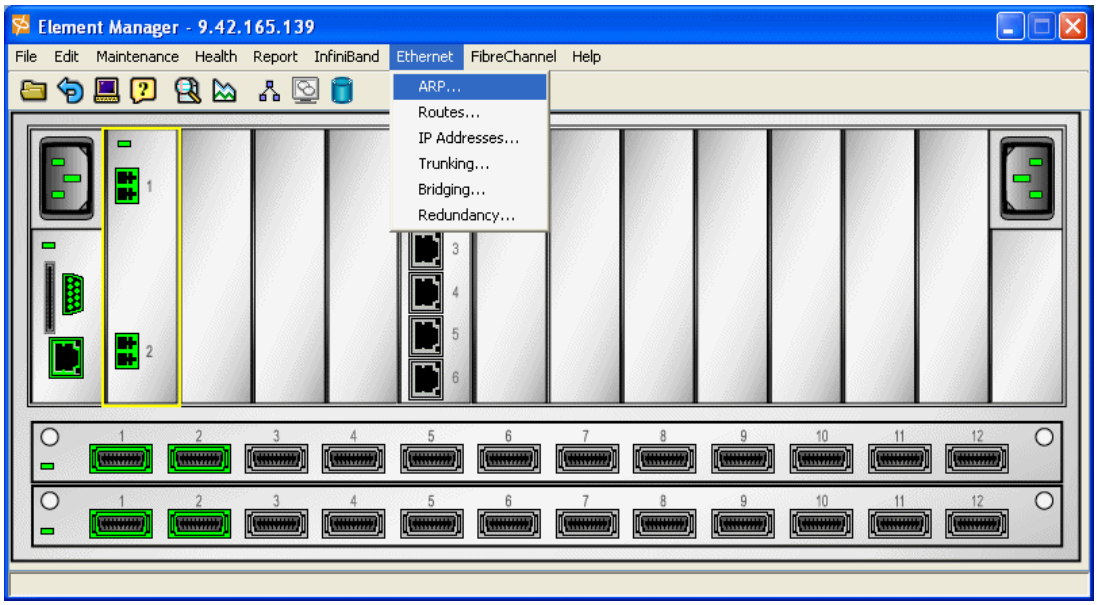

図 6-43 Element Manager: Ethernet

Element Manager は、構成の変更を表示するために動的に更新されます。カードおよびポートが構成されると、対応する実行ライトとポート・フレームが、緑色に変わって変更を表します。Element Manager の「Preference」の設定に応じて、構成変更が Element Manager の画面に表示されるのに数秒かかる場合があります。図 6-43 と 69 ページの図 6-44 の表示が、Topspin InfiniBand Switch Module の実際の表示（図 6-42）と異なる点は、イーサネット・スイッチ・ポートとファイバー・チャネル・ポートを挿入する拡張スロットの位置です。

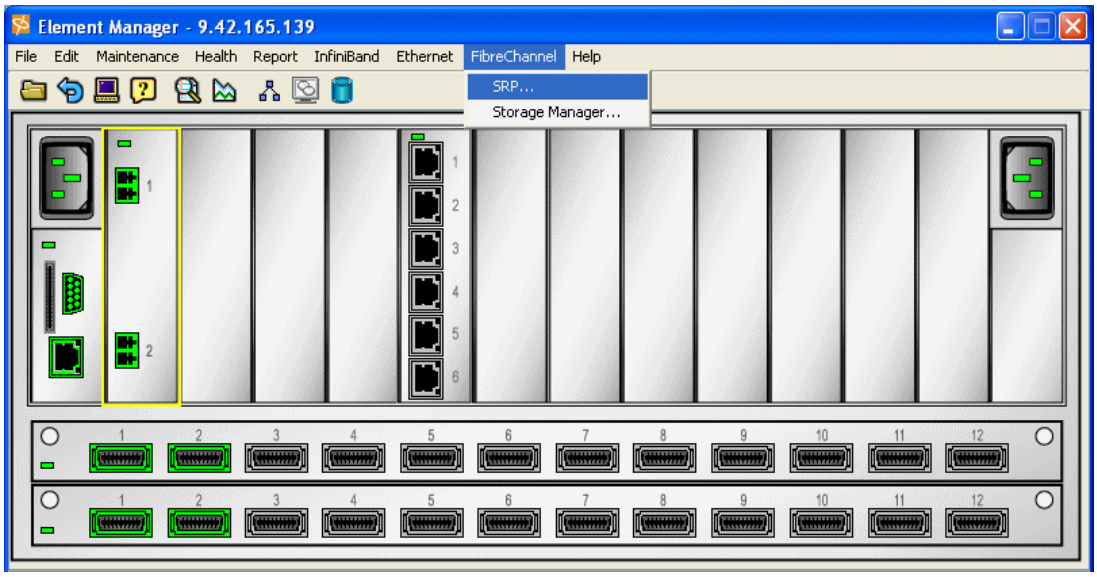

図 6-44　Element Manager: Fibre Channel

この節では、FibreChannel → Storage Manager を表示する図 6-44 に焦点を当てます。このファイバー・チャネル構成では、フル装備の IBM TotalStorage DS4500 を使用し、Element Manager は、管理される DS4500 のドライブの内容（図 6-45）を表示します。

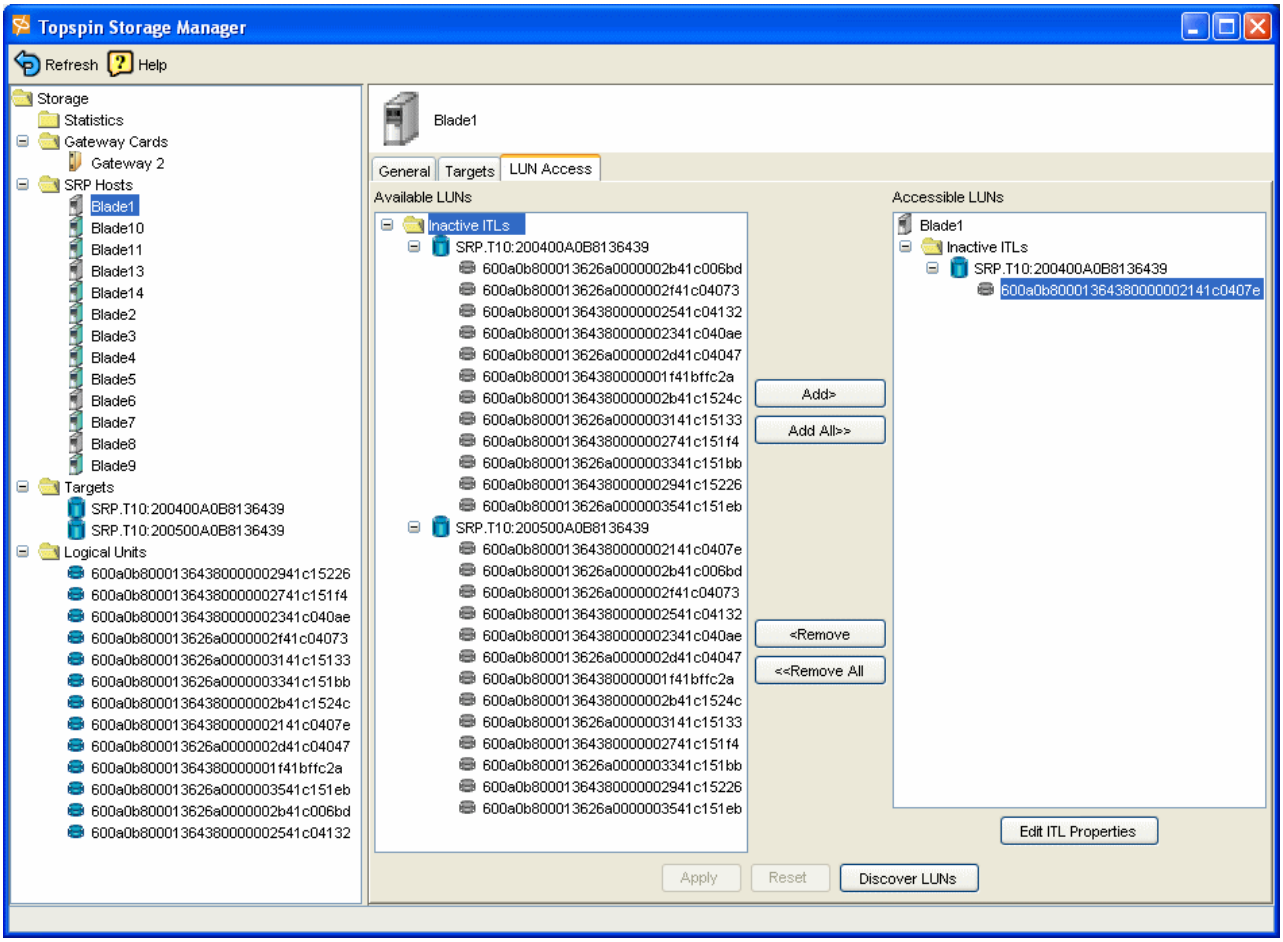

図 6-45　Topspin 360 FC Storage Manager

6.4 Chassis Manager

Topspin Chassis Manager（CM）は、サーバー・スイッチで直接実行され、各種管理タスクを素早くかつ容易に実行するのに役立ちます。この章では、インターフェースの各種コンポーネントについて説明しています。Chassis Manager（図 6-46）は、すべての Topspin サーバー・スイッチで実行されます。

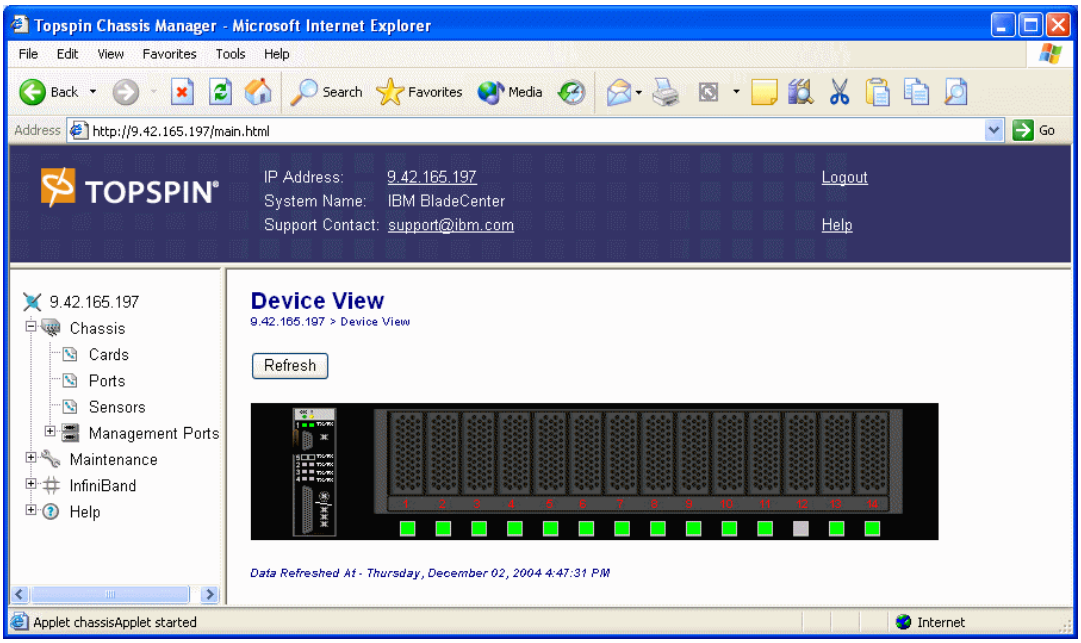

図 6-46　Topspin InfiniBand Switch Module ビュー

図 6-46 では、Chassis Manager の Topspin InfiniBand Switch Module の「Device View」と左側のパネルが、Element Manager のメニュー・バーと Topspin InfiniBand Switch Module ビュー（図 6-47）とほぼ同じであることに注目してください。Element Manager は、Topspin 装置用の管理ソフトウェアとしてお勧めしますが、Chassis Manager を使用して Topspin InfiniBand Switch Module の必要な構成アクティビティーを実現することもできます。

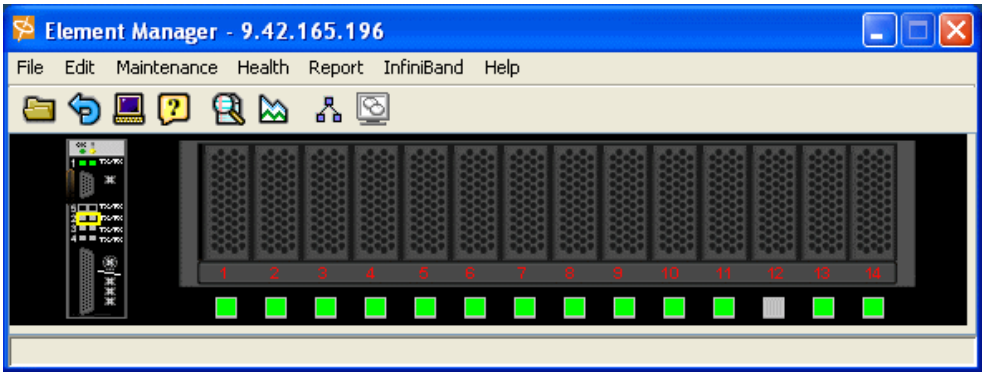

図 6-47　Element Manager スイッチ・モジュールとビュー

以下の図では、左側のパネルの項目と実現可能なタスクについて説明します。

サーバー・スイッチ内のハードウェアを表示し、構成するには、「Chassis」アイコン（図 6-48）をクリックします。このアイコンにアクセスすると、装置上のすべての FRU（技術員により交換される部品）の状況を表示できます。

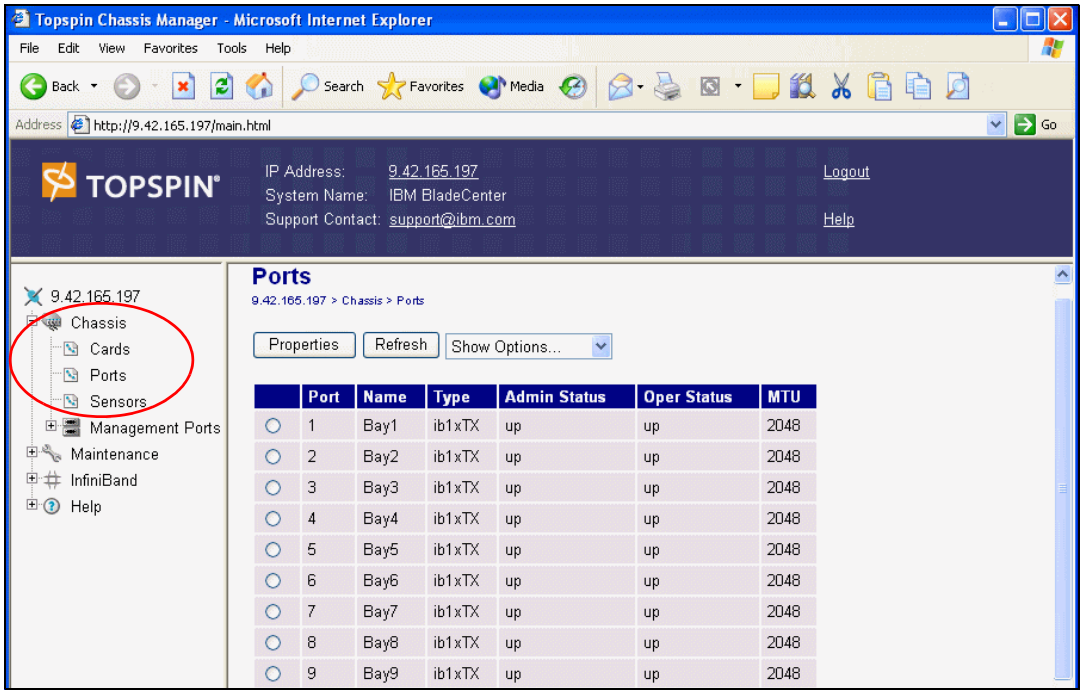

図 6-48　Chassis → Ports

「Maintenance」アイコン（図 6-49）には、サーバー・スイッチで基本的な管理タスクを実行するための分岐が含まれています。このアイコンにアクセスすると、NTP サーバーの構成、boot-config ファイルの割り当て、ファイル・システムの内容の表示などを行うことができます。

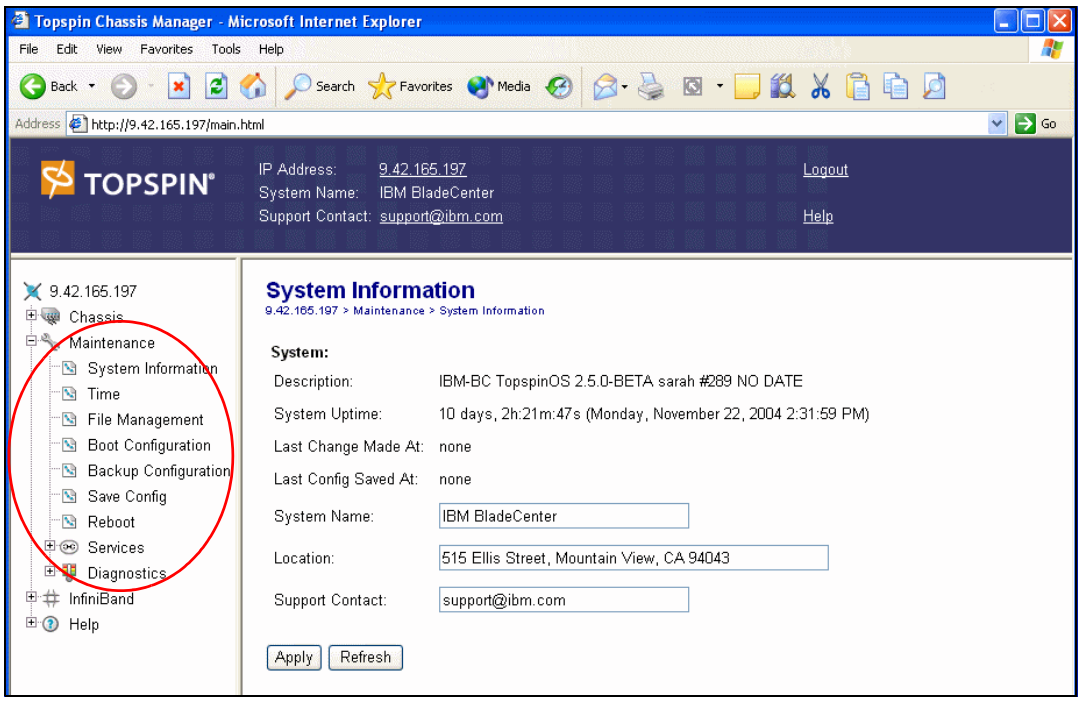

図 6-49　Maintenance → System Information

「InfiniBand」アイコン（図 6-50）は、Subnet Manager と I/O の詳細を表示します。基本的な SM プロパティーを構成するには、このアイコンの「**Subnet Managers**」分岐をクリックしてください。

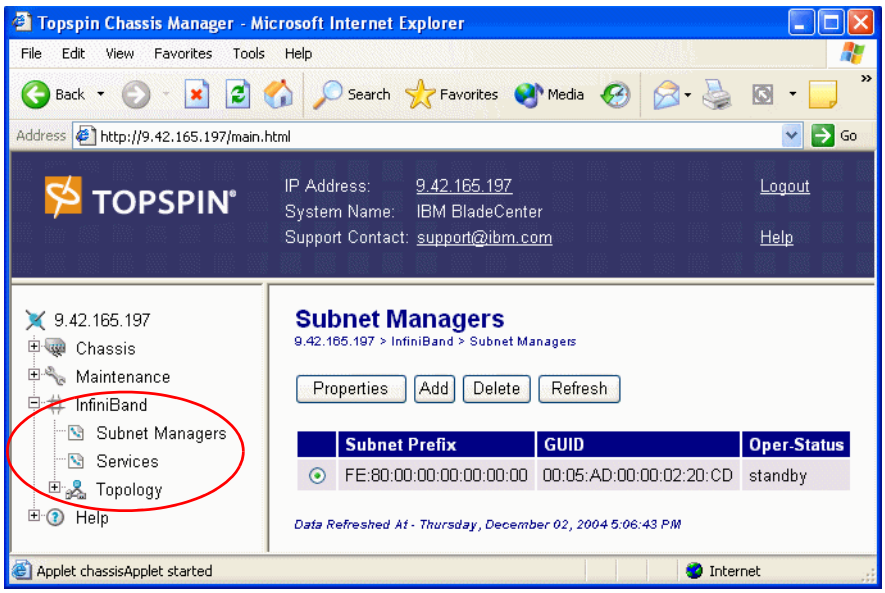

図 6-50　Chassis Manager: - InfiniBand → Subnet Managers

「Ethernet」アイコン（Topspin 360 などのハードウェア・プラットフォームの選択のみで使用可能）を使用すると、サーバー・スイッチ上の IP トラフィックの多くの特徴を表示し、構成することができます。

「Fibre Channel」アイコン（Topspin 360 などのハードウェア・プラットフォームの選択のみで使用可能）は、SRP ホストと FC ストレージの詳細を表示し、グローバル・ポリシーの構成を可能にします。

「Help」アイコンを使用すると、オンライン・ヘルプとサポート・リソースにアクセスできます。

7

IBM eServer BladeCenter システムの初期セットアップと構成

この章では、IBM @server BladeCenter 用の Topspin InfiniBand Switch Module の導入時に役立つことを目的に、今回構成したテスト環境のネットワーク構成、ハードウェアの設定について説明します。

7.1 IBM eServer BladeCenter システム

ここでは、BladeCenter のセットアップについて説明します。

7.1.1 管理モジュールのファームウェア

必要なハードウェアが BladeCenter に取り付けられた後、BladeCenter - Management Module Firmware Update Version 1.16 またはそれ以降を使用して、管理モジュールを更新する必要があります。旧バージョンの管理モジュール・ファームウェアは、InfiniBand ハードウェアを認識せず、電源を投入しません。ファームウェアを取得するには、次の Web サイトにアクセスしてください。

http://www-1.ibm.com/servers/eserver/support/bladecenter/chassis/downloadinghwonly.html

Readme ファイルのインストールとセットアップの説明に従って導入を行ってください。拡張子が .pkt というファイルをインストールしてください。（図 7-1）インストール後、管理モジュールの再始動が必要です。

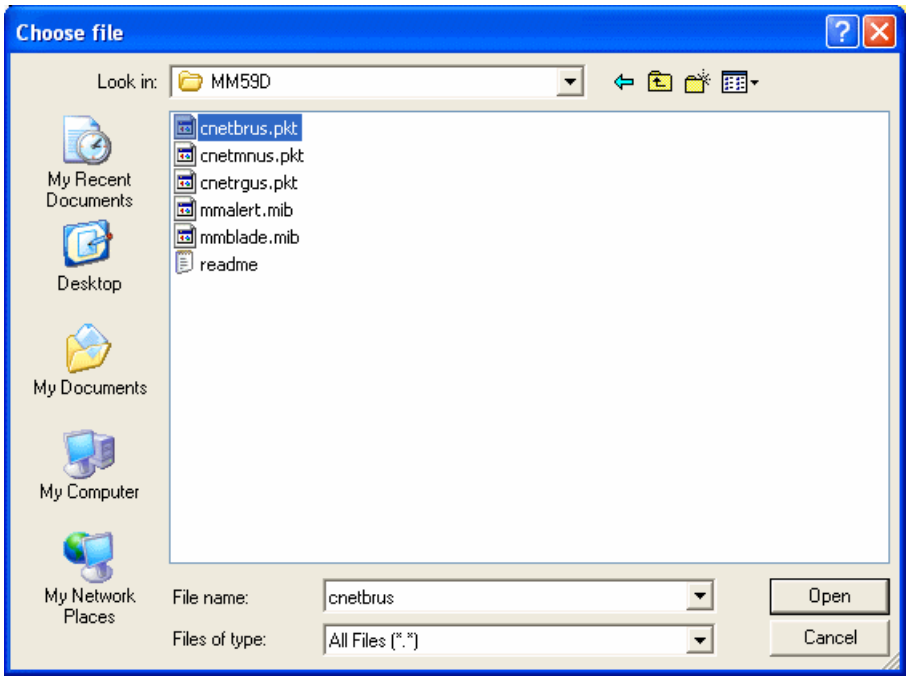

図 7-1　管理モジュール・ファームウェア更新ファイル

7.1.2 管理モジュールのネットワーク・インターフェース

ここでは、管理サブネット上に存在する管理モジュールの外部および内部ネットワーク・インターフェースを構成します。外部ネットワーク・インターフェース IP アドレスは、BladeCenter 外部のネットワークに接続されます。このアドレスは、外部装置から管理モジュールと通信するのに使用されます。

管理モジュールとの物理接続の確立

管理モジュールの管理方法は、モジュールの前面にある外部 10/100 Mbps イーサネット・ポートを使用する方法です。管理モジュールとの物理接続を確立するには、次の方法のどちらかを使用します。

