

# 