- カテゴリー 3、4、5、またはそれ以上の対より線（シールドなし）（UTP）ストレート・ケーブルを使用して、管理モジュールのイーサネット・ポートを、アクセス可能な管理ステーションを持つネットワーク内のスイッチに接続します。
- カテゴリー 3、4、5、またはそれ以上のクロスオーバー・ケーブルを使用して、管理ステーション（PC やラップトップなど）を管理モジュールの外部イーサネット・ポートに直接接続します。

管理モジュール Web インターフェースへのアクセス

管理モジュールとの物理接続を確立した後、管理モジュールと同じサブネット内の使用可能な IP アドレスを持つ管理ステーションを構成します。デフォルトでは、サブネットは 192.168.70.0/24 です。

管理モジュールの基本的な管理方法には次の 2 通りの方法があります。

- HTTP Web インターフェース
- IBM Director

ここでは、管理モジュール Web インターフェースを使用して、管理モジュールの初期構成とスイッチ・モジュール構成を示します。

管理モジュールとの管理セッションを確立し、初期スイッチ・モジュール設定を構成する手順は、次のとおりです。

1. Web ブラウザーを開き、構成済みの IP アドレスを使用して管理モジュールに接続する。管理モジュールのデフォルトの動作は、DHCP プロトコルを使用して IP アドレスの取得を試みることです。2 分しても IP アドレスを受信しない場合、管理モジュールは、デフォルトの IP アドレス 192.168.70.125 を外部インターフェースに適用します。内部インターフェースのデフォルト IP アドレスは 192.168.70.126 であることに注意してください。
2. ユーザー ID とパスワードを入力する。デフォルトは `USERID` と `PASSWORD` です（大文字小文字の区別があり、文字 0 ではなく数字のゼロです）。「**OK**」をクリックします。
3. 初期ウィンドウで、「**Continue**」をクリックして管理セッションにアクセスする。

また、BladeCenter documentation CD に収録されている「*BladeCenter Management Module User's Guide*」も参照できます。

管理モジュール・ネットワーク・インターフェースの構成

管理モジュール Web インターフェースにアクセスした後、外部および内部のネットワーク・インターフェースを構成できます。BladeCenter 管理モジュールの Web インターフェースから、「**MMControl**」→「**Network Interfaces**」をクリックします。

BladeCenter 管理モジュールは、デフォルトで IP アドレス 192.168.70.125 に設定されます。管理ネットワーク上に複数の BladeCenter がある場合は、外部ネットワーク・インターフェース（eth0）を変更する必要があります。変更しないと、IP アドレスの競合により、管理モジュールにアクセスできなくなります。図 7-2 では、固定 IP アドレスを使用して同じデフォルト管理サブネット上にある外部インターフェースを構成しました。

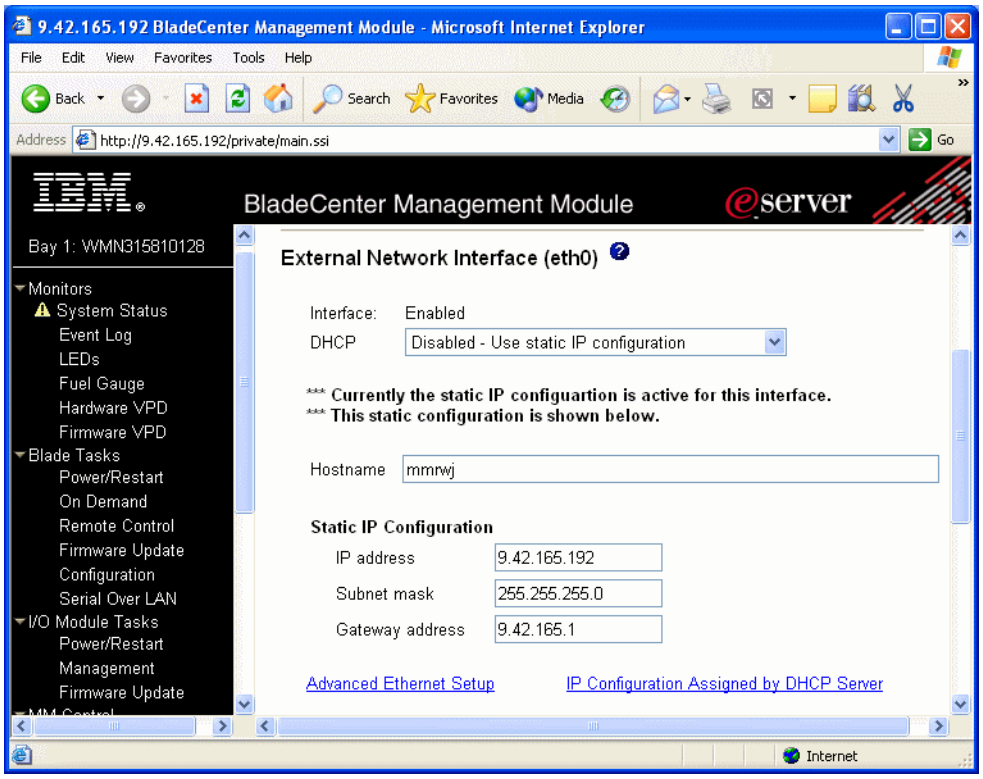

図 7-2　Management Module External Network Interface ウィンドウ

外部インターフェースが構成された後、別の固定 IP アドレスを使用して内部インターフェース（図 7-3）を構成する必要があります。内部ネットワーク・インターフェース（eth1）の目的は、イーサネット・リンク上で BladeCenter 装置と通信することです。外部インターフェースと同じネットワーク上に内部インターフェースを構成しない場合、管理モジュールとスイッチ・モジュールとの間で接続できないことに注意してください。

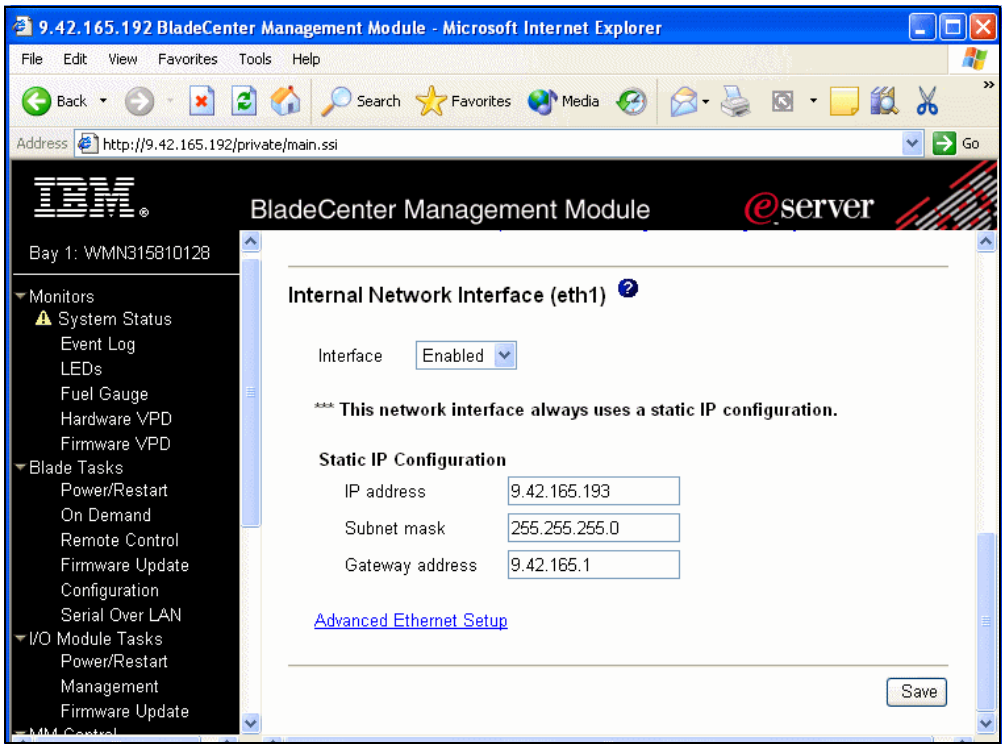

図 7-3　Management Module Internal Network Interface ウィンドウ

「**Save**」をクリックします。変更を反映するには、管理モジュールの再始動が必要です。

7.1.3 I/O モジュール管理タスク

管理モジュールでは、Topspin InfiniBand Switch Module を表示し、管理できます。しかし、84 ページの 7.4、『1 つまたは 2 つの InfiniBand スイッチを備えたシャーシの構成』および 87 ページの 7.5、『InfiniBand スイッチ上のファームウェアの更新』の手順を使用して、ファームウェアをロードすると共に、スイッチを管理することをお勧めします。

7.2 Topspin HCA とスイッチ・モジュールの取り付けおよび構成

以下の節では、BladeCenter HS20 の作動の準備を行います。

1. InfiniBand ドーター・カードをブレードに取り付ける。
2. InfiniBand スイッチをシャーシに取り付ける。
3. 各 Topspin InfiniBand Switch Module から外部 Topspin サーバー・スイッチまでを InfiniBand ケーブルで接続する。
4. Topspin ファイバー・チャネル・ゲートウェイから DS4000 までをファイバー・チャネル・ケーブルで接続する。
5. Topspin 360、BladeCenter シャーシ、および DS4000 の電源を投入する。

6. InfiniBand ソフトウェアをブレードにインストールする。（本書の作成時点で、HCA は Version 3.02.0000 Build 3.0.0-126 でした。Topspin 360 は 2.1.0-Build170 でした。）

7.3 ブレードへの InfiniBand HCA の取り付け

InfiniBand ファブリックの一部として機能するために、IBM ブレード・サーバーには Topspin InfiniBand HCA およびドライバーがインストールされていなければなりません。

ブレード・サーバーに I/O 拡張カードを取り付ける手順は、次のとおりです。

1. BladeCenter 格納装置にブレード・サーバーが取り付けられ、稼働している場合は、オペレーティング・システムをシャットダウンしてから、ブレード・サーバーの電源をオフにする。（カードの取り付け方法については、ご使用のブレード・サーバーのインストールとユーザーの手引きを参照してください。）
2. ブレード・サーバーを BladeCenter 格納装置から取り外し、ブレード・サーバー・カバーを開いて、I/O 拡張カードを取り付けるコネクターを見つける。

重要 :

- I/O 拡張カードを取り付けるために IDE ハード・ディスクを取り外す必要があるときに、このディスク・ドライブに保存したい情報が入っている場合は、その情報を別のストレージ・デバイスにバックアップしてください。
- I/O 拡張カードを取り付けるために IDE ハード・ディスクを取り外す必要があるときに、このディスク・ドライブが RAID アレイに含まれている場合、この RAID アレイの構成を解除してから、ハード・ディスクを取り外してください。その方法については、ご使用のオペレーティング・システムの資料を参照してください。

3. I/O 拡張カードを取り付ける IDE コネクター位置に IDE ハード・ディスクが取り付けられている場合は、そのドライブ、ライザー・カード、およびトレイを取り外す（トレイとシステム・ボードを固定しているねじを保管しておく）。取り付けられていない場合は、IDE コネクターの近くに 2 本のねじがあれば、それらのねじを取り外します。
4. I/O 拡張カードに付属の I/O 拡張トレイを取り付ける。オプション・キットからのねじを使用して、トレイをブレード・サーバーに固定します。
5. 帯電防止パッケージから I/O 拡張カードを取り出す。
6. I/O 拡張カード・コネクター J2 から保護カバーが取り外されていることを確認してから、I/O 拡張カードの狭い方の端を、I/O 拡張トレイの飛び出ているフックに差し込む。
7. I/O 拡張カードの I/O 拡張カード・コネクター（J2 および J3）を、ブレード・サーバー上の I/O 拡張オプション・コネクターの位置と合わせてから、カードをコネクターにそっと押し込む。
8. このブレード・サーバーに他のオプションを取り付ける場合は、すぐに取り付ける。他のオプションがない場合は、ブレード・サーバー・カバーを閉じ、BladeCenter 格納装置に取り付けます。

7.3.1 Windows を実行するシステム用の HCA ファームウェアの更新

Windows を使用して InfiniBand HCA をインストールする手順は、次のとおりです。

1. 「スタート」メニューから、「Topspin utilities」ウィンドウを開く。「**スタート**」→「**プログラム**」→「**Topspin InfiniBand SDK**」→「**Utlities**」をクリックします。

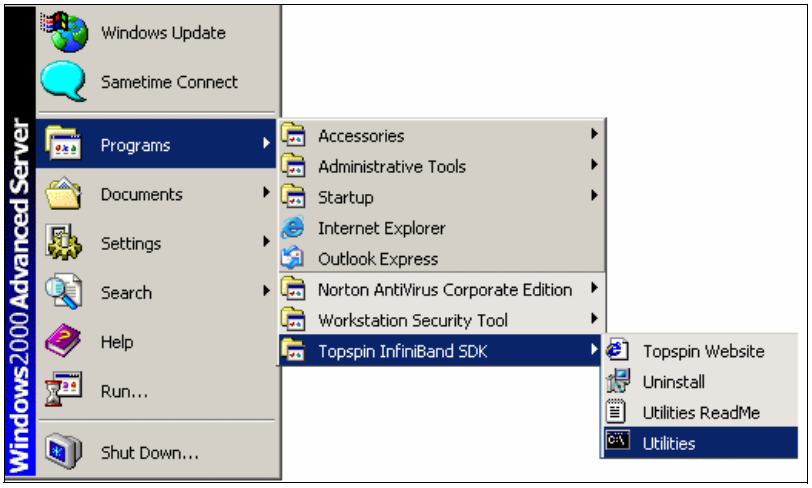

図 7-4　始動

図 7-5 のようなコマンド・プロンプトが表示されます。

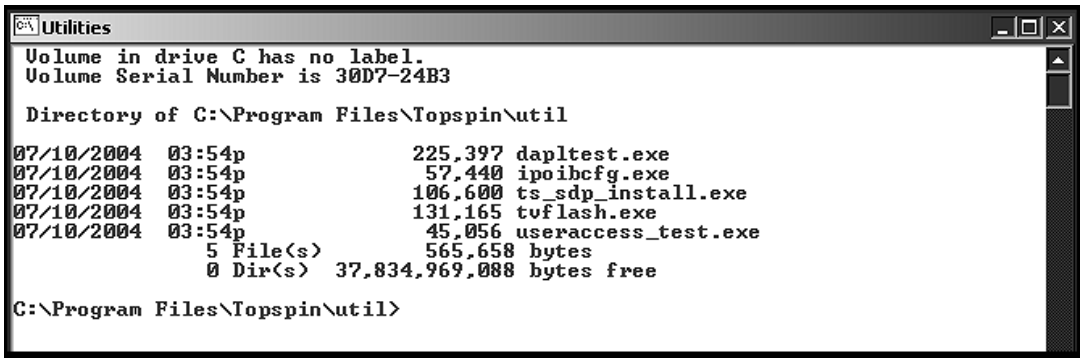

図 7-5　ファームウェア・ユーティリティー

> **注 :**Windows コマンド・プロンプトを開き、tvflash が入っているディレクトリー（デフォルトでは C:¥Program Files¥Topspin¥util）にディレクトリーを変更することもできます。

2. コマンド・プロンプトで **tvflash -i** を入力して、ドーター・カード上のファームウェアのレベルを確認する。

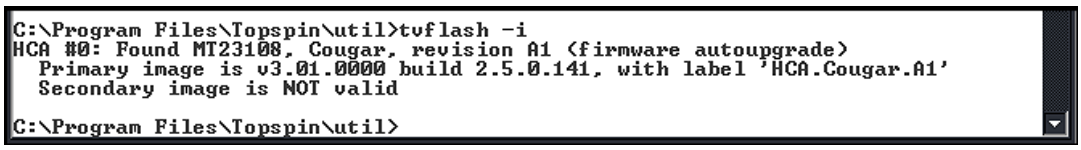

図 7-6　ファームウェア・レベルの確認

このコマンドは、システム上の各 HCA のフォーム・ファクターおよび 1 次と 2 次ファームウェア・イメージを示します。BladeCenter HCA フォーム・ファクターは cougar とついているものです。ファームウェアを更新する際は、正しいフォーム・ファクターのファームウェアであるかどうかを確認してください。

3. **tvflash** コマンドの後に、ファームウェア名を続けて入力して（図 7-7）、Enter を押す。

```
C:\Program Files\Topspin\util>tvflash ..\tavorFW\fw-cougar-a1-3.2.0.bin
New Node  GUID = 0005ad0000020da8
New Port1 GUID = 0005ad0000020da9
New Port2 GUID = 0005ad0000020daa
Programming Tavor Microcode... Flash Image Size = 342944
........................................
Erasing all sectors.....
..
Writing the failsafe image....
--------------------------------------------------------------------------------
................
Verifying the image.....
..
Finishing the failsafe image......
Flash verify passed!

C:\Program Files\Topspin\util>_
```

図 7-7　フラッシュ・プロセス

4. デスクトップから、Web ブラウザーで「BladeCenter Management Module」アクセス・ウィンドウを開く。ビューが開いて「System Status」を表示します（図 7-8）。このビューで、ブレードに InfiniBand ドーター・カードが取り付けられているかどうかを確認することができます。

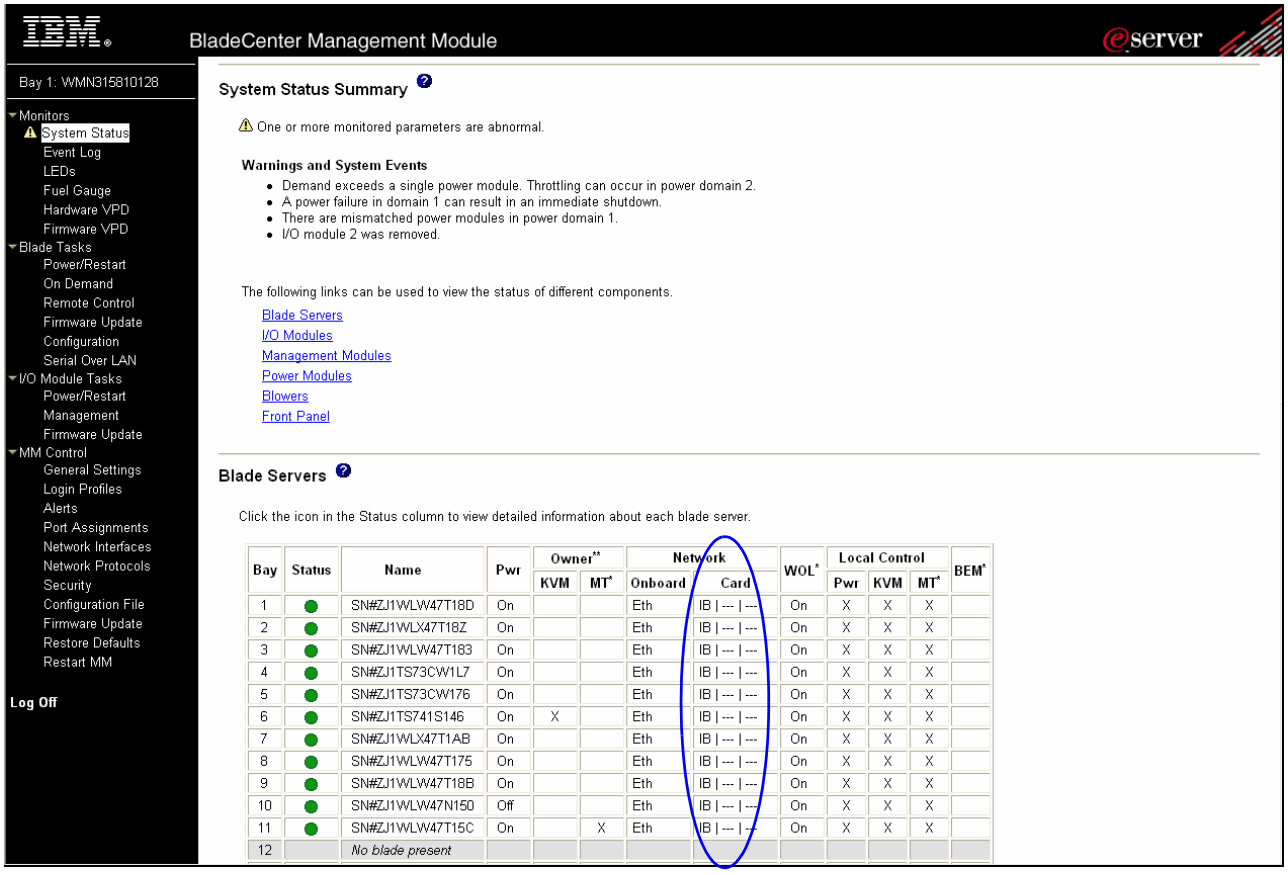

図 7-8　IBM eServer BladeCenter 管理モジュールのシステム状況の要約

7.3.2 Linux を実行するシステム用の HCA ファームウェアの更新

Linux を使用して InfiniBand HCA をインストールする手順は、次のとおりです。

1. カードにインストールされているファームウェアのレベルを確認する。Linux 端末で、**tvflash -i** コマンドの前に、tvflash 実行可能ファイルの絶対パスを付けて実行します。ドライバーがインストールされている場合、/usr/local/topspin/sbin/tvflash にコピーがあります。

```
[root@vframe-director /]# /usr/local/topspin/sbin/tvflash -i
HCA #0: Found MT23108, Cougar, revision A1 (firmware autoupgrade)
  Primary image is v3.02.0000 build 2.5.0.279, with label 'HCA.Cougar.A1.Boot'
  Secondary image is v3.02.0000 build 2.5.0.277, with label 'HCA.Cougar.A1.Boot'
[root@vframe-director /]#
```

図 7-9　ファームウェア・レベルの確認

このコマンドは、システム上の各 HCA のフォーム・ファクターおよび 1 次と 2 次ファームウェア・イメージを表示します。BladeCenter HCA フォーム・ファクターは、cougar で識別されます。ファームウェアを更新する際は、正しいフォーム・ファクターのファームウェアであるかどうかを確認してください。

2. 次のコマンドを入力し、Enter を押す。

```
{path for tvflash}/tvflash {firmware path}/{firmware name}
```

```
[root@vframe-director /]# /usr/local/topspin/sbin/tvflash /usr/local/topspin/share/fw-cougar-a1-3.2.0.bin
New Node  GUID = 0005ad0000020ef0
New Port1 GUID = 0005ad0000020ef1
New Port2 GUID = 0005ad0000020ef2
Programming Tavor Microcode... Flash Image Size = 342944
Failsafe   [==========================================================================]
Erasing    [==========================================================================]
Writing    [==========================================================================]
Verifying  [==========================================================================]
Flash verify passed!
[root@vframe-director /]#
```

図 7-10　フラッシュ・プロセス

3. デスクトップから、Web ブラウザーで「BladeCenter Management Module」アクセス・ウィンドウを開く。ビューが開いて「System Status」が表示されます（82 ページの図 7-8）。このビューで、ブレードに InfiniBand ドーター・カードが取り付けられているかどうかを確認することができます。

7.4 1 つまたは 2 つの InfiniBand スイッチを備えたシャーシの構成

ブレード・サーバーを外部 InfiniBand ファブリックに接続するには、1 つまたは 2 つの InfiniBand スイッチを備えた BladeCenter シャーシを構成します。

1. BladeCenter シャーシに 1 つまたは 2 つのスイッチを取り付ける。
2. Web ブラウザーで管理モジュール・アクセス・ウィンドウを開く。ビューが開いて「System Status」が表示されます。このビューで、InfiniBand スイッチがインストールされていることを確認できます（図 7-11）。

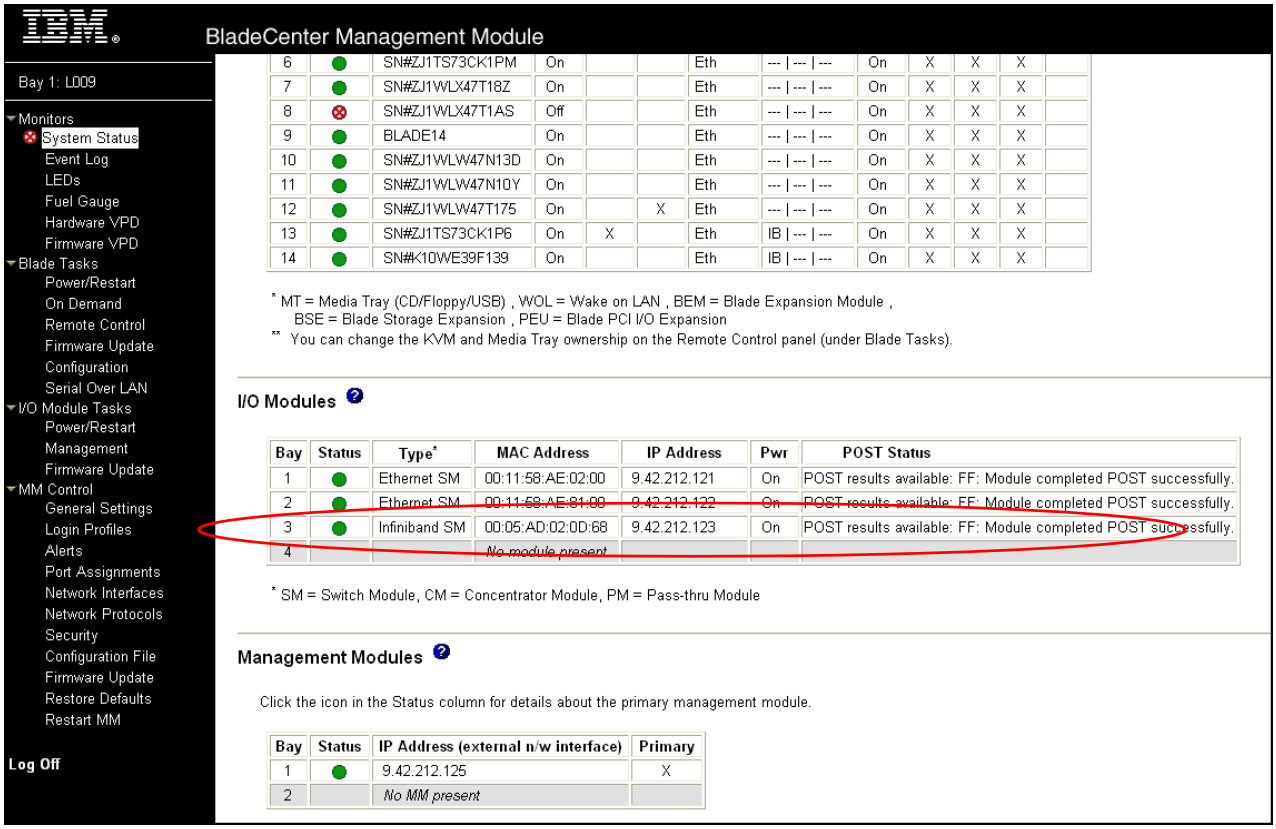

図 7-11　BladeCenter Management Module: I/O Modules

3. 管理モジュール・アクセス・ウィンドウの左側のナビゲーション・バーで、「**Firmware VPD**」をクリックする。「**I/O Module Firmware VPD**」までスクロールダウンして、スイッチ・ファームウェアの現行レベルを確認します（図 7-12）。

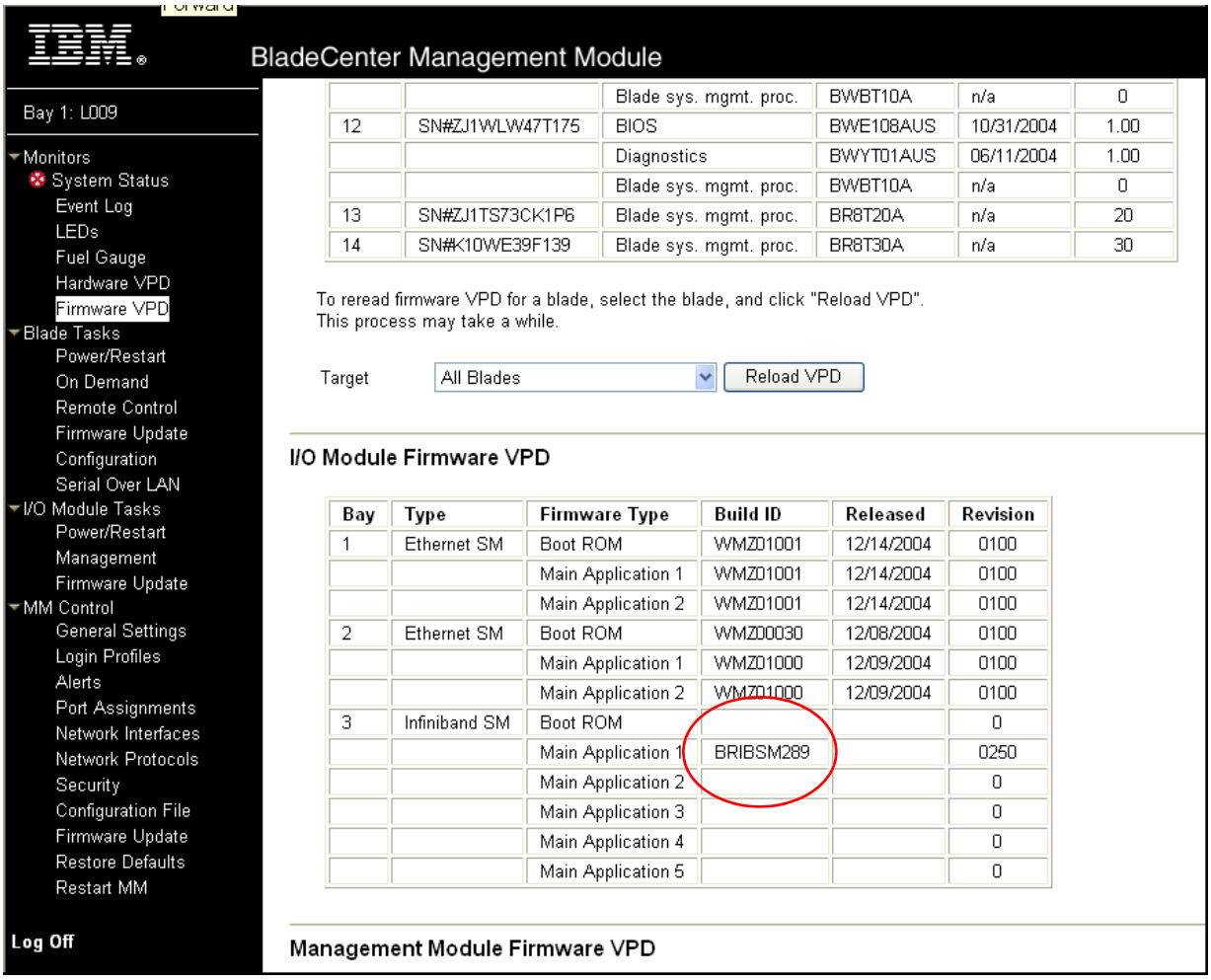

図 7-12　BladeCenter Management Module: I/O Module Firmware VPD

4. ping 要求を送信して、スイッチがアクティブであるかどうかを確認する（図 7-13）。管理モジュールから、「**I/O Module Tasks**」→「**Management**」をクリックします。「Bay 3 InfiniBand SM」ウィンドウから、「**Advanced Management**」をクリックします。「Advanced Management for I/O Module 3」ウィンドウから、「**Send Ping Requests**」→「**Ping Switch Module**」をクリックします。ping が成功したら、スイッチは InfiniBand ファブリックにいつでも接続できます。

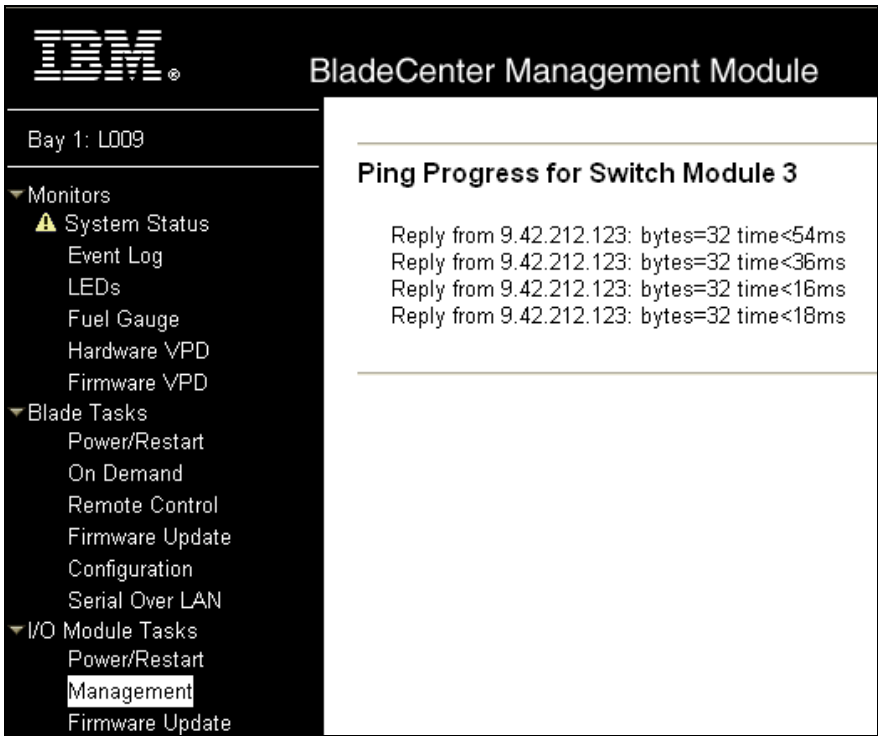

図 7-13　スイッチ・モジュールの Ping

注 : コマンド・プロンプトからスイッチ・モジュールを ping することもできます（図 7-14）。

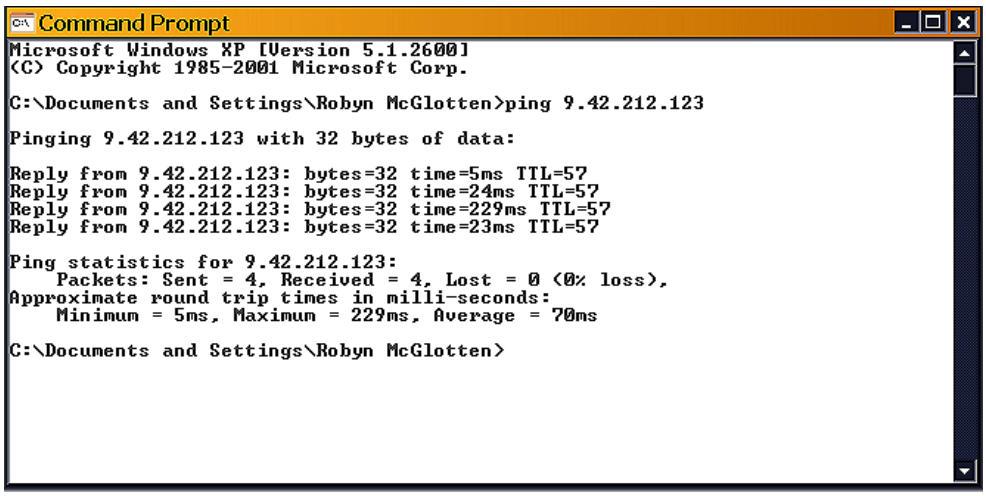

図 7-14　コマンド・プロンプトからのスイッチの Ping

7.5 InfiniBand スイッチ上のファームウェアの更新

BladeCenter InfiniBand スイッチのファームウェアを更新するには、複数の方法があります。下記の方法の中から、ニーズに最も適した方法を選択してください。

7.5.1 Element Manager の使用

Topspin スイッチ・モジュールの管理に使用できる複数のツールがあります。

- Element Manager（GUI）
- Chassis Manager（Web インターフェース）
- コマンド・ライン・インターフェース

ここでは、これらのツールの使用例をいくつか提示していますが、Topspin Element Manager の使用をお勧めします。

1. Topspin Element Manager を開き、BladeCenter スイッチの IP アドレスと SNMP コミュニティーの名前（デフォルト：`secret`）を入力する。

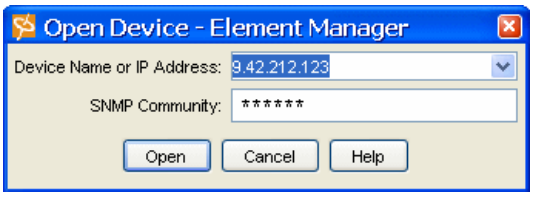

図 7-15　Open Device - Element Manager ログオン

ログオン後、Element Manager（図 7-16）は、BladeCenter シャーシ内のブレードと InfiniBand スイッチのグラフィカル表現を表示します。

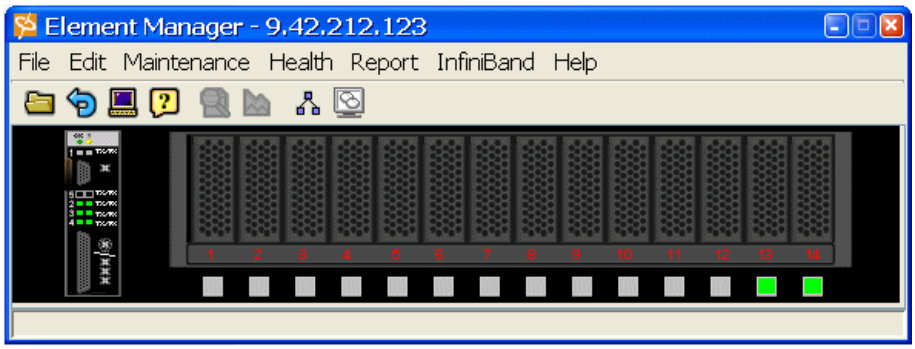

図 7-16　Element Manager

スイッチ上のアクティブ・ポートはすべて、スイッチの LED を表す緑色の正方形で強調表示されます。InfiniBand HCA が作動しているすべてのブレードには、下に緑色の正方形が表示されます。

2. 図 7-17 に表示されているように「**Maintenance**」→「**File Management**」をクリックする。

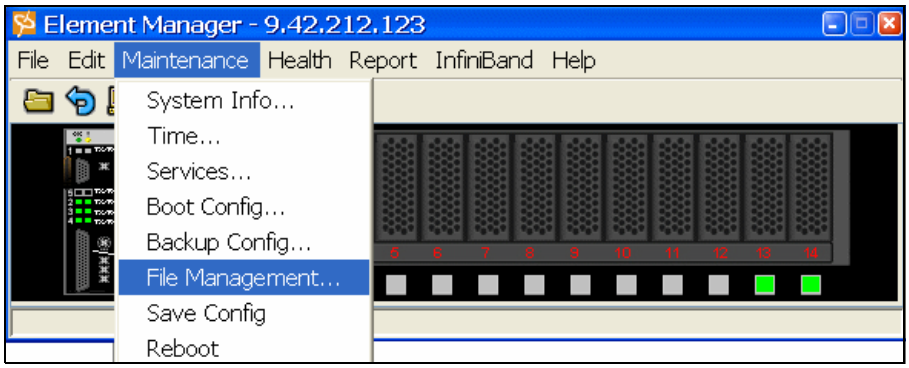

図 7-17 Element Manager: Maintenance

スイッチ上の現行ファイルのリストが、図 7-18 のようなウィンドウに表示されます。

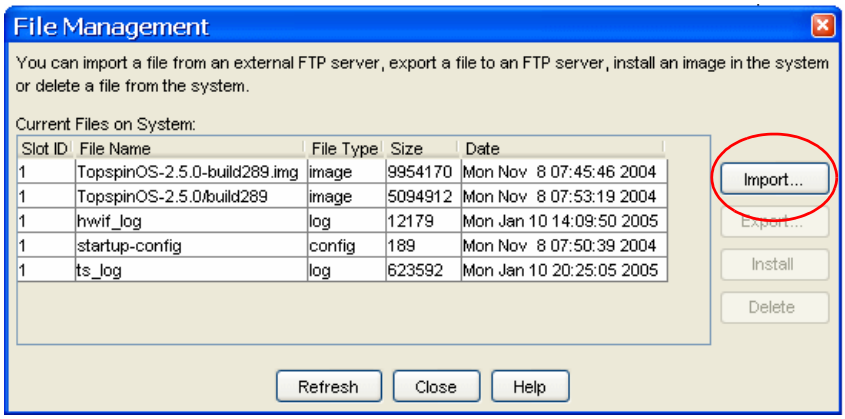

図 7-18 File Management

3. 「**Import**」をクリックして、BladeCenter スイッチにインストールするファイルを、FTP サーバーまたはローカル・システムからインポートする。これで、図 7-19 のようなウィンドウが開きます。

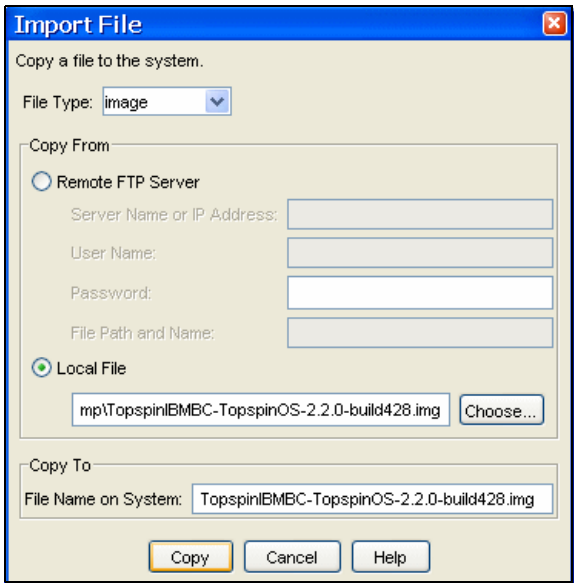

図 7-19　Import File

> **注 :** システム上のシステム・イメージ・ファイルが複数ある場合、スイッチに新しいファイルをインポートできません。「File Management」ウィンドウ（88 ページの図 7-18）で古いイメージ・ファイルの 1 つをクリックして強調表示し、「**Delete**」をクリックしてください。古いファイルが削除された後、ファイルのインポートを再試行してください。

4. 「File Management」ウィンドウ（88 ページの図 7-18）で、新しいイメージをクリックして強調表示し、「**Install**」をクリックする。
5. インストールが完了したら、「**Maintenance**」→「**Boot Config**」を選択する（図 7-21）。

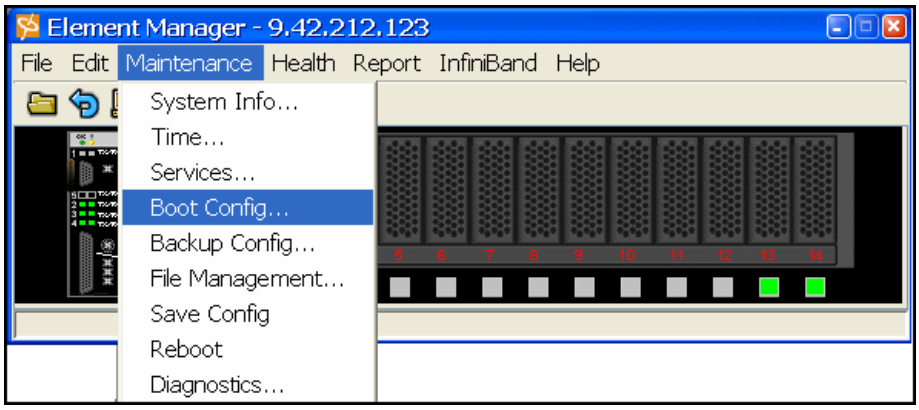

図 7-20　Element Manager → Boot Config

6. 「Image Source for Next Reboot」のプルダウン矢印をクリックし、新しいファームウェア・イメージを選択する。

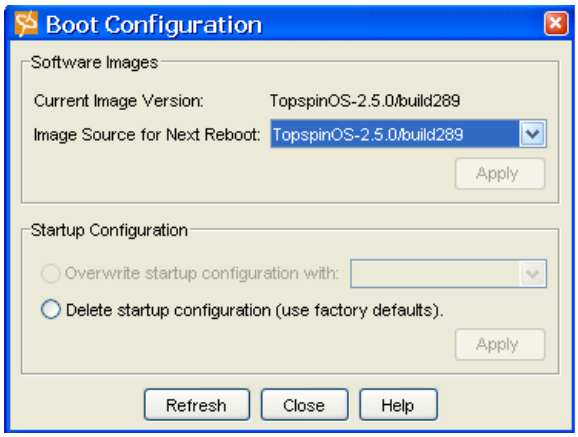

図 7-21　Boot Configuration

7.「**Apply**」をクリックしてから、「**Close**」をクリックする。

8.「**Maintenance**」→「**Reboot**」を選択する（図 7-22）。

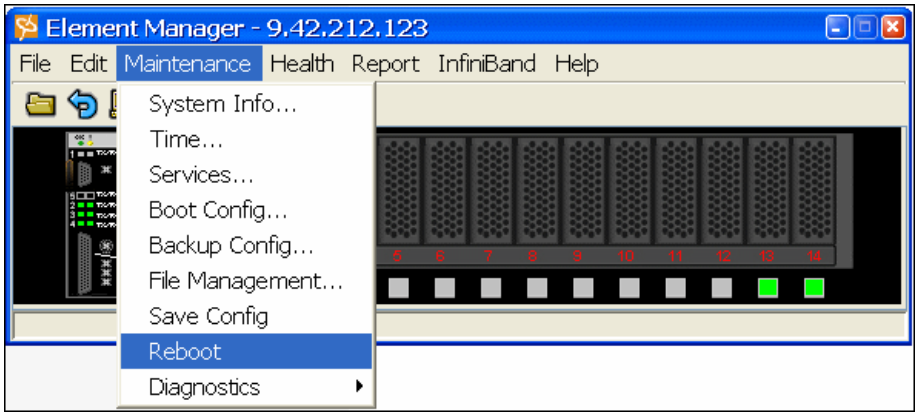

図 7-22　Element Manager: Maintenance: Reboot

9. システム構成の変更を保管するかどうかをたずねられたら、「**Yes**」をクリックする。

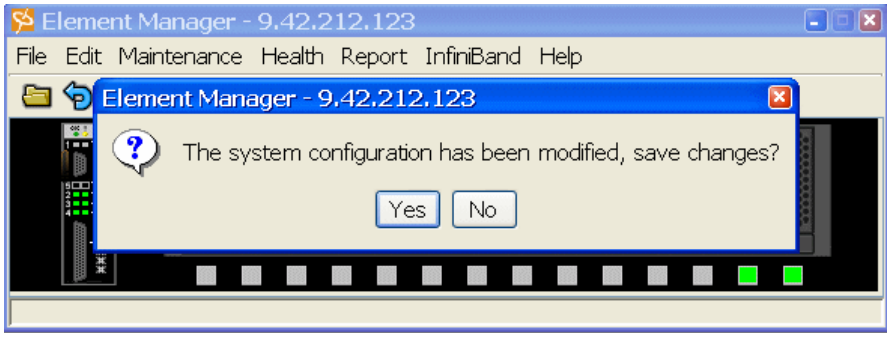

図 7-23　Element Manager の保管

10.「**OK**」をクリックして、シャーシをリブートする。

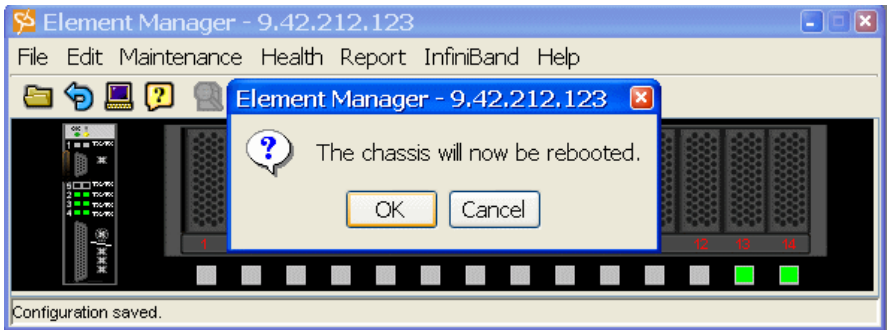

図 7-24　シャーシのリブート

11. スイッチのリブート後、「Maintenance」メニューでブート構成を調べて、更新されているかどうかを確認する。

7.5.2 Chassis Manager の使用

Chassis Manager を使用してスイッチのファームウェアを更新する手順は、次のとおりです。

1. Web ブラウザーを開き、BladeCenter InfiniBand スイッチの IP アドレスを入力する。これで、Chassis Manager のログイン・ウィンドウ（図 7-25）が開きます。

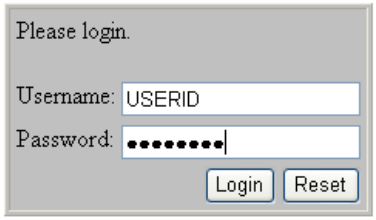

図 7-25　Chassis Manager - Login

ログオンすると、右側のフレームに「Device View」（図 7-26）が表示されます。このビューは、BladeCenter シャーシ内のブレードと InfiniBand スイッチのグラフィカル表現を表示します。スイッチ上のアクティブ・ポートはすべて、スイッチの LED を表す緑色の正方形で強調表示されます。InfiniBand HCA が作動しているすべてのブレードには、下に緑色の正方形が表示されます。

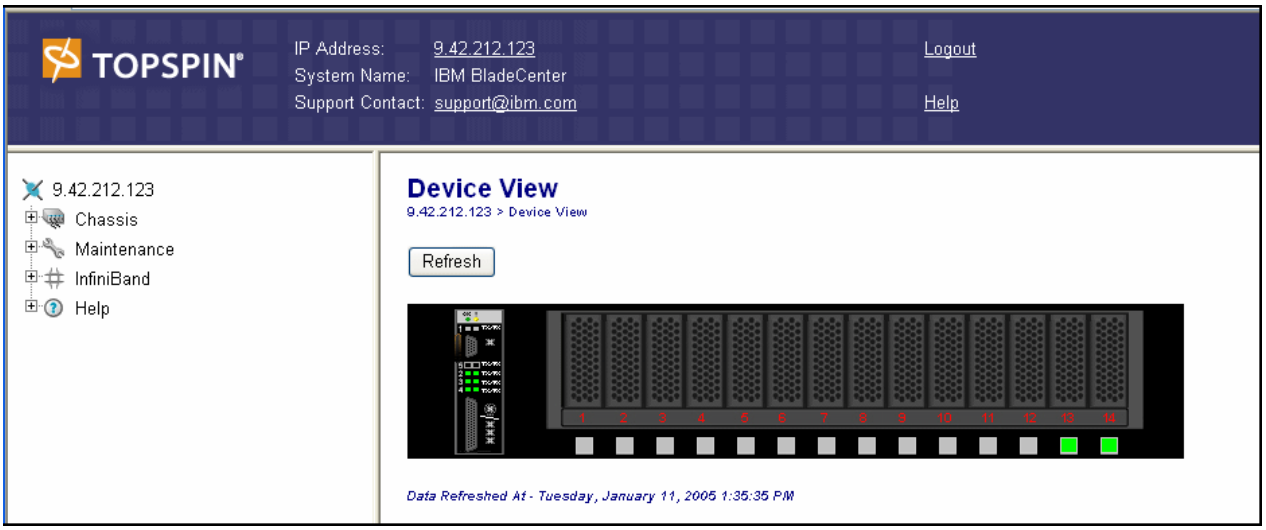

図 7-26 Chassis Manager: Device View

2. 左側のナビゲーション・メニュー（図 7-27）で、「**Maintenance**」→「**File Management**」をクリックする。右側のフレームの「File Management」ビューは、スイッチ上のすべてのファイルを表示します。

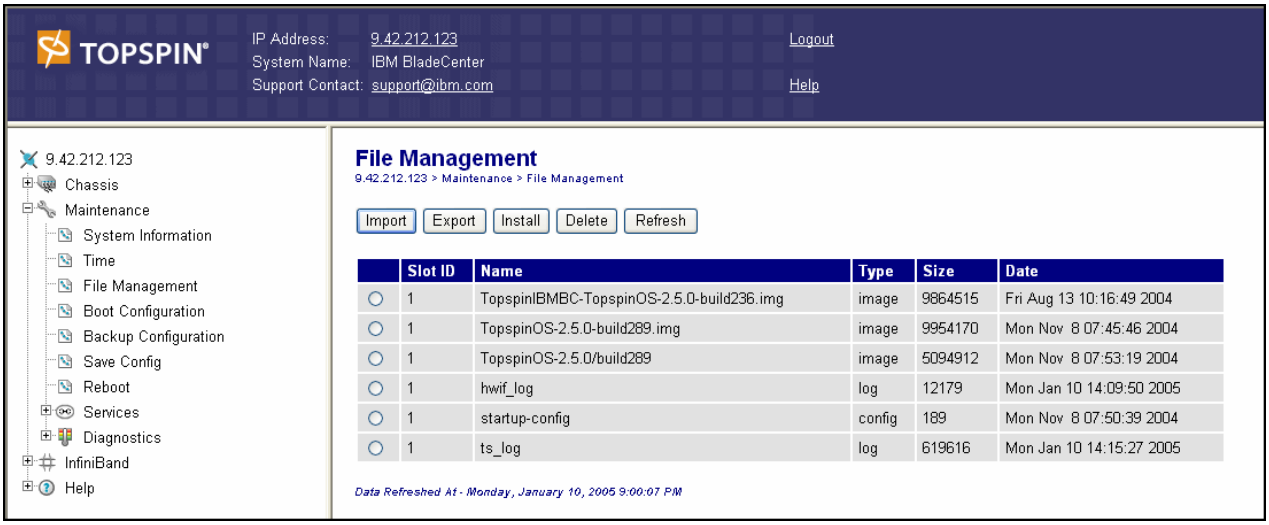

図 7-27 Chassis Manager: File Management

3. 「**Import**」をクリックして、FTP サーバーに保管されているファイルをスイッチにインポートする。これで、図 7-28 のようなウィンドウが開きます。

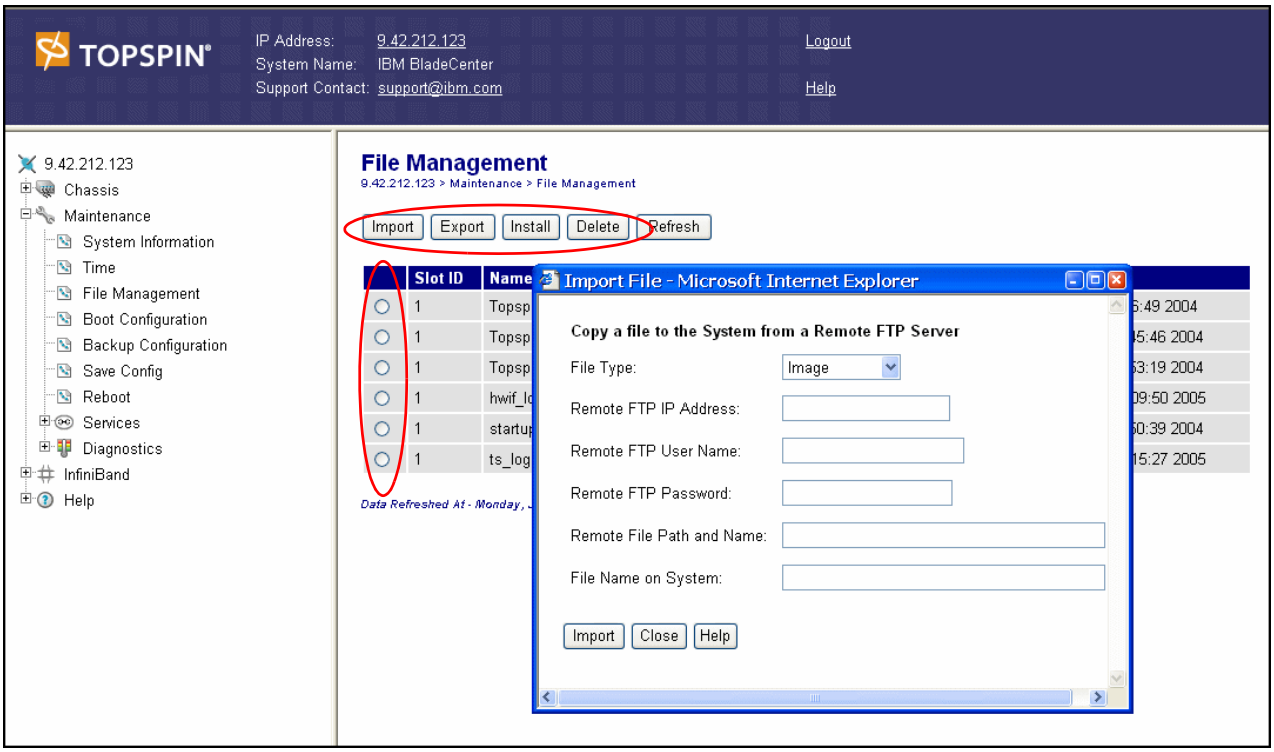

図 7-28　Chassis Manager: File Management

4. システム上のシステム・イメージ・ファイルが多過ぎる場合、スイッチに新しいファイルをインポートできません。古いイメージ・ファイルの 1 つの横にあるラジオ・ボタンをクリックして、「**Delete**」をクリックします。古いファイルが削除された後、ファイルのインポートを再試行してください。
5. 新しいファームウェア・イメージ・ファイルの横にあるラジオ・ボタンをクリックし、「**Install**」をクリックする。
6. インストールが完了したら、左側のナビゲーション・メニューで「**Maintenance**」→「**Boot Configuration**」を選択する（図 7-29）。

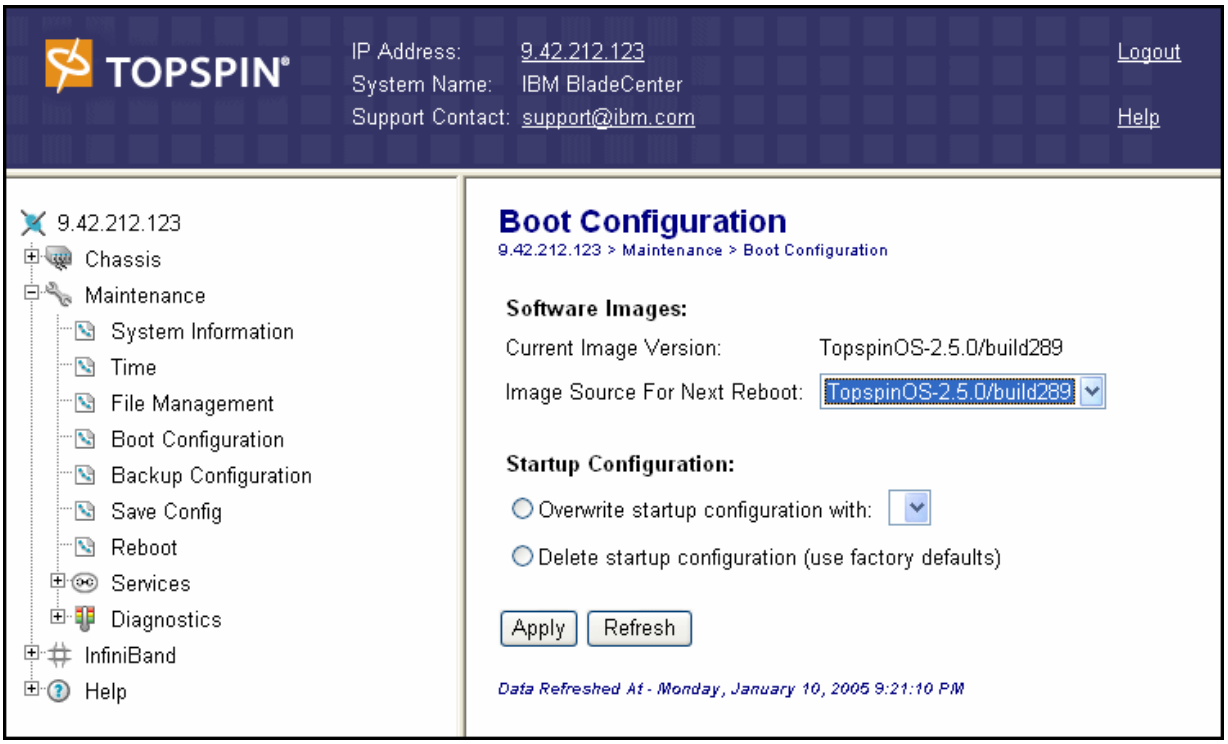

図 7-29 Chassis Manager: Boot Configuration

7. 「Image Source for Next Reboot」の横にあるプルダウン矢印をクリックし、新しいファームウェア・イメージを選択する。
8. 「**Apply**」をクリックする。
9. 「**Maintenance**」→「**Reboot**」を選択する。「Reboot」ウィンドウの「**Reboot**」ボタンをクリックします。

図 7-30 Chassis Manager: Reboot

10. スイッチのリブート後、「Boot Configuration」でスイッチのファームウェア・レベルを調べて、更新されているかどうかを確認する。

7.5.3 CLI の使用

ここでは、CLI を使用してファームウェアを更新する方法を説明します。

1. BladeCenter InfiniBand スイッチとの Telnet セッションを開き、ログオンする。

   Login USERID
   Password PASSWORD

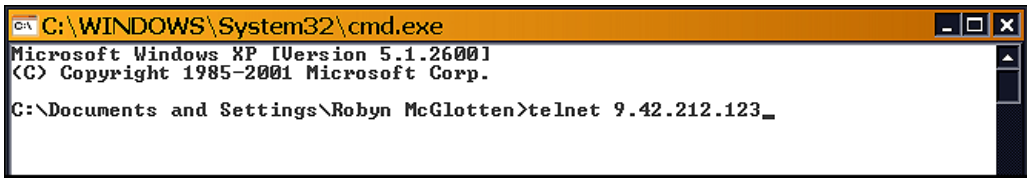

図 7-31 Telnet ログオン

スイッチにログオンしたら、User Execute モードに入ります。このモードでは、数個のコマンドのみに制限されます。CLI プロンプトで、`enable` を入力し、Enter を押して、Privileged Execute モードに入ります。このモードでは、Topspin Switch へのアクセス権が増えます。詳細については、「*Topspin CLI Reference Guide*」を参照してください。

2. CLI **copy** コマンドを使用して、FTP サーバーから BladeCenter スイッチにファームウェア・イメージをコピーする（図 7-32）。

```
IBM BladeCenter# copy ftp://userid:passw0rd@9.42.212.114/TopspinIBMBC-TopspinOS-
2.2.0-build428.img image:TopspinIBMBC-TopspinOS-2.2.0-build428.img
operation completed successfully
IBM BladeCenter# _
```

図 7-32 CLI copy

3. スイッチ上のイメージが多過ぎると、新しいファイルをインポートできません。**delete** コマンドを使用して、古いイメージを除去します。古いファイルが削除された後、ファイルのインポートを再試行してください。スイッチ上のすべてのイメージを表示するには、**dir** コマンドを使用します。

```
IBM BladeCenter# copy ftp://userid:passw0rd@9.65.243.135/TopspinIBMBC-TopspinOS-
2.2.0-build428.img image:TopspinIBMBC-TopspinOS-2.2.0-build428.img
Error: Exceeding the maximum number of system image files allowed.
IBM BladeCenter# dir image


================================================================================
                          Existing Boot-Images on System
================================================================================
slot date-created               size      file-name
--------------------------------------------------------------------------------
1    Fri Aug 13 10:16:49 2004  9864515   TopspinIBMBC-TopspinOS-2.5.0-build236.i
mg
1    Mon Nov  8 07:45:46 2004  9954170   TopspinOS-2.5.0-build289.img
1    Mon Nov  8 07:53:19 2004  5094912   TopspinOS-2.5.0/build289

IBM BladeCenter# delete image:TopspinIBMBC-TopspinOS-2.5.0-build236.img
Delete file 1:TopspinIBMBC-TopspinOS-2.5.0-build236.img? [yes(default) | no] yes
******
IBM BladeCenter# dir image


================================================================================
                          Existing Boot-Images on System
================================================================================
slot date-created               size      file-name
--------------------------------------------------------------------------------
1    Mon Nov  8 07:45:46 2004  9954170   TopspinOS-2.5.0-build289.img
1    Mon Nov  8 07:53:19 2004  5094912   TopspinOS-2.5.0/build289

IBM BladeCenter#
```

図 7-33 CLI delete

4. **install** コマンドを使用して、イメージをインストールする。

```
IBM BladeCenter# install image:TopspinIBMBC-TopspinOS-2.2.0-build428.img
Proceed with install? [yes(default) | no] yes
********************************************** operation completed successfully
IBM BladeCenter#
```

図 7-34 CLI install

5. **dir** コマンドをもう一度使用して、作成された新しいイメージ・フォルダーを確認する。

```
IBM BladeCenter# dir image


================================================================================
                          Existing Boot-Images on System
================================================================================
slot date-created               size      file-name
--------------------------------------------------------------------------------
1    Tue Jan 11 08:42:26 2005  9716464   TopspinIBMBC-TopspinOS-2.2.0-build428.i
mg
1    Tue Jan 11 08:48:37 2005  4913152   TopspinOS-2.2.0/build428
1    Mon Nov  8 07:45:46 2004  9954170   TopspinOS-2.5.0-build289.img
1    Mon Nov  8 07:53:19 2004  5094912   TopspinOS-2.5.0/build289

IBM BladeCenter#
```

図 7-35 CLI dir の確認

6. **boot-config** コマンドを使用して、新しいイメージを boot config イメージに設定する。

```
IBM BladeCenter(config)# boot-config primary-image-source TopspinOS-2.2.0/build4
28
IBM BladeCenter(config)#
```

図 7-36 CLI boot-config

7. **reload** コマンドを使用して、システムをリブートする。プロンプトが表示されたら、必ず構成を保管してください。

```
IBM BladeCenter# reload
System configuration is modified. Save? [yes(default) | no | filename] yes
Proceed with reload? [yes(default) | no] yes


Connection to host lost.
```

図 7-37 CLI reboot

8. 数分待ってから、スイッチに再度ログオンする。**enable** と入力し、Enter を押してから、**dir image** コマンドを使用して、古いイメージ・フォルダーが削除されていることを確認します。**show card** コマンドを使用して、新しいイメージがブート・イメージになっていることを確認します（97 ページの図 7-38）。

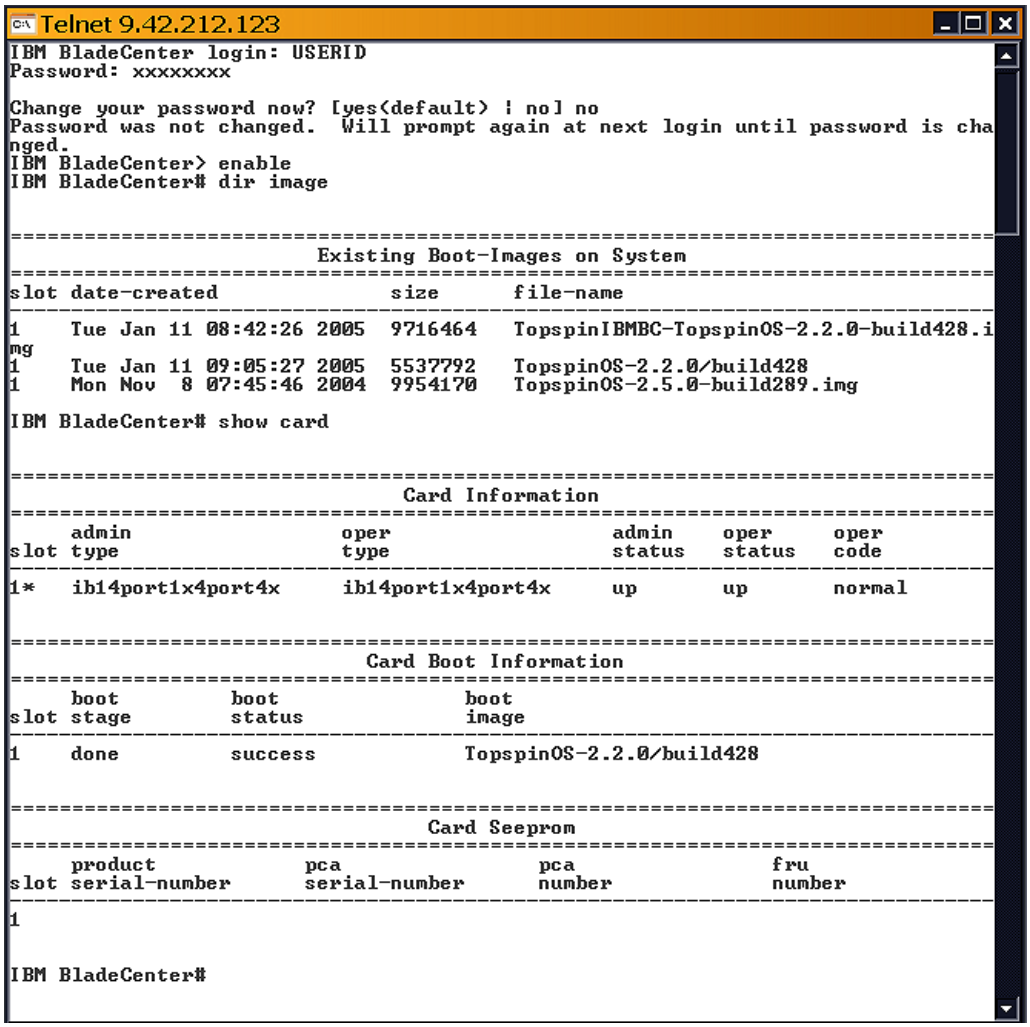

図 7-38 CLI の確認

7.6 外部スイッチの構成

外部 Topspin スイッチに割り当てられているデフォルト IP アドレスはないので、シリアル接続を通じて CLI を使用して初期構成を完了する必要があります。

1. シリアル・ポートを備えたシステムから、外部スイッチ上の管理ポートにシリアル・ケーブルを接続する。
2. 次のパラメーターを使用して端末セッションを開く。

ボー	9600bps
データ・ビット	8
パリティー	なし
フロー制御	なし
ストップ・ビット	1

3. CLI で、外部スイッチにログオンする。

Login	`super`
Password	`super`

注 :CLI プロンプトを表示するには、Enter を押す必要がある場合があります。

4. 管理カードが作動可能であることを確認する。
   a. スイッチにログオンしたら、User Execute モードに入ります。このモードでは、数個のコマンドのみに制限されます。CLI プロンプトで、`enable` を入力し、Enter を押して、Privileged Execute モードに入ります。このモードでは、Topspin Switch へのアクセス権が増えます。（詳しくは、「Topspin CLI reference」を参照してください。）
   b. **`show card`** コマンドを入力し、Enter を押す。これは、スイッチ上のすべてのカード（管理カード、ゲートウェイ・カード、スイッチ・カード）の状況とブート・イメージ情報を表示します。

```
Topspin-360# show card

================================================================================

                              Card Information
================================================================================

     admin                oper                 admin    oper     oper
slot type                 type                 status   status   code
--------------------------------------------------------------------------------

1    controller           controller           up       up       normal
2    en6port1G            en6port1G            up       up       normal
5    fc2port2G            fc2port2G            up       up       normal
6    fc2port2G            fc2port2G            up       up       normal
14   controller           controller           up       up       standby
15   ib12port4x           ib12port4x           up       up       normal
16   ib12port4x           ib12port4x           up       up       normal


================================================================================

                            Card Boot Information
================================================================================

     boot         boot               boot
slot stage        status             image
--------------------------------------------------------------------------------

1    done         success            TopspinOS-2.1.0/build132
2    done         success            TopspinOS-2.1.0/build132
5    done         success            TopspinOS-2.1.0/build132
6    done         success            TopspinOS-2.1.0/build132
14   done         success            TopspinOS-2.1.0/build132
15   done         success            TopspinOS-2.1.0/build132
16   done         success            TopspinOS-2.1.0/build132


================================================================================

                                Card Seeprom
================================================================================

     product            pca                pca                fru
slot serial-number      serial-number      number             number
--------------------------------------------------------------------------------

1    US2044300132       C2044300104        95-00005-01-C2     98-00001-01
2    US2043900022       C2043700178        95-00025-01-C0     98-00022-01
5    USC041400084       CS-0413-000718     95-00008-02-C3     98-00021-01
6    USC041400078       CS-0413-000760     95-00008-02-C3     98-00021-01
14   US2044300130       C2044300106        95-00005-01-C2     98-00001-01
15   US2043600065       C2044200028        95-00006-01-B2     98-00002-01
16   US2044300067       C2044200027        95-00006-01-B2     98-00002-01
```

図 7-39　360 または 90 の show card 情報

5. スイッチの IP アドレスを入力する（図 7-39）。
   a. Privileged Execute モードから、`configure terminal` を入力し、Enter を押して、Global Configuration モードに入る。

b. そのモードで、interface mgmt-ethernet を入力してから、Enter を押す。これで、スイッチのイーサネット管理ポートにアクセスできます。

c. プロンプトで、ip address xx.xx.xx.xx yy.yy.yy.yy と入力する。ここで、xx.xx.xx.xx はスイッチの IP アドレスであり、yy.yy.yy.yy はスイッチのサブネット・マスクです。

```
Topspin-90# configure terminal
Topspin-90(config)# interface mgmt-ethernet
Topspin-90(config-if-mgmt-ethernet)# ip address 10.10.10.10 255.255.255.0
Topspin-90(config-if-mgmt-ethernet)# no shutdown
Topspin-90(config-if-mgmt-ethernet)#
```

図 7-40 IP アドレスの設定

6. コマンド・プロンプトで no shutdown と入力し、Enter を押して管理ポートを使用可能にする。

7. 現行の構成を保管する。CLI プロンプトで、copy running-config startup-config を入力して、スイッチの始動構成に加えた変更を保管します。

```
Topspin-90# copy running-config startup-config
```

図 7-41 実行構成の保管

これで、Telnet、Chassis Manager、または Element Manager を通じてスイッチにアクセスできるようになりました。

7.7 Element Manager のセットアップ

Element Manager がネットワーク上で Topspin スイッチにアクセスする前に、外部管理インターフェースを構成する必要があります。この手順を実行しないと、Topspin スイッチにアクセスし、構成する方法は、直接接続されたシリアル・ケーブルと端末プログラム（Microsoft HyperTerminal など）を介した方法しかありません。

これを実行する手順は、次のとおりです。

1. 端末プログラムがあるマシンにシリアル・ケーブルを接続する。このプログラムは、次のように構成する必要があります。

ボー	9600
データ・ビット	8
パリティー	なし
ストップ・ビット	1
フロー制御	なし

2. ユーザー ID super、パスワード super を使用してログオンしてから、enable と入力する。

3. **config term** と入力してから、次を入力する。

   – interface mgmt-ethernet xxx.xxx.xxx.xxx yyy.yyy.yyy.yyy を入力する。ここで、x は IP アドレスであり、y はサブネット・マスクです（例 : 9.42.165.139 255.255.255.0）。

   – no shutdown を入力して、ポートを使用可能にする。

     Ctrl+z を押して、構成を実行状態に保管します。

   – save を入力して、構成を不揮発性メモリーに書き込む。

これで、Element Manager は 360 シャーシに接続できるようになりました。

7.7.1 外部ハード・ディスクの接続

ここでは、外部ハード・ディスクを接続した方法を説明します。

図 7-42 は構成を表示しています。BladeCenter に 14 枚のブレードを取り付け、各ブレードは InfiniBand HCA ドーター・カードを備えています。ブレード 1 から 6 には、Windows 2000 Server Advanced がロードされています。ブレード 7 から 14 には、Red Hat Linux Advanced Server 3.0 がロードされています。BladeCenter にはベイ 3 と 4 に、2 つの InfiniBand スイッチが取り付けられています。これらのスイッチはそれぞれ、12x ケーブルで Topspin 360 スイッチに接続されています。1 つのファイバー・チャネル・ゲートウェイにより、DS4500 との接続が可能になり、1 つのイーサネット・ゲートウェイにより、コア・ネットワークへのアクセスが可能になっています。

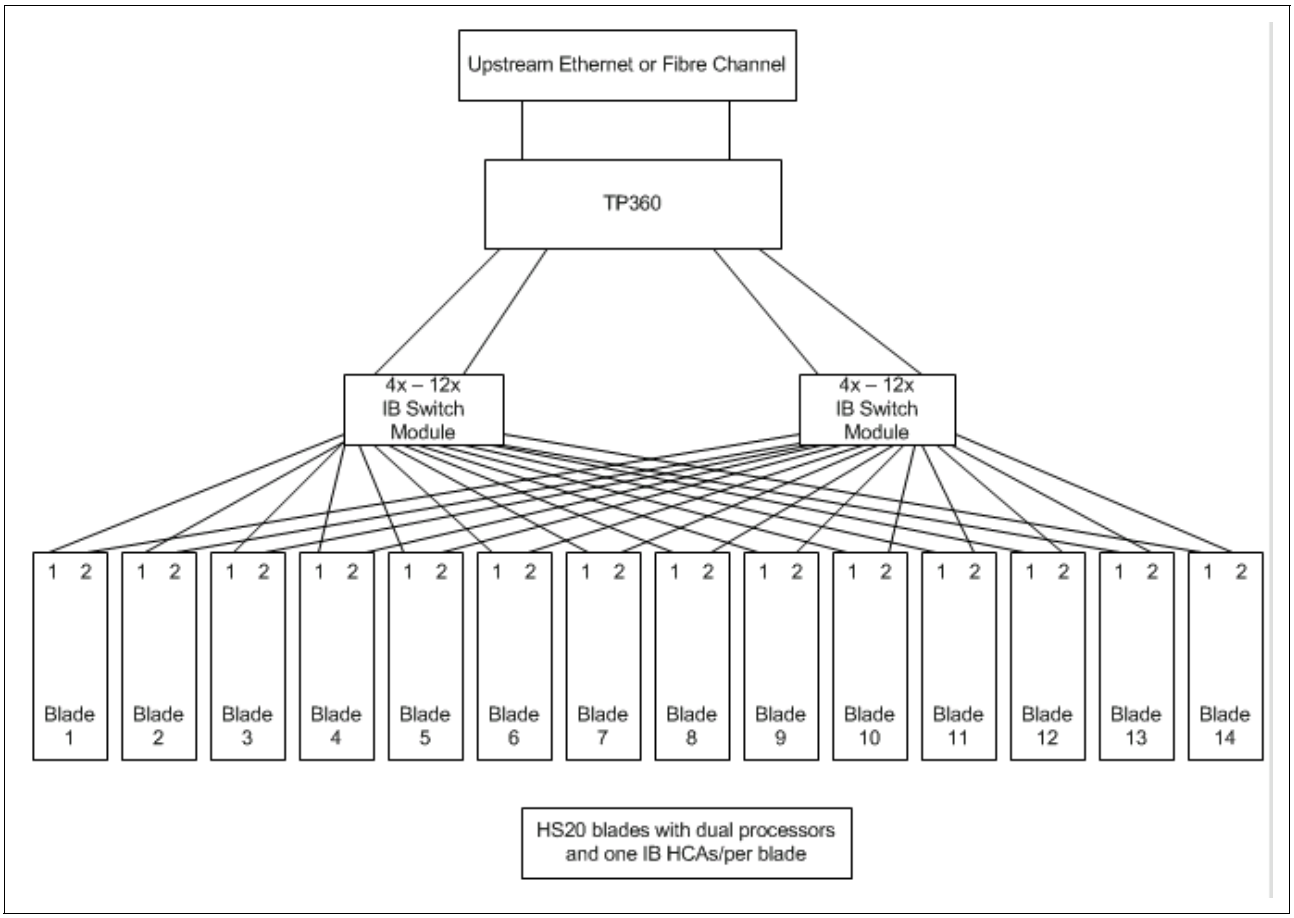

図 7-42　BladeCenter - InfiniBand Switch Module の図

Windows オペレーティング・システムは、Topspin 360 スイッチに取り付けられているファイバー・チャネル・ゲートウェイ・モジュールを使用して、InfiniBand ネットワークを通じて外部ハード・ディスクにアクセスできます。すべての物理接続が実行された後の構成プロセスは、次のとおりです。

1. すべての機器の電源を投入する。
   - BladeCenter、および InfiniBand HCA が取り付けられているブレード
   - 外部 Topspin 360 サーバー・スイッチ
   - IBM TotalStorage DS4500
2. 管理 PC から Topspin Element Manager を始動する。管理するスイッチの IP アドレスを要求するウィンドウが開きます。スイッチのアドレスを入力します。このウィンドウでは複

数のスイッチ・アドレスを保存できますが、一度に表示できるのは 1 つのスイッチのみです。複数のスイッチを管理する場合は、別の Element Manager アプリケーションを始動してください。

3. Element Manager で「**FibreChannel**」→「**Storage Manager**」を選択する（図 7-43）。

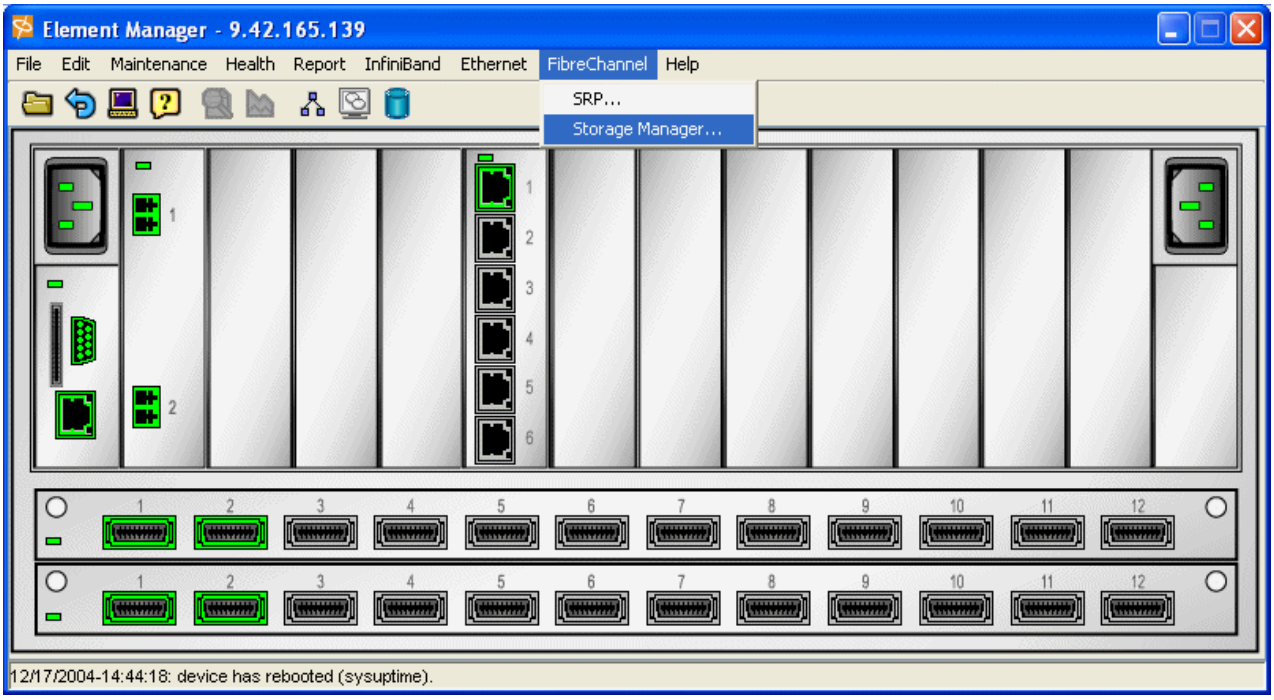

図 7-43　Element Manager

4. これで、図 7-44 のようなウィンドウが開きます。「Gateway Port Access」と「LUN Access」の「**Restricted**」チェック・マークを外し、「**Apply**」をクリックします。

> **注 :** Topspin スイッチがトラフィックを渡すには、これらの制限を取り除く必要があります。

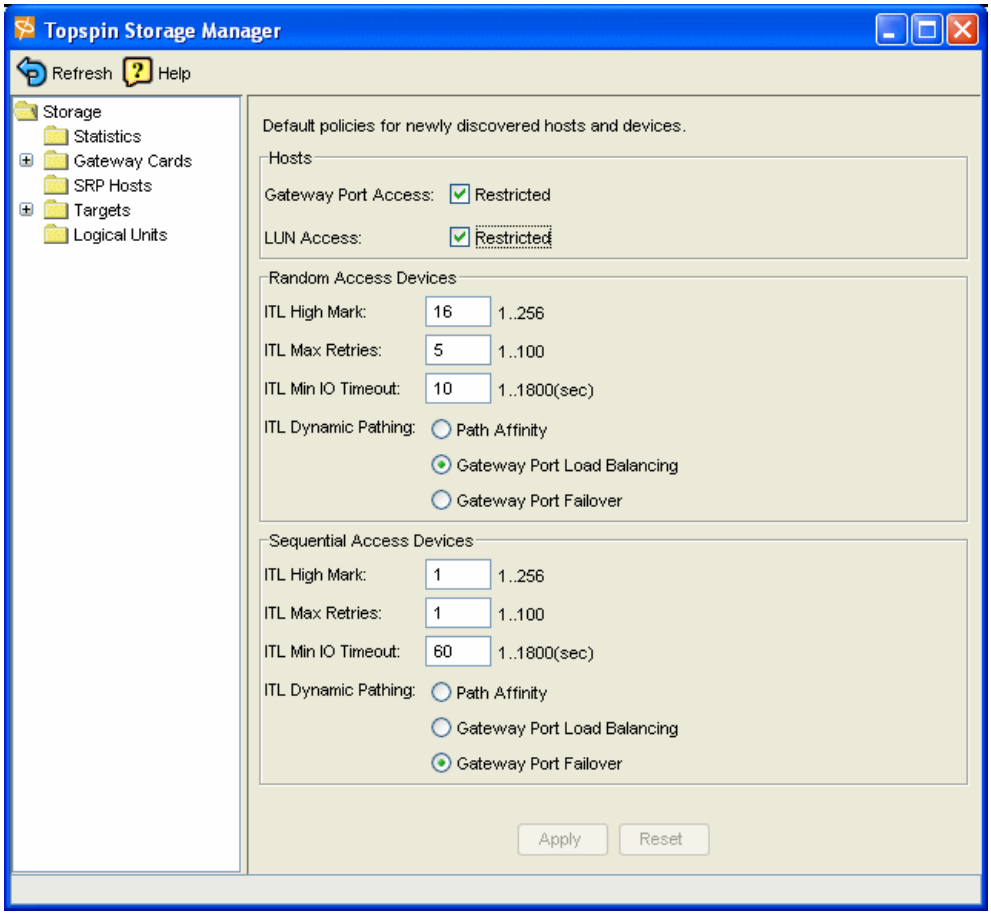

図 7-44　Topspin Storage Manager

5. 左側のナビゲーション・バーで、「**Storage**」→「**SRP Hosts**」を選択する。

6. 図 7-45 のようなウィンドウが開きます。「**Define New**」をクリックします。

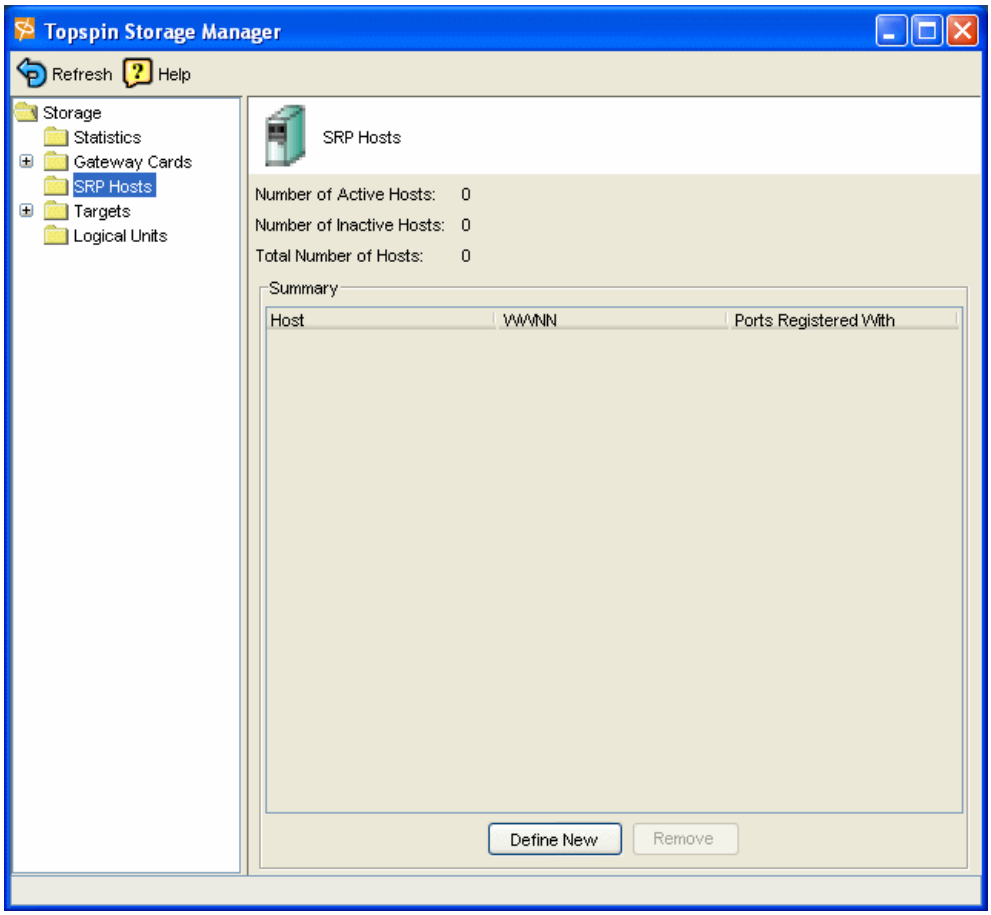

図 7-45　Topspin Storage Manager

7. これで、図 7-46 のようなウィンドウが開きます。追加するブレードの GUID を選択するか、手動で入力します。GUID は、ドーター・カードで見つかり、ブレードのブート時に BIOS で表示されます。
8. 「Description」フィールドに装置の名前を入力する。「**Next**」をクリックします。

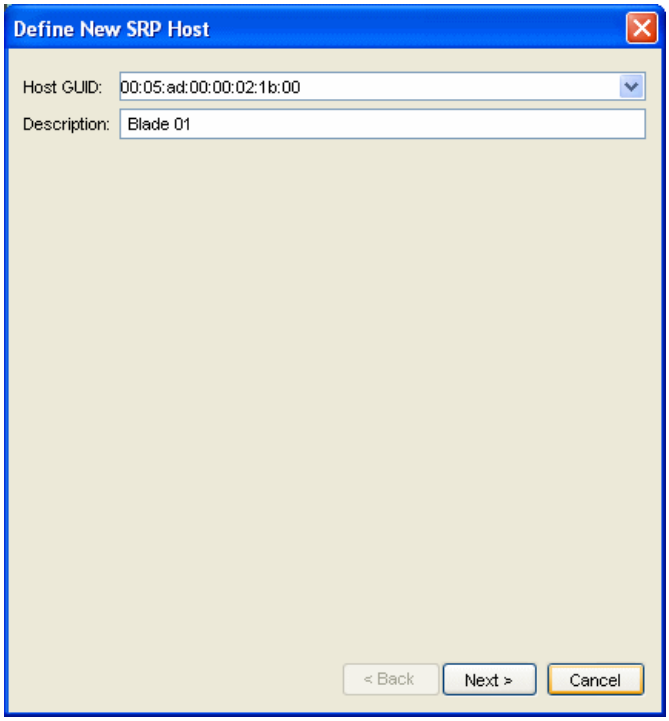

図 7-46 Define New SRP Host

9. 「**Finish**」をクリックする。
10. すべてのホストについてこの手順を繰り返す。
11. 完了したら、このウィンドウを閉じる。

7.7.2 ファイバー・チャネル・パスの構成

これで Topspin スイッチは、サーバーがファイバー・チャネル装置に接続することをいつでも許可できるようになりました。次に、物理ハード・ディスクとのファイバー・チャネル・パスを構成します。

1. FAStT Storage Manager を始動する。
2. IBM FAStT Storage Subsystem Manager を、インストールされている管理 PC から起動する。この例では、システム管理デスクトップです。

3. 図 7-47 のようなウィンドウが開きます。「**Logical Drive**」→「**Create**」を選択して、論理ドライブを作成します。

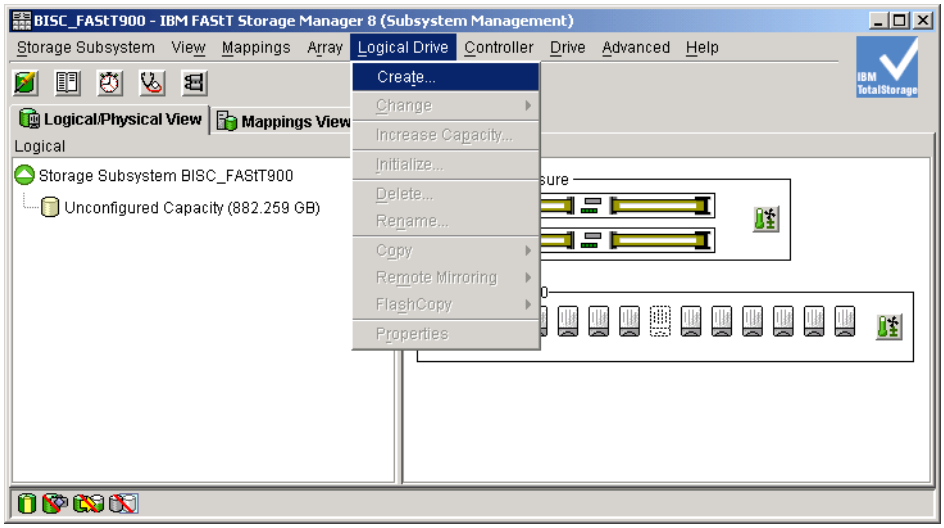

図 7-47　論理ドライブの作成

4. 図 7-48 のようなウィンドウが表示されます。ホスト・オペレーティング・システムのタイプを定義します。（この例の場合、「**Windows 2000/Server 2003 Non-Clustered**」を選択しました。）「**OK**」をクリックします。

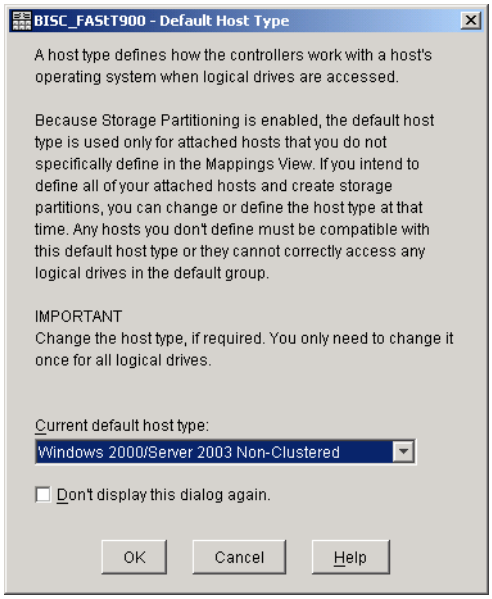

図 7-48　デフォルトのホスト・タイプ

5. 図 7-49 のようなウィンドウが表示されます。「**Unconfigured capacity**」を選択して、新しいアレイを作成します。デフォルトを受け入れ、「**Next**」をクリックします。

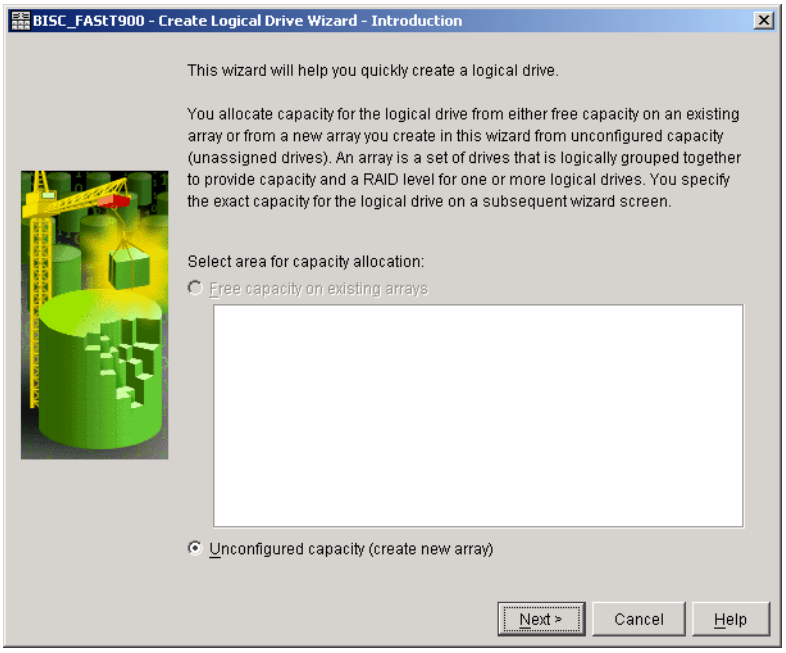

図 7-49 新規アレイの作成

6. 図 7-50 に表示されているウィンドウで、RAID レベル、および使用するハード・ディスクを定義する。
   a. RAID レベルを選択する。ここでは、データの損失やハード・ディスクの障害を気にしないので、**RAID 0** を選択しました。
   b. ドライブを選択する。ここでは、手動で選択することを選択しました。目的のドライブを強調表示します。（この例では、ブレードごとに 1 つのドライブのみを使用しました。ブレード 1 はドライブ 1 に割り当てられ、ブレード 2 はドライブ 2 に割り当てられ、以下同様です。）「**Apply**」をクリックします。

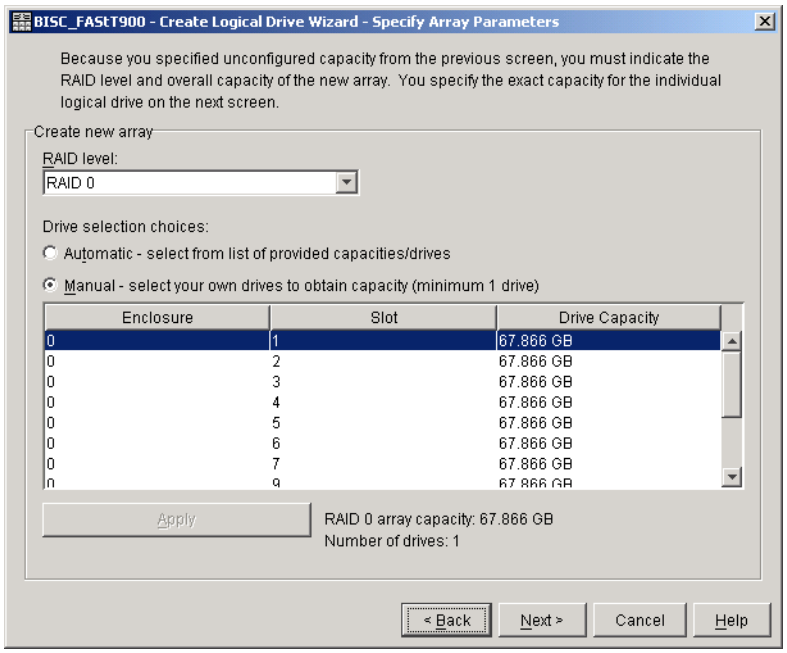

図 7-50　アレイ・パラメーターの指定

c. 「**Next**」をクリックして、図 7-51 に表示されているウィンドウを開く。

7. ハード・ディスクのパラメーターとその名前を定義する。

   a. 必要に応じて容量を変更する。

   b. もっと分かりやすい名前に変更する。ここでは、ハード・ディスクの名前に、割り当てられるブレード番号を付けました。

   c. 「**Finish**」をクリックする。

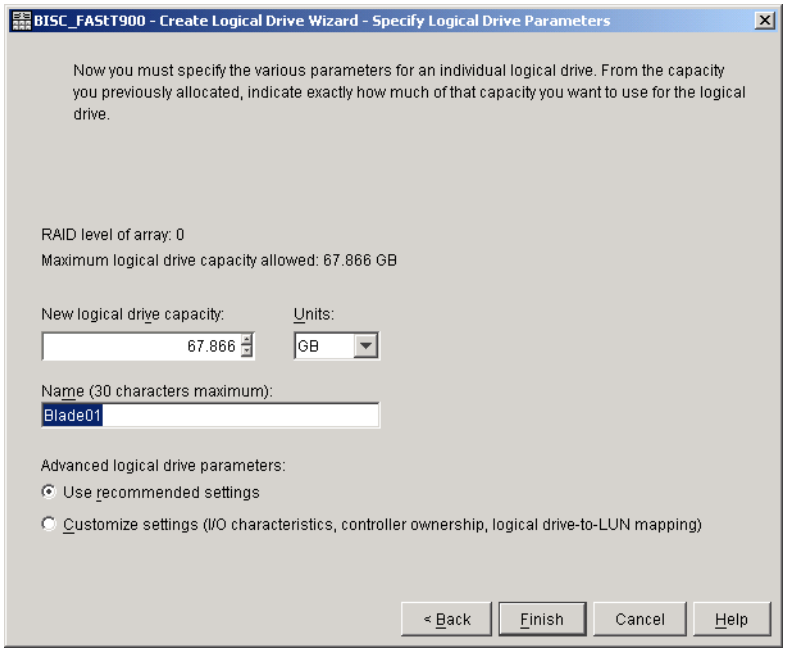

図 7-51　論理ドライブ・パラメーターの指定

d. 図 7-52 に表示されているウィンドウが開きます。「**Yes**」をクリックすると、必要な数のアレイに対してこの手順を実行できます。終了する場合は、「**No**」をクリックして閉じます。

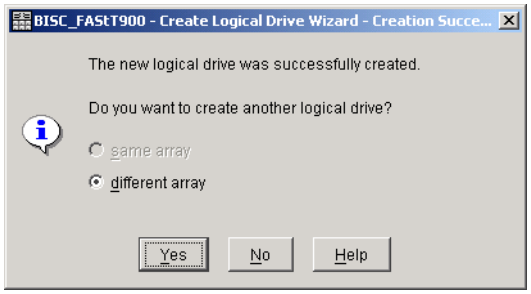

図 7-52　作成成功

8. 完了ウィンドウが表示されます。「**OK**」をクリックします。

ハード・ディスクの物理ビューが完成しました。次に、論理ビューを構成します。

9. マッピングとホスト・グループを定義するために、「**Mappings View**」タブの選択から開始する。

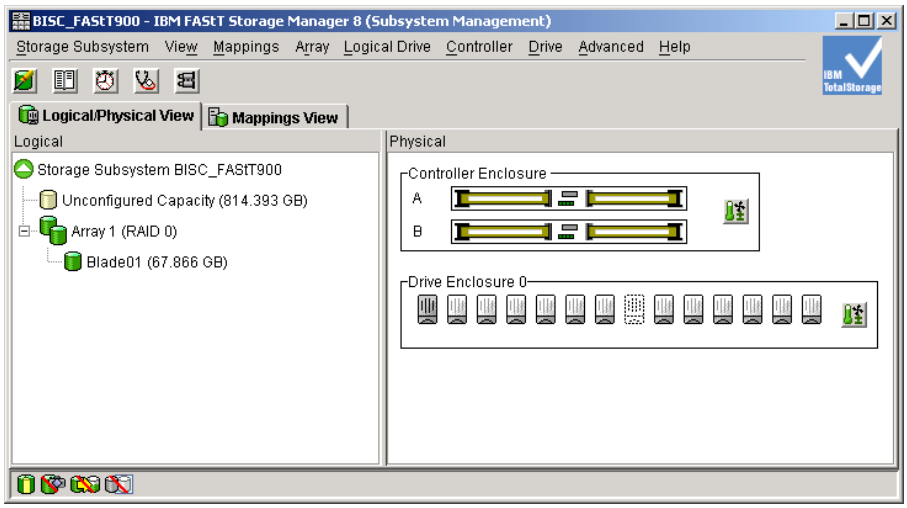

図 7-53　Logical Physical View（後）

10. 図 7-53 に表示されているウィンドウで、「**Mappings**」→「**Define**」→「**Host Group**」を選択する。

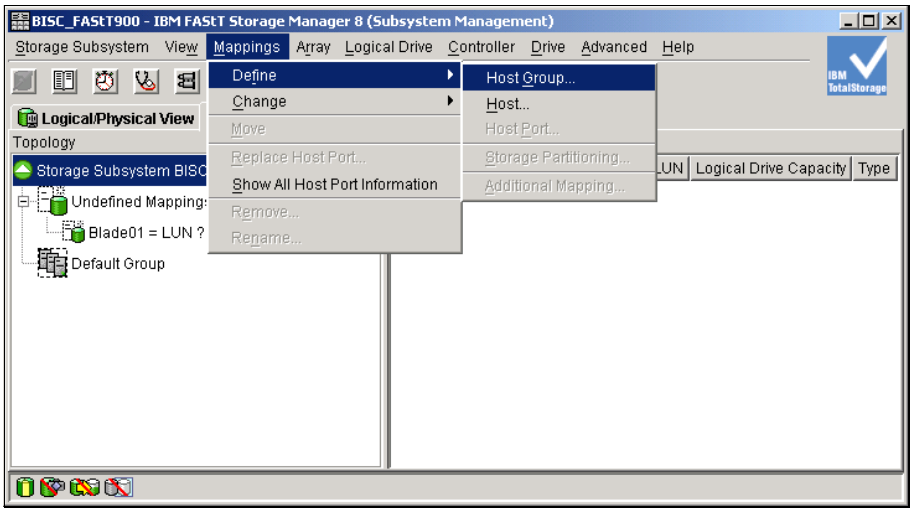

図 7-54　マッピングの定義

11. ホスト・グループ（ハード・ディスクのグループ）に名前を付けて、「**Add**」をクリックする（図 7-55）。（ラボ環境で使用しているハード・ディスクのグループを表すために IBRedpaper を使用しました。）「**Close**」をクリックします。

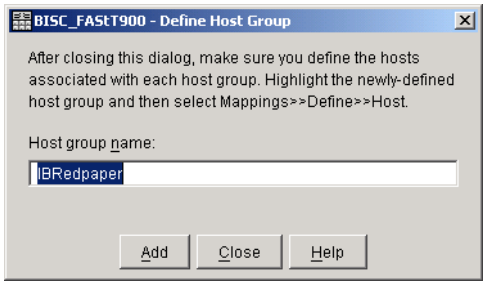

図 7-55　ホスト・グループの命名

12. ホストを追加するために、作成したばかりのホスト・グループを選択し、「**Mappings**」→「**Define**」→「**Host**」を選択する。これで、図 7-56 のようなウィンドウが開きます。
13. Blade 01 と入力し、「**Add**」をクリックして、ホスト名を定義する。必要な数のホストを入力できます。終了したら、「**Close**」をクリックします。

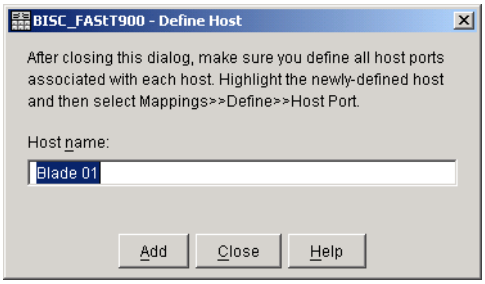

図 7-56　ホストの定義

14.「**Host Blade 01**」をクリックし、「**Mappings**」→「**Define - Host Port**」を選択して、ホスト・ポートを定義する（図 7-57）。

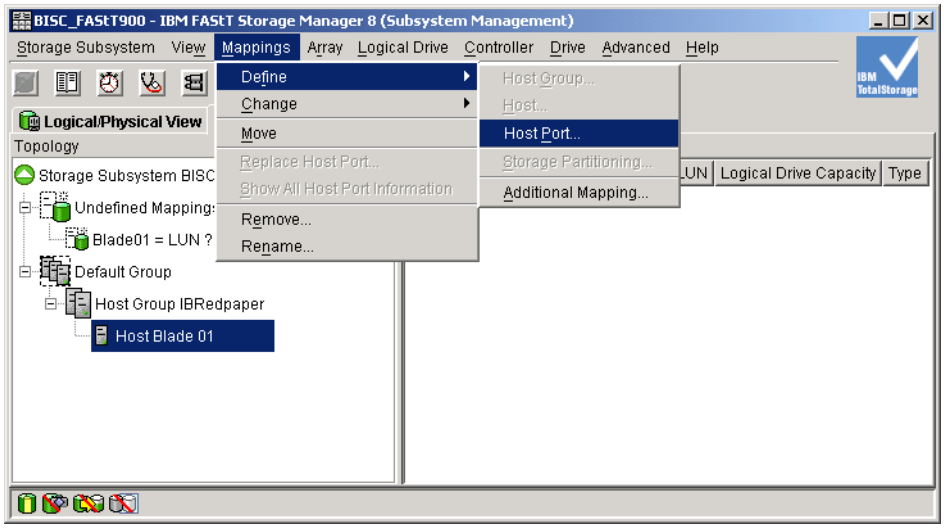

図 7-57　ホスト・ポートの定義（パート 1）

15. ポートの World Wide Name を識別する（110 ページの図 7-58）。

a. ホスト・ポート ID およびホスト・タイプ（オペレーティング・システム・タイプ）を選択する。

b. ポート名を変更しないでください。この名前は固有でなければなりません。デフォルト名のままにするのが最善の手順です。

c.「**Add**」をクリックする。

d. ドーター・カードの 2 番目のポートを選択する。

e.「**Add**」をクリックする。

f.「**Close**」をクリックする。

g. 入力したすべてのホストについて手順を繰り返す。

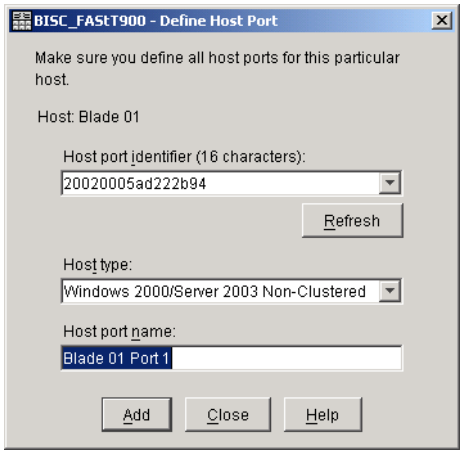

図 7-58　ホスト・ポートの定義

16. アレイに対するマッピングを定義するために、「**Undefined Mappings**」の下で LUN を強調表示し、「**Mappings**」→「**Define**」→「**Additional Mapping**」を選択する（図 7-58）。

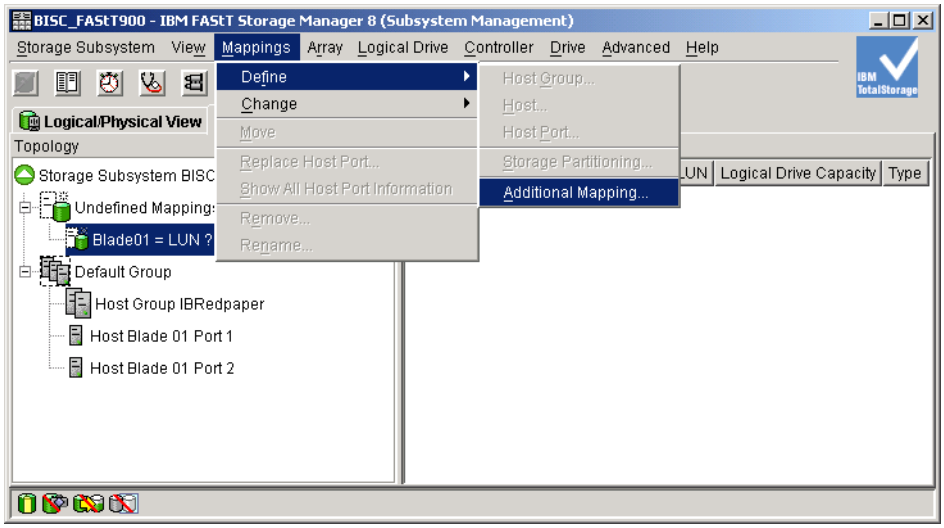

図 7-59　Mappings → Define → Additional Mapping

17. 図 7-60 のウィンドウが開きます。追加のマッピングを選択します。
    a. 作成したホスト・グループを選択し、論理ドライブの 1 つを選択する。「**Add**」をクリックします。
    b. すべての論理ドライブについてこの手順を繰り返し、終了したら「**Close**」をクリックする。

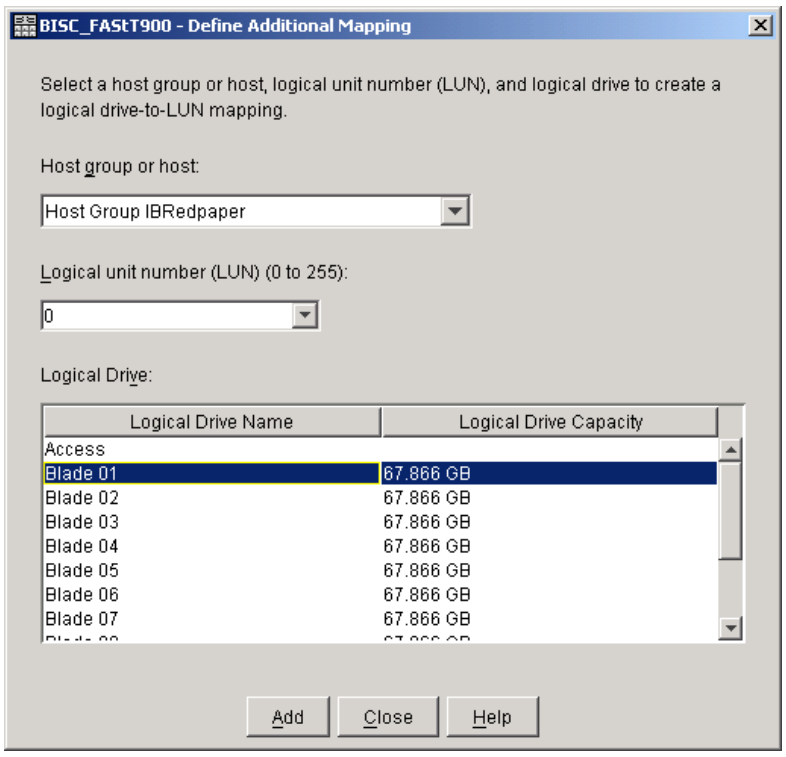

図 7-60　追加マッピングの定義

これで、各ブレードからハード・ディスクが使用可能になりました。ブレードが 1 つのハード・ディスクのみにアクセスするように制限できます。これを行うには、Topspin Element Manager を使用します。

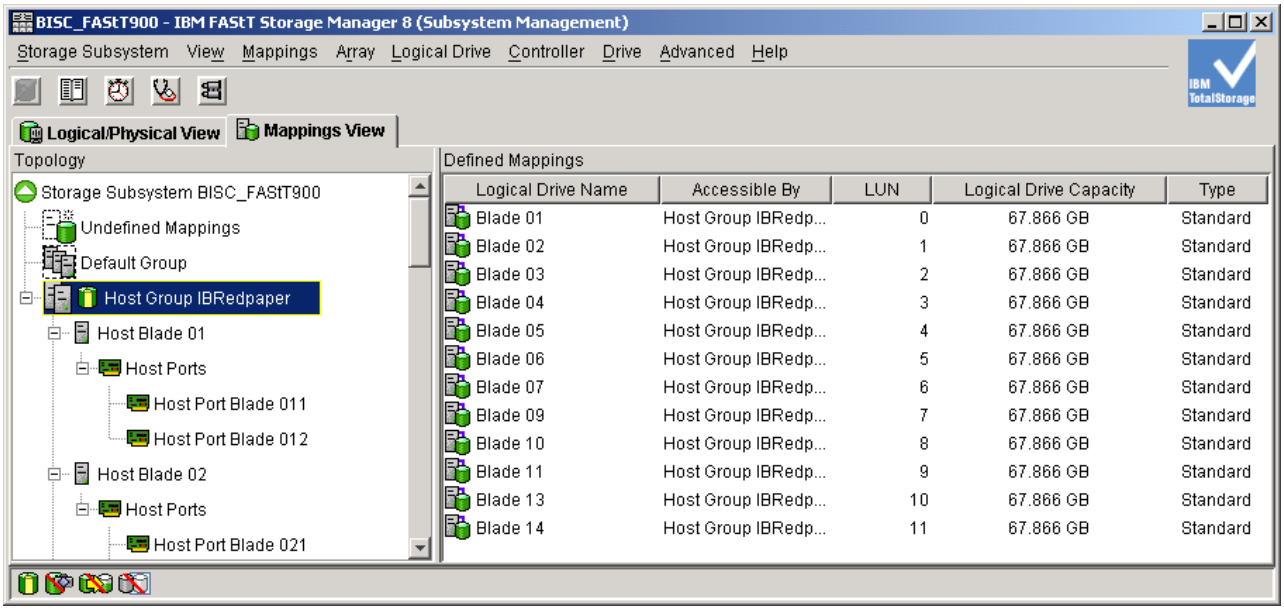

図 7-61　Mappings View

18. Topspin 360 を使用したハード・ディスク・アクセスを制限するために、Topspin Element Manager を始動し、「**FibreChannel**」→「**Storage Manager**」を選択する。
    a. 「**SRP Hosts**」を展開し、構成するホストを選択する。
    b. 「**LUN Access**」タブをクリックする。
    c. アクセスしようとする LUN を「Accessible LUNs」フィールドに移し、アクセスしない LUN を「Available LUNs」に移す。
    d. 「**Apply**」をクリックする。
    e. 必要に応じて、その他のホストについてこの手順を繰り返す。

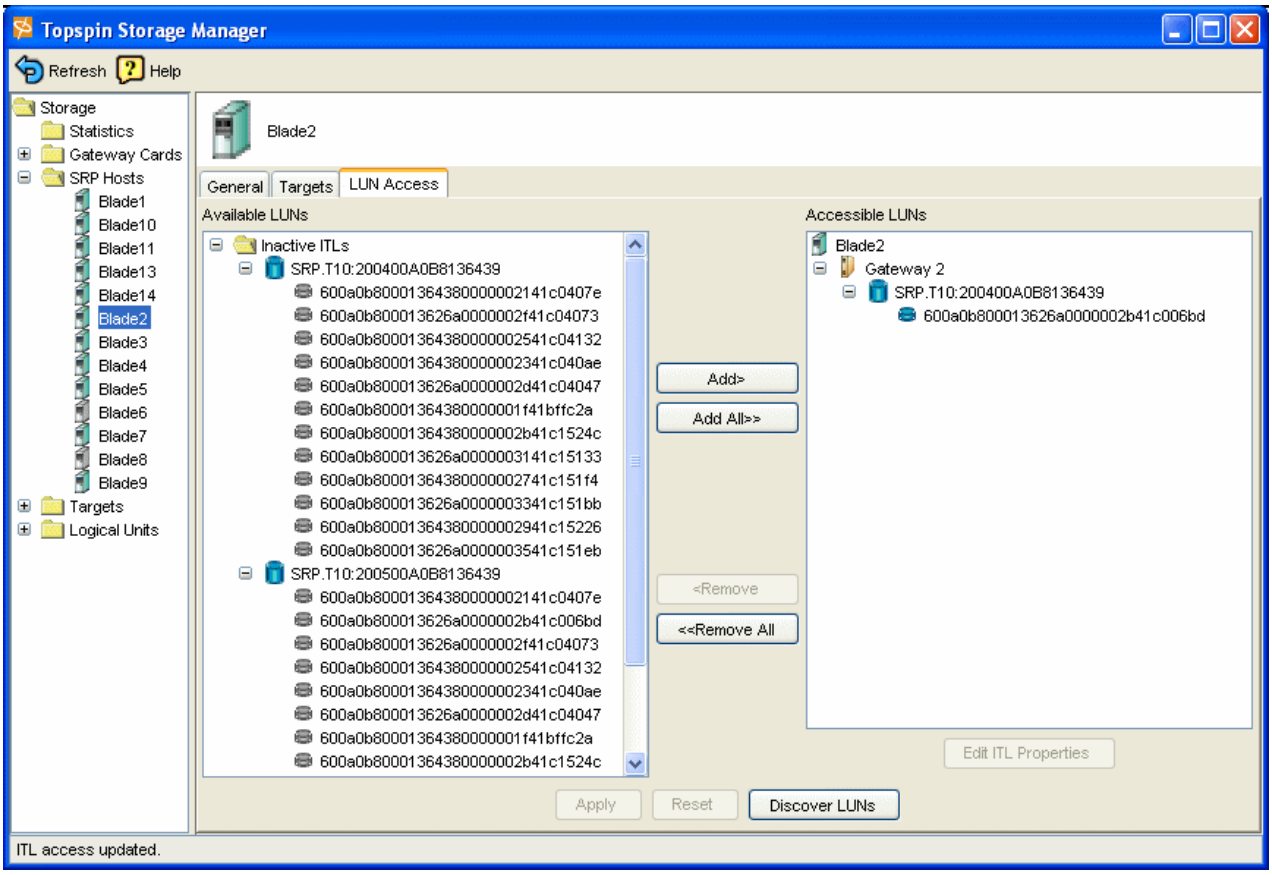

図 7-62　Topspin Storage Manager

19. サーバーでハード・ディスクが使用可能であることを確認する。Windows サーバーからこれを行うには、「コンピュータの管理」ウィンドウを開き、「**デバイスマネージャ**」を選択します。「**ディスクドライブ**」を選択すると、SCSI ディスク装置が表示されます。

8

標準的な構成

この章では、Windows および Linux 環境における標準的な InfiniBand の構成方法を説明します。

8.1 InfiniBand を介した Windows のブート

Windows には、サーバーの外部にあるブート可能なハード・ディスクを作成する固有の方法がありません。ここでは Paragon Partition Manager 6.0 を使用して、ロードされたハード・ディスクを DS4500 内の外部ハード・ディスクにコピーしました。複製が作成された後、ブレードのハード・ディスクを取り外し、InfiniBand を介してブレードを起動できます。

注 : Windows には、Service Pack 4 またはそれ以上が適用されていなければなりません。

1. ハード・ディスクを正常に構成し、オペレーティング・システムに接続した後、Paragon Partition Manager を起動します。
2. 「**Hard Disk**」→「**Copy hard disk**」を選択します。

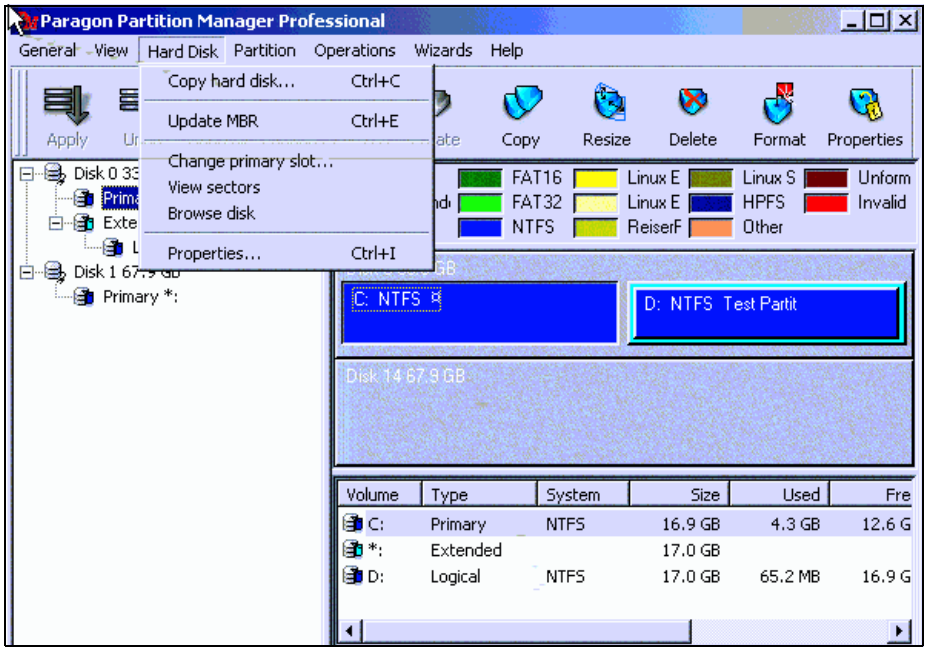

図 8-1 Paragon Partition Manager Professional

3. 「Copy Hdd」ウィンドウ（図 8-2）で、宛先ディスクを選択し、「**OK**」をクリックします。

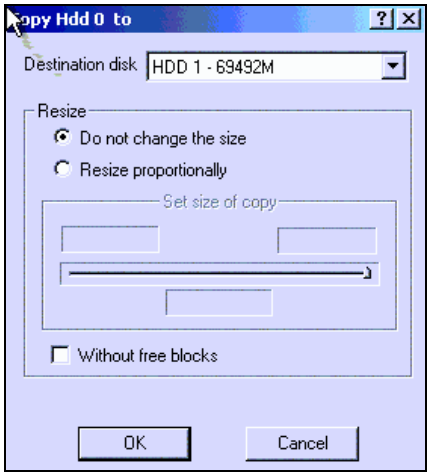

図 8-2 ハード・ディスクのコピー

4. 「**Apply**」をクリッし、再起動します。これによりリブートします。コピーした後、ブレードの電源を遮断します。
5. Element Manager を始動し、「**FibreChannel**」→「**Storage Manager**」を選択します。
6. 「**SRP Hosts**」を展開し、ホストを選択します。（ここでは、図 8-3 に表示されているように、Blade2 を選択しました。）
7. 「Boot Target WWPN」と「Boot FC LUN」をプルダウン・メニューから選択し、「**Apply**」をクリックします。
8. 「**LUN Access**」タブをクリックします。

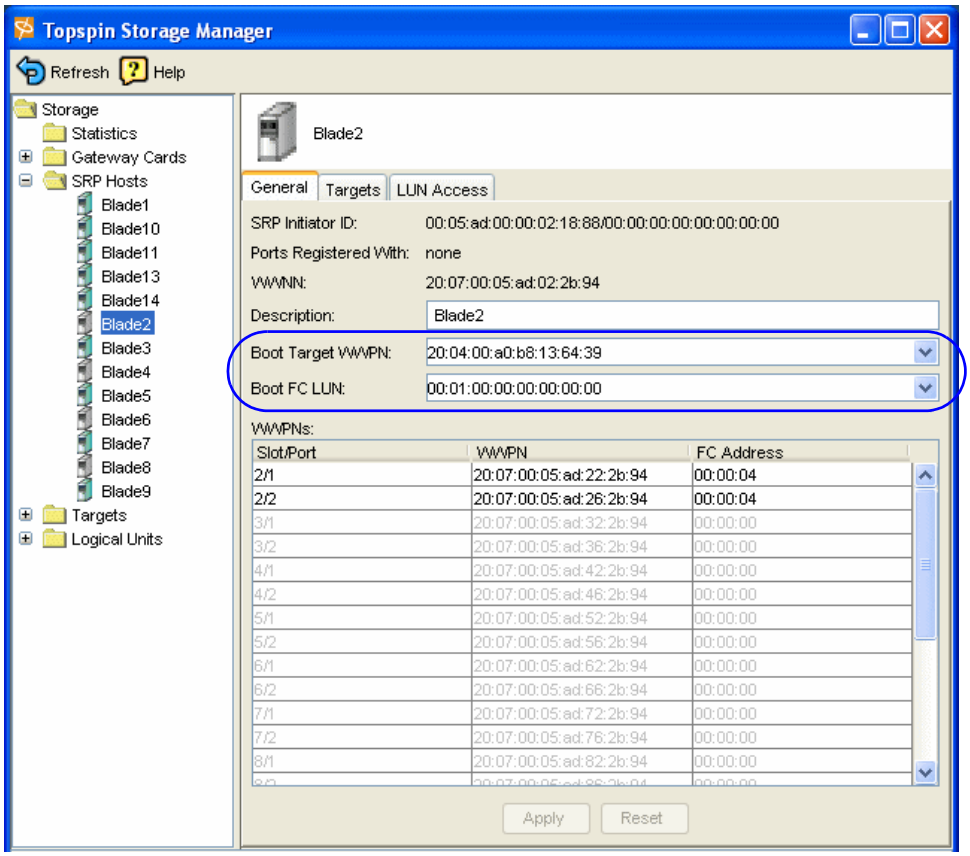

図 8-3　Topspin Storage Manager

9. SAN HDD を選択します。
10. 「**Edit ITL Properties**」をクリックします。

11. 図 8-4 に表示されているウィンドウで、SRP LUN ID が 00:00:00:00:00:00:00:00 であることを確認します。Windows のブートにはこの ID が必要です。Windows は 00 LUN から起動し、他の LUN からは起動しません。

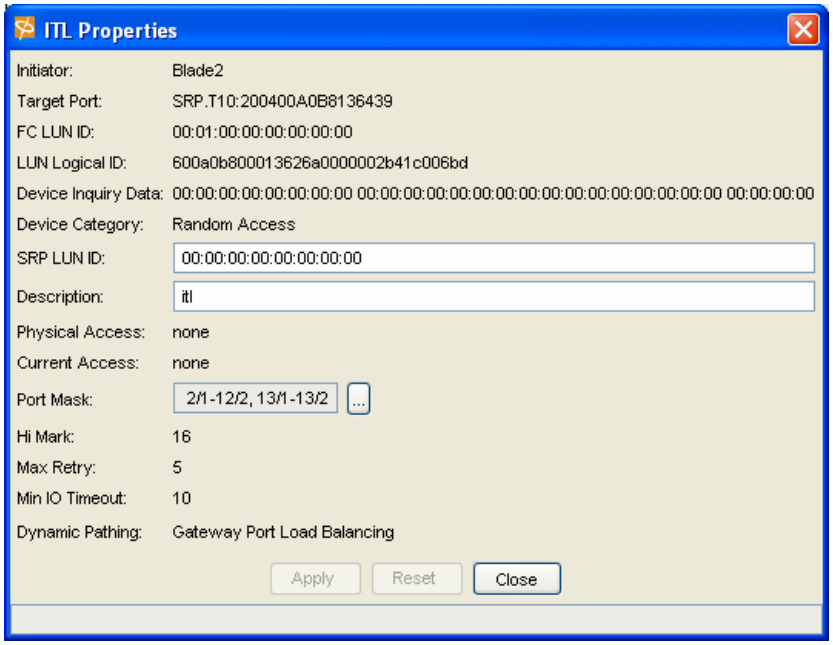

図 8-4 ITL Properties

12. ブレードを起動します。SAN Disk がブートオフされます。

8.2 InfiniBand を介した Linux のブート

InfiniBand を介してブートするように Linux を構成できます。これを行うには、外部ハード・ディスクをブレードに接続し、そのハード・ディスクにオペレーティング・システムをインストールします。Windows とは異なり、Linux は、リモート・ハード・ディスクに直接インストールできます。次の手順を実行してください。

1. ストレージ・アレイを接続するために、87 ページの 7.5、『InfiniBand スイッチ上のファームウェアの更新』および 99 ページの 7.7、『Element Manager のセットアップ』の手順を実行したことを確認する。
2. Linux CD を使用してブレードをブートする。この例では、Red Hat Linux Advanced Server 3.0 を使用します。
   a. プロンプトで、`linux dd askmethod` を入力する。`askmethod` コマンドを使用すると、ネットワーク・インストールが開始します。
   b. プロンプトで「Do you have a disk driver」と聞かれたら、「**Yes**」を選択する。

      この例では、`BoIB 2_4_21_20_EL` を使用しました。
   c. ドライバーのソースを選択するように要求があればそのソースを選択します。この例では、CD の `scd0` を使用しました。
   d. ドライバーのインストールが完了したら、「**No**」を選択し、Linux のロードを続行する。

8.3 Linux への外部ハード・ディスクのマウント

InfiniBand を介してハード・ディスクをマウントするように Linux を構成できます。これを実行するには、InfiniBand ファブリックを通じて外付けハード・ディスクをブレードに接続し、Linux オペレーティング・システムで定義します。

1. 87 ページの 7.5、『InfiniBand スイッチ上のファームウェアの更新』および 99 ページの 7.7、『Element Manager のセットアップ』の手順を実行して、ストレージ・アレイを接続します。
2. 例 8-1 の手順を実行して、外部ハード・ディスクを Linux にマウントします。（太字の単語はユーザーが入力する項目です。）

例 8-1 外部ハード・ディスクのマウント例

```
[root@localhost root]# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF disklabel
Building a new DOS disklabel. Changes will remain in memory only,
until you decide to write them. After that, of course, the previous
content won't be recoverable.


The number of cylinders for this disk is set to 8859.
There is nothing wrong with that, but this is larger than 1024,
and could in certain setups cause problems with:
1Åj software that runs at boot timeÅie.g., old versions of LILOÅj
2Åj booting and partitioning software from other OSs
  Åie.g., DOS FDISK, OS/2 FDISKÅj
Warning: invalid flag 0x0000 of partition table 4 will be corrected by wÅiriteÅj

CommandÅim for helpÅj: p

Disk /dev/sdb: 72.8 GB, 72870526976 bytes
255 heads, 63 sectors/track, 8859 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

   Device Boot    Start       End    Blocks   Id  System

CommandÅim for helpÅj: n
Command action
   e   extended
   p   primary partitionÅi1-4Åj
p
Partition numberÅi1-4Åj: 1
First cylinderÅi1-8859, default 1Åj:
Using default value 1
Last cylinder or +size or +sizeM or +sizeKÅi1-8859, default 8859Åj: 1246

CommandÅim for helpÅj: n
Command action
   e   extended
   p   primary partitionÅi1-4Åj
p
Partition numberÅi1-4Åj: 2
First cylinderÅi1247-8859, default 1247Åj:
Using default value 1247
Last cylinder or +size or +sizeM or +sizeKÅi1247-8859, default 8859Åj: +10240M

CommandÅim for helpÅj: p

Disk /dev/sdb: 72.8 GB, 72870526976 bytes
```

```
255 heads, 63 sectors/track, 8859 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

   Device Boot    Start       End    Blocks   Id  System
/dev/sdb1             1      1246  10008463+  83  Linux
/dev/sdb2          1247      2492  10008495   83  Linux

CommandÅim for helpÅj: w
The partition table has been altered!

Calling ioctlÅiÅj to re-read partition table.
Syncing disks.
[root@Blade11 root]# mkfs /dev/sdb2
mke2fs 1.32Åi09-Nov-2002Åj
Filesystem label=
OS type: Linux
Block size=4096Åilog=2Åj
Fragment size=4096Åilog=2Åj
1251712 inodes, 2502123 blocks
125106 blocksÅi5.00%Åj reserved for the super user
First data block=0
77 block groups
32768 blocks per group, 32768 fragments per group
16256 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632

Writing inode tables: done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 27 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
[root@Blade11 root]# ls /
bin   dev  home    lib    lost+found  mnt  proc  sbin      tftpboot  usr
boot  etc  initrd  lib64  misc        opt  root  srpdisk1  tmp       var
[root@Blade11 root]# mount /dev/sdb2 /srpdisk1
[root@Blade11 root]# df
Filesystem           1K-blocks      Used Available Use% Mounted on
/dev/sda2             32890776   2658840  28561176   9% /
/dev/sda1               101089     13846     82024  15% /boot
none                    504616         0    504616   0% /dev/shm

/dev/sdb2             60191008        20  57133420   1% /srpdisk1
```

8.4 IP over InfiniBand の構成と 2 つの BladeCenter シャーシの接続

IBM @server BladeCenter スイッチ用の Topspin InfiniBand ドライバーを使用すると、InfiniBand ネットワーク上で IP トラフィックを流すことができます。

8.4.1 Windows 2000 を実行するブレード・サーバーへのドライバーのインストール

Windows を実行するブレード・サーバーに IPoIB ドライバーをインストールするには、次の手順を実行します。

1. ブレード・サーバーをシャットダウンし、HCA をインストールし、ブレードを再起動します。旧ドライバーがすでにインストールされている場合は、現行のドライバーをアンインストールしてから、ブレード・サーバーをシャットダウンします。

   「スタート」→「プログラム」→「**Topspin InfiniBand SDK**」→「**Uninstall**」をクリックします。

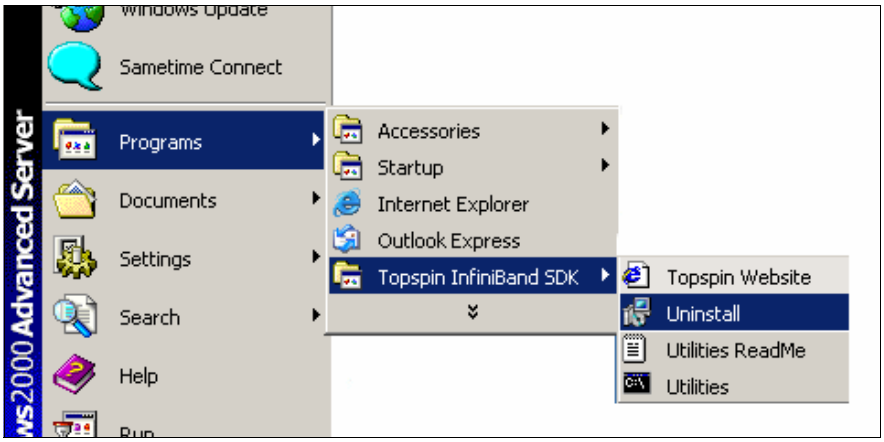

図 8-5 Uninstall

2. Topspin Uninstall ユーティリティーが開きます（図 8-6）。「**Next**」をクリックして、旧 InfiniBand ドライバーを除去します。

図 8-6　Topspin InfiniBand Product Uninstall Setup

3. 「Removal Complete」ウィンドウ（図 8-7）が表示されたら、「**Finish**」をクリックします。

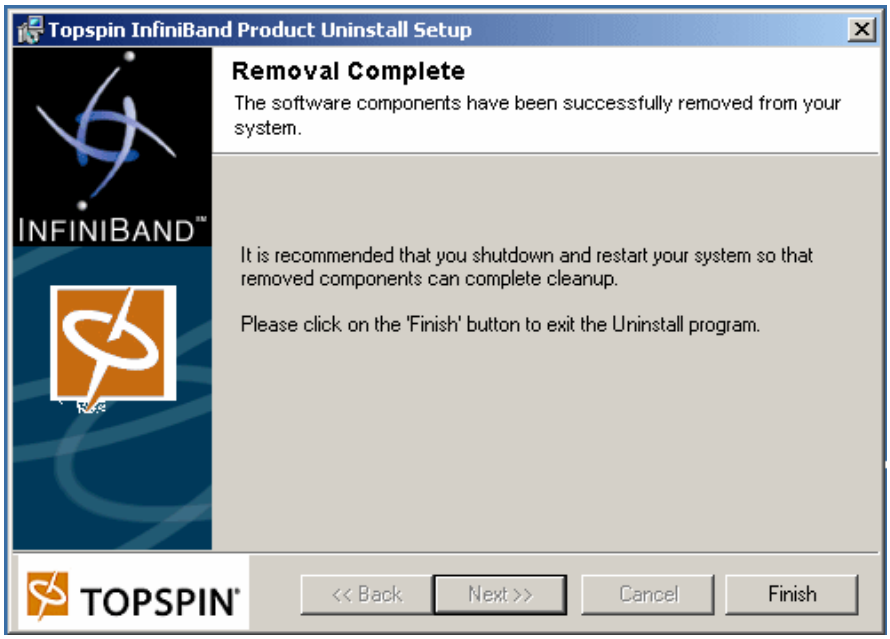

図 8-7　Topspin InfiniBand Product Uninstall Setup

4. ブレード・サーバーを再起動します。Windows は新しいハードウェアを検出し、ドライバーのインストールを試みます。このウィザードはドライバーを正しくインストールしないので、「**Cancel**」をクリックします。

図 8-8　Found New Hardware Wizard

5. ドライバー実行可能ファイルをダブルクリックし、ウィザードの指示に従います。

図 8-9　Topspin InfiniBand Product Install ウィンドウ

6.「**Finish**」をクリックして、ブレード・サーバーを再起動します。

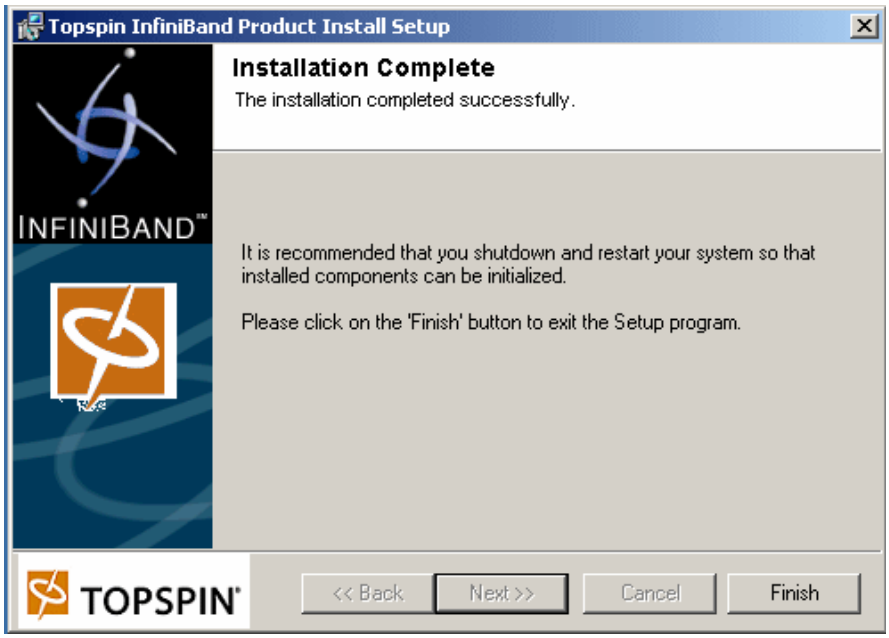

図 8-10　Installation Complete（この時点でリブート）

7. ブレード・サーバーが再起動した後、「Network and Dial-up Connections」に 2 つの InfiniBand ホスト・アダプター・ポートが表示されます（図 8-11）。

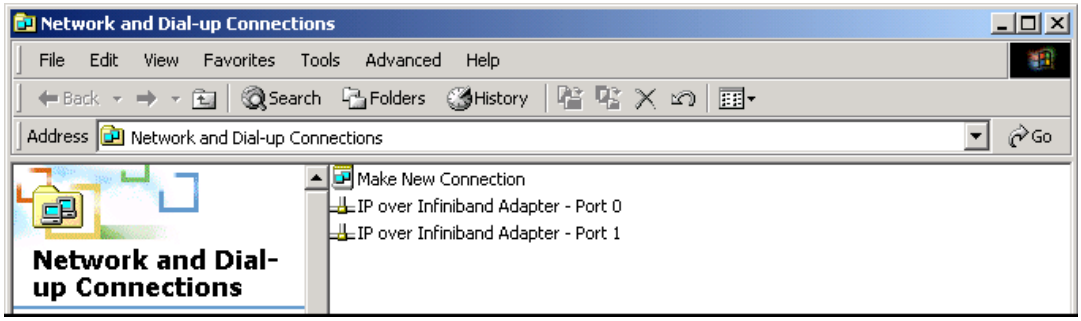

図 8-11　Network and Dial-up Connections（ネットワーク接続内の InfiniBand ホスト）

8.4.2 Linux を実行するブレード・サーバーへのドライバーのインストール

Linux を実行するブレード・サーバーに IPoIB ドライバーをインストールするには、次の手順を実行します。

1. 次の端末コマンドを使用して、ドライバーがブレードにインストールされているかどうかを確認する。

```
rpm -qa | grep topspin
```

```
[root@Blade11 root]# rpm -qa | grep topspin
topspin-ib-mod-rhel3-2.4.21-20.EL-3.0.0-126
topspin-ib-rhel3-3.0.0-126
```

図 8-12　ドライバーの確認

2. 旧ドライバーがすでにインストールされている場合は、現行のドライバーをアンインストールする。

```
rpm -e {driver name}
```

```
[root@Blade11 root]# uname -a
Linux Blade11 2.4.21-20.EL #1 SMP Wed Aug 18 20:34:58 EDT 2004 x86_64 x86_64 x86_64 GNU/Linux
```

図 8-13　ドライバーの削除

3. 基本ドライバー・サポートをインストールする。

   ```
   rpm -ihv {driver name}
   ```

   Intel 32 ビット・プロセッサーの場合、*.i686.rpm を使用します

   Intel EMT 64 ビット・プロセッサーの場合、*.x86_64.rpm を使用します

```
[root@Blade11 TS-IB-RHAS3]# rpm -ihv topspin-ib-rhel3-3.0.0-126.x86_64.rpm
Preparing...                ########################################### [100%]
   1:topspin-ib-rhel3       ########################################### [100%]
[root@Blade11 TS-IB-RHAS3]#
```

図 8-14　基本ドライバーのインストール

4. 特定のカーネル・サポートをインストールする。

   ```
   rpm -ihv topspin-ib-mod{linux version}{kernel version}
   ```

   Intel 32 ビット・プロセッサーの場合、*.i686.rpm を使用します

   Intel EMT 64 ビット・プロセッサーの場合、*.x86_64.rpm を使用します

```
[root@Blade11 redhat]# rpm -ihv topspin-ib-mod-rhel3-2.4.21-20.EL-3.0.0-126.x86_64.rpm
Preparing...                ########################################### [100%]
   1:topspin-ib-mod-rhel3-2.########################################### [100%]
```

図 8-15　カーネル・サポートのインストール

注 : 誤ったドライバー名を入力すると、「Failed Dependencies」エラーと推奨される解決法が表示されます（図 8-16）。

```
[root@Blade11 TS-IB-RHAS3]# rpm -ihv topspin-ib-mod-rhel3-2.4.21-20.ELsmp-3.0.0-126.x86_64.rpm
error: Failed dependencies:
        kernel-smp = 2.4.21-20.EL is needed by topspin-ib-mod-rhel3-2.4.21-20.ELsmp-3.0.0-126
    Suggested resolutions:
        kernel-smp-2.4.21-20.EL.x86_64.rpm
[root@Blade11 TS-IB-RHAS3]#
```

図 8-16　Failed dependencies

8.4.3 Windows を実行するブレード上の InfiniBand ポートへの IP アドレスの割り当て

ここでは、Windows 環境で InfiniBand ポートに IP アドレスを割り当てる方法を説明します。

1. 「ネットワーク接続」を開きます。
2. 最初の InfiniBand ポートをダブルクリックして、「Port Properties」ウィンドウを開く（図 8-17）。「Components」の下で、「**Internet Protocol（TCP/IP）**」を強調表示し、「**Properties**」をクリックします。

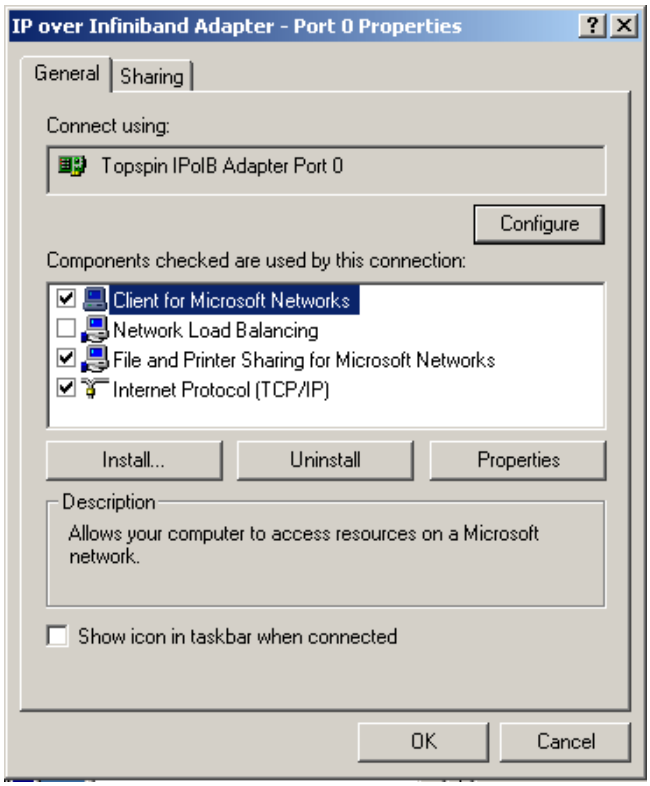

図 8-17 IP over InfiniBand Adapter - Port 0 Properties

3. 「Internet Protocol properties」ウィンドウが開きます（図 8-18）。「Use the following IP address」の横にあるラジオ・ボタンを選択します。
4. InfiniBand ポートの IP アドレスとサブネット・マスクを入力します。「**OK**」をクリックします。

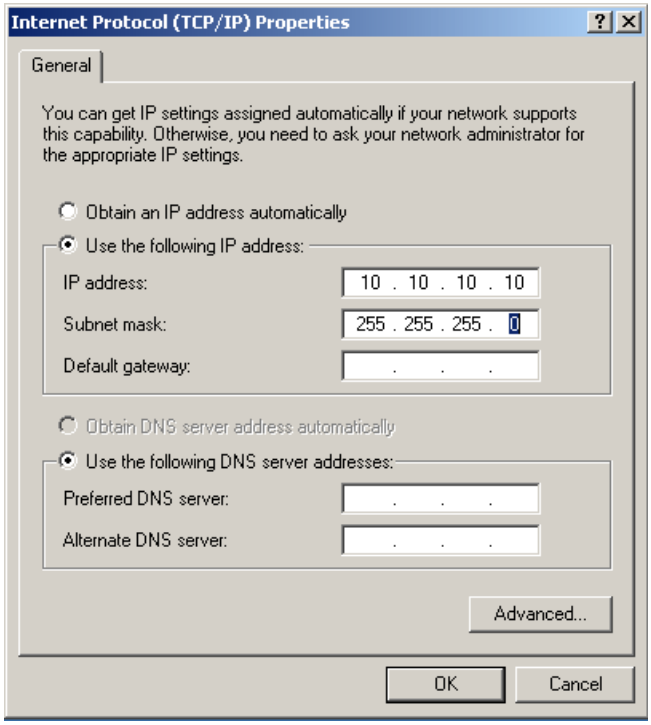

図 8-18　Internet Protocol（TCP/IP）Properties

5. 「Connection Properties」ウィンドウで「**OK**」をクリックします。
6. 2 番目の InfiniBand ポートに対して 125 ページの ステップ 1 から 5 を繰り返します。

8.4.4 Linux を実行するブレード上の InfiniBand ポートへの IP アドレスの割り当て

ここでは、Linux 環境で InfiniBand ポートに IP アドレスを割り当てる方法を示します。

1. 端末ウィンドウを開きます。
2. コマンド **`ifconfig ib0`** xx.xx.xx.xx **`netmask`** yy.yy.yy.yy を入力します。

   xx.xx.xx.xx はポートの IP アドレスであり、yy.yy.yy.yy はポートのサブネット・マスクです。
3. ポートを使用可能にするために、コマンド **`ifconfig ib0 up`** を入力します。
4. ib1 に対してステップ 2 と 3 を繰り返します。
5. Linux システムが再起動するときに、ポートが自動的に再起動することはありません。Linux がリブートするたびにポートを使用可能にするには、システムが始動するたびに始動するようにポートを構成する必要があります。ネットワーク構成ウィンドウを開くには、メインメニュー・アイコン →「**System Settings**」→「**Network**」をクリックします。

6. 図 8-19 に表示されているウィンドウが開きます。InfiniBand ポートをダブルクリックして、装置構成ウィンドウを開きます。

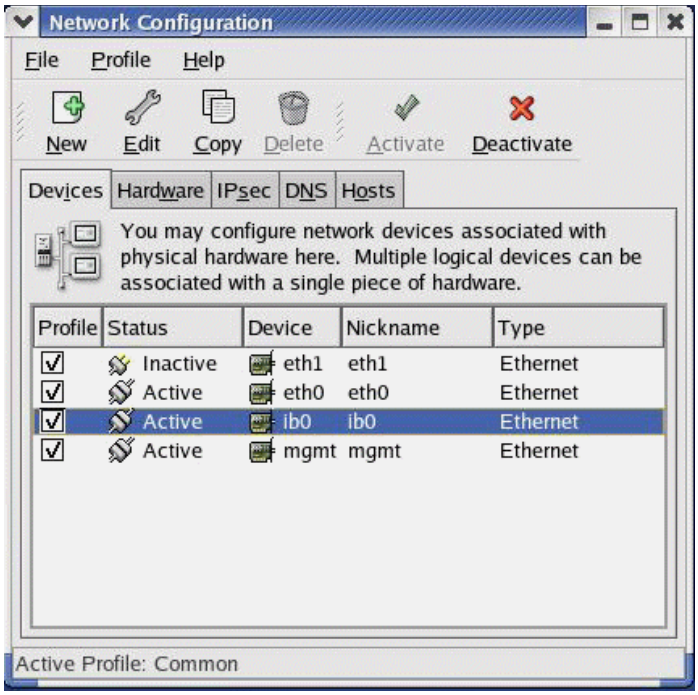

図 8-19 Network Configuration ウィンドウ

7. 図 8-20 のようなウィンドウが表示されます。「**Activate device when computer starts**」のチェック・ボックスを選択します。「**OK**」をクリックして、ウィンドウを閉じ、設定を保管します。

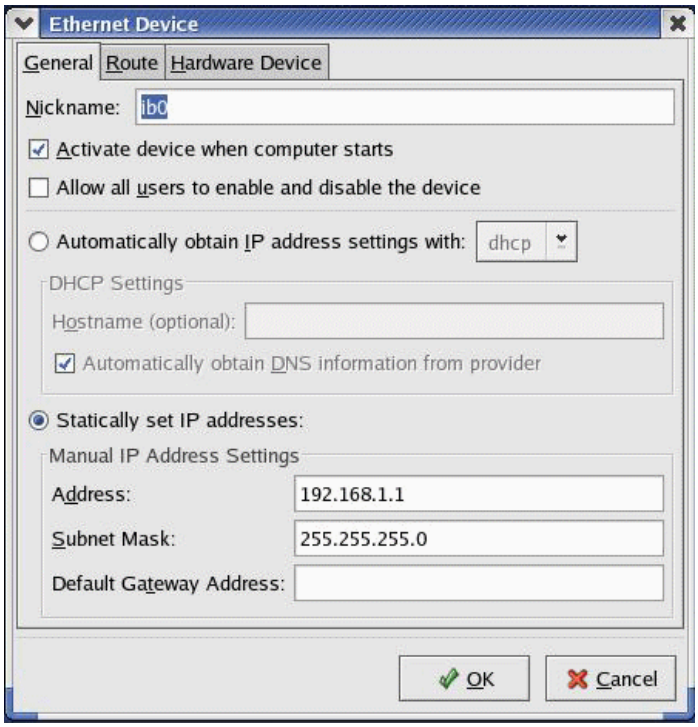

図 8-20 Ethernet Device ウィンドウ

8.4.5 InfiniBand スイッチ相互の接続（直接または外部スイッチ経由）

IBM @server BladeCenter InfiniBand スイッチを使用して直接、または Topspin 外部 InfiniBand スイッチを使用して、2 つの BladeCenter シャーシを接続することができます。

InfiniBand スイッチの直接接続

2 つの IBM @server BladeCenter InfiniBand スイッチを接続するには、4x ケーブルを使用して 4x ポートを接続するか、12x ケーブルを使用して 12x ポートを接続するか、オクトパス・ケーブルを使用して、一方の BladeCenter スイッチの 12x ポート内の 4x 接続の 1 つを、もう一方の BladeCenter InfiniBand スイッチ内の 4x ポートに接続します。

トポロジー・ビュー（図 8-21）は、2 つの BladeCenter InfiniBand スイッチが直接接続されている様子を示しています。2 番目の InfiniBand スイッチは外部スイッチに接続されます。

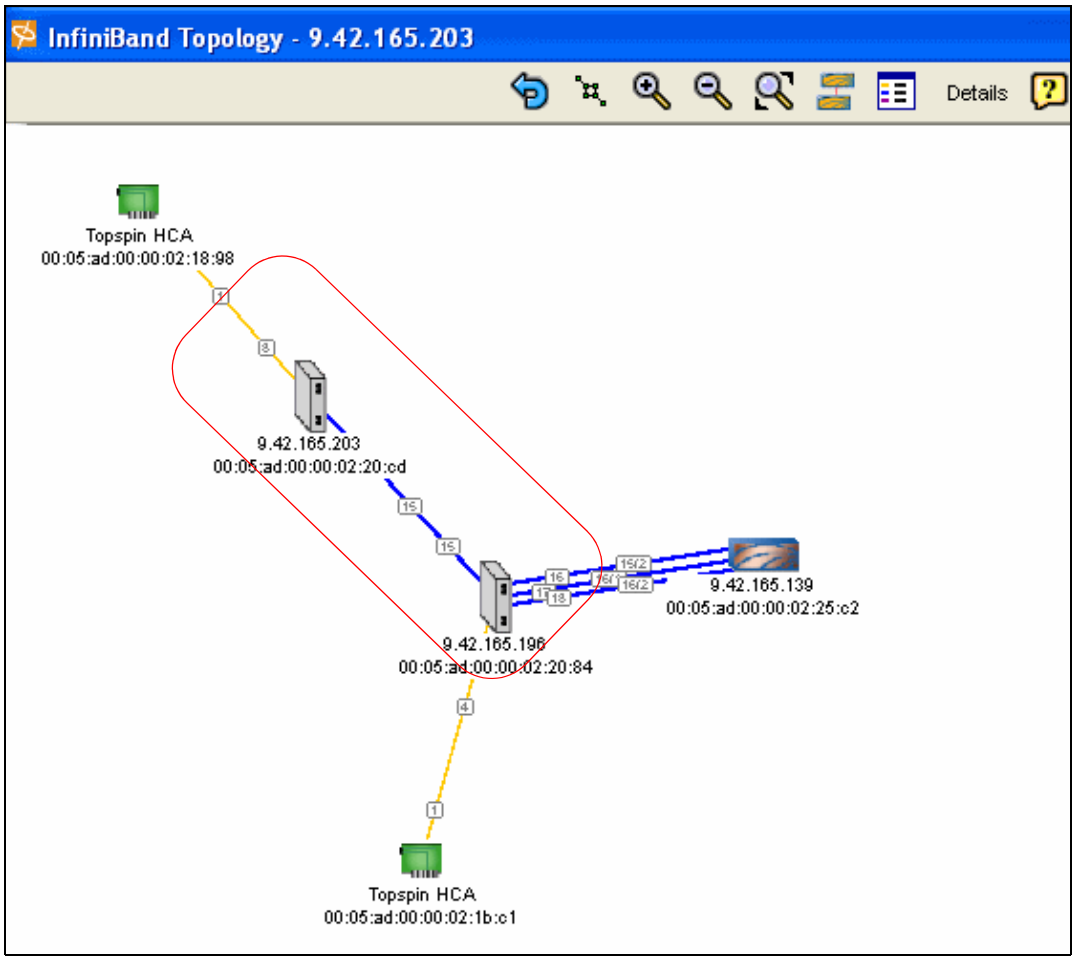

図 8-21　Element Manager: 直接 InfiniBand 接続

外部スイッチを使用した接続

外部 Topspin InfiniBand スイッチを使用して 2 つの BladeCenter シャーシを接続するには、4x ケーブルを使用して、BladeCenter InfiniBand スイッチの 4x ポートを外部スイッチの 4x ポートの 1 つに接続するか、オクトパス・ケーブルを使用して、BladeCenter スイッチの 12x ポート内の 4x 接続の 1 つを、外部スイッチの 4x ポートの 1 つに接続します。

トポロジー・ビュー（図 8-22）は、2 つの BladeCenter InfiniBand スイッチが外部 InfiniBand スイッチを経由して接続されている様子を示しています。

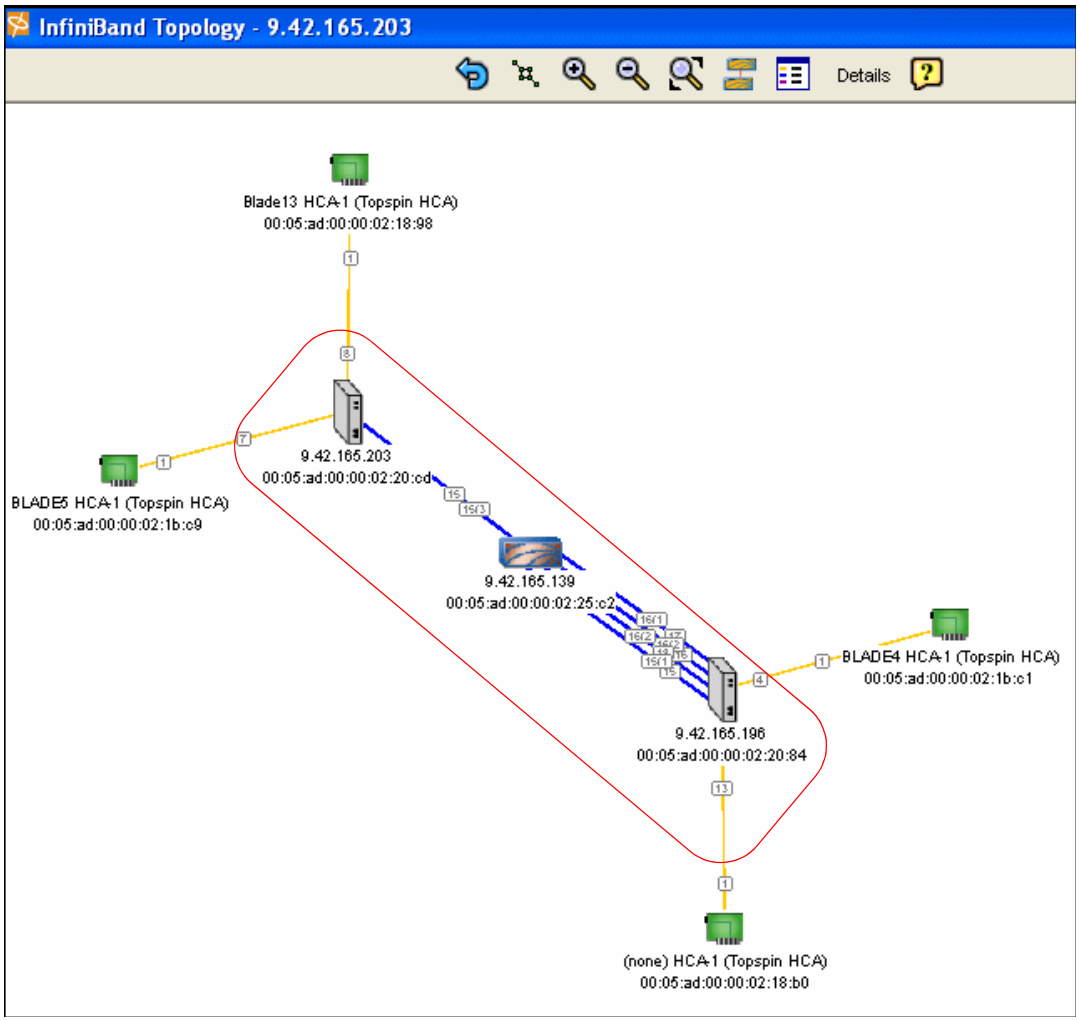

図 8-22　Element Manager: クラスター・トポロジー・ビュー

8.4.6 シャーシ 2 のブレードからのシャーシ 1 のブレードの Ping

Windows コマンド・プロンプトまたは Linux 端末を開きます。最初のシャーシ内のブレードの 1 つに、2 番目のシャーシ内のブレードから Ping コマンドを実行します。

8.4.7 Topspin HCA 拡張カードのプロトコル構成

ホスト・ドライバーをインストールすると、インストール・プロセスにより、使用可能なすべてのドライバーがホスト上にインストールされます。この Redpaper のリリース時点で、HCA 拡張カードは、Linux ホスト上のすべてのプロトコルをサポートし、Windows ホスト上の IPoIB および SRP をサポートします。

8.5 InfiniBand とイーサネットの接続

外部スイッチでイーサネット・ゲートウェイを構成すると、InfiniBand ドーター・カードの有無にかかわらず、ブレード間で IP トラフィックを流すことができます。

8.5.1 外部 InfiniBand スイッチへのイーサネット・ゲートウェイの取り付け

ここでは、外部 InfiniBand スイッチにイーサネット・ゲートウェイを取り付け、BladeCenter シャーシのベイ 1 またはベイ 2 にイーサネット・スイッチを取り付ける方法を説明します。

1. 標準のイーサネット・ケーブルを使用して、外部スイッチのイーサネット・ゲートウェイを、BladeCenter シャーシ内のイーサネット・スイッチに接続します。
2. Element Manager を使用して、外部 InfiniBand スイッチにログオンします。アクティブなリンクはすべて、緑色に強調表示されます。

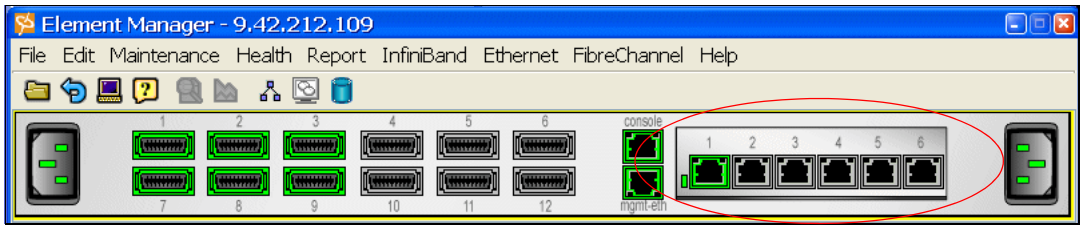

図 8-23　90 の最初のビュー

3. イーサネット・ゲートウェイを表す GUI の部分を右クリックし、「**Properties**」を選択します。

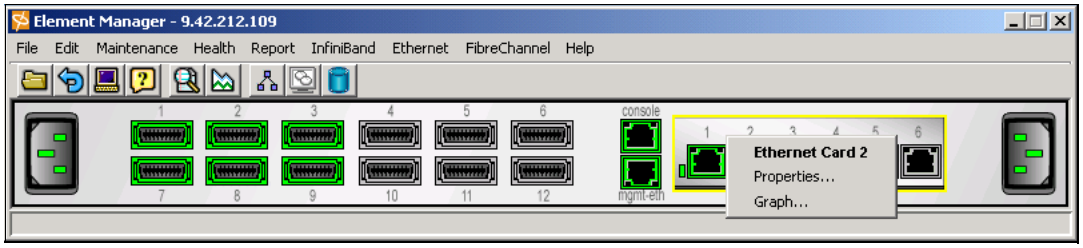

図 8-24　イーサネット・ゲートウェイ

4. 現行のカード状況が **up** であり、稼働状況が **normal** であることを確認します。

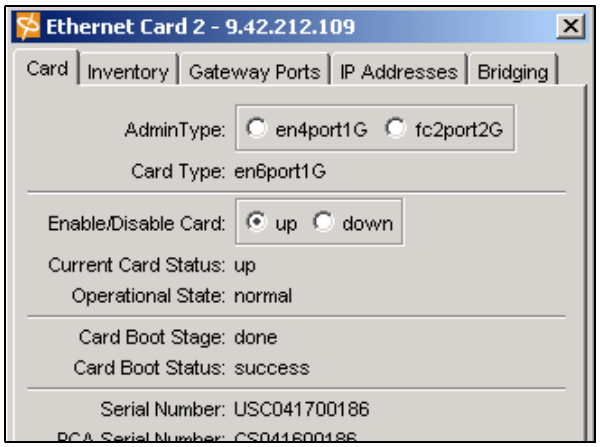

図 8-25　ゲートウェイのプロパティー

8.5.2 イーサネット・ゲートウェイ上のブリッジ・グループの構成

ここでは、イーサネット・ゲートウェイでブリッジ・グループを構成します。これにより、イーサネット・ファブリック上の装置は、InfiniBand ファブリック上の装置と通信できます。

1. Element Manager で、「**Ethernet**」→「**Bridging**」を選択します（図 8-26）。

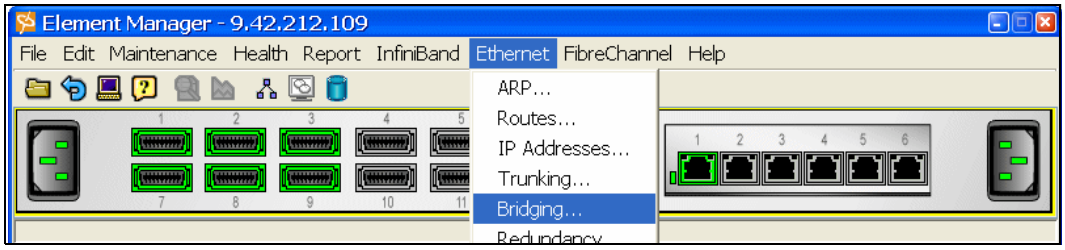

図 8-26 Bridging

2. ゲートウェイで構成しているブリッジ・グループを表示するウィンドウが開きます（図 8-27）。ブリッジ・グループを選択し、「**Edit**」をクリックします。

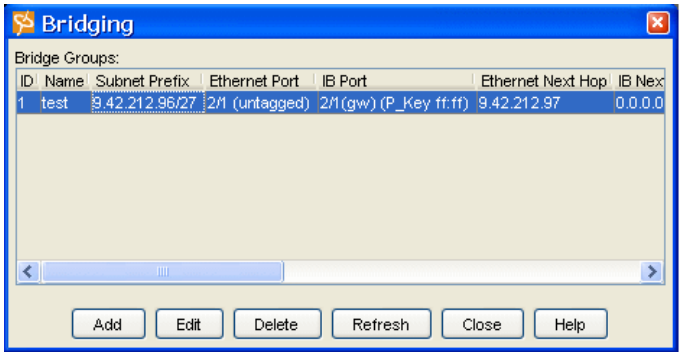

図 8-27 ブリッジ・グループ

3. 選択したブリッジ・グループのパラメーターを示すウィンドウが開きます。イーサネット・ファブリックと InfiniBand ファブリックを接続するサブネットの新しいパラメーターを入力します。「**OK**」をクリックします。

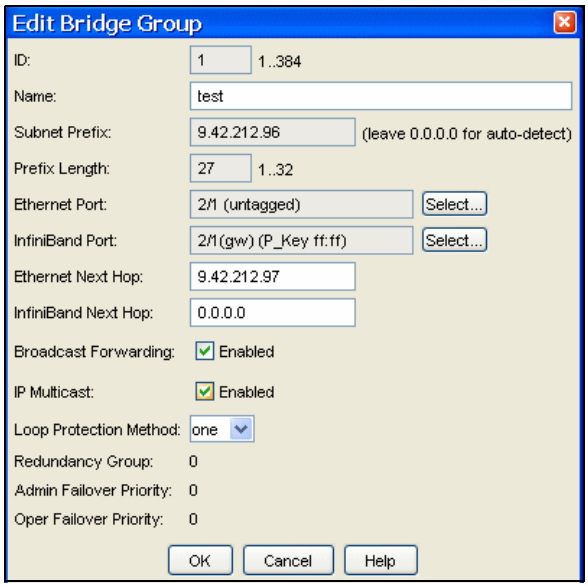

図 8-28　Edit Bridge Group

8.5.3 イーサネット・ファブリックと InfiniBand ファブリック間の接続のテスト

InfiniBand HCA のないブレードから、InfiniBand HCA のあるブレードを ping コマンドを実行します。

注 : ゲートウェイに接続されているイーサネット・スイッチに関連したオンボード NIC を使用可能にしてください。

関連資料

ここに掲載する資料は、このレッドブックで取り上げている問題をさらに詳しく検討するのに特に適していると考えられる資料です。

IBM Redbooks

これらの資料の注文方法については、『IBM Redbook の入手方法』（136 ページ）を参照してください。

- *The Cutting Edge: IBM @server* BladeCenter, REDP-3581

  http://www.redbooks.ibm.com/redpapers/abstracts/redp3581.html

その他の資料

次の Topspin の資料も関連の情報源として利用できます。

- *InfiniBand Host Channel Adapter Expansion Card for BladeCenter User Guide*
- *InfiniBand User Guide*
- *InfiniBand Switch Module for BladeCenter User Guide*
- *Site Deployment Planning Guide Multi-Fabric I/O and VFrame*
- *Configuring IBM BladeCenter for VFrame*

オンライン・リソース

次の Web サイトも関連の情報源として役立ちます。

- BladeCenter home page

  http://www.ibm.com/servers/eserver/bladecenter/index.html
- Storage for xSeries home page

  http://www.pc.ibm.com/us/eserver/xseries/storage.html
- Systems management for xSeries and BladeCenter

  http://www.ibm.com/servers/eserver/xseries/systems_management/xseries_sm.html
- IBM Tivoli Intelligent Orchestrator home page

  http://www.ibm.com/software/tivoli/products/intell-orch/
- IBM TotalStorage DS4300 home page

  http://www.ibm.com/servers/storage/disk/ds4000/ds4300/index.html
- IBM TotalStorage DS4400 home page

  http://www.ibm.com/servers/storage/disk/ds4000/ds4400/index.html
- IBM TotalStorage DS4500 home page

  http://www.ibm.com/servers/storage/disk/ds4000/ds4500/index.html

- IBM TotalStorage Product Guide

  ftp://ftp.software.ibm.com/common/ssi/rep_sp/n/TSB00364USEN/TSB00364USEN.PDF

- The Programmable Server Switch

  http://www.topspin.com/solutions/index.html

- Support for BladeCenter chassis

  http://www.ibm.com/servers/eserver/support/bladecenter/chassis/downloadinghwonly.html

IBM Redbook の入手方法（英語版のみ）

次の Web サイトでは、Redbook、Redpaper、ヒント、ドラフト資料、追加資料などを、検索、表示、またはダウンロードできます。

ibm.com/redbooks

IBM のヘルプ

IBM サポートおよびダウンロード

ibm.com/support

IBM グローバル・サービス

ibm.com/services

索引

J

K

L

M

N

O

P

Q

R

S

T

IBM @server BladeCenter および Topspin InfiniBand スイッチ・テクノロジー

InfiniBand スイッチ機能を追加する eServer BladeCenter

スループットの向上とコストの削減に役立つ構成

Topspin VFrame および Tivoli Intelligent Orchestrator のデモンストレーション

BladeCenter 用の Topspin ソリューションは、IBM @server BladeCenter シャーシとの 80 GB 接続、RDMA（Remote Direct Memory Access）、および単一の I/O ファブリック上でのシャーシからのクラスタリング、LAN、および SAN トラフィックの統合機能を提供して、スループットの増加とコストの削減を可能にします。

Topspin InfiniBand ソリューションの基盤を整えるために、この Redpaper では、InfiniBand、BladeCenter、および IBM TotalStorage のテクノロジーについて記述します。その後、Topspin InfiniBand のアーキテクチャー、スイッチ・モジュール、およびホスト・チャネル・アダプター・カード、ならびに Element Manager と Chassis Manager の使用について詳しく説明します。

この Redpaper は、Topspin InfiniBand ソリューション・コンポーネントを使用した IBM eServer BladeCenter の複数の構成について詳しく説明します。本書は、お客様が HPC または大規模エンタープライズ環境に固有の BladeCenter InfiniBand ソリューションを構築する基盤になります。

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

実際的経験に基づく技術情報の作成

IBM Redbook は IBM International Technical Support Organization によって作成されます。世界中の IBM、お客様、およびパートナーの専門家が、現実的なシナリオに基づいてタイムリーな技術情報を作成します。ご使用の環境に IT ソリューションを効果的にインプリメントするのに役立つ具体的な推奨事項を提供します。

詳細情報の参照先：
ibm.com/redbooks

SG88-8548-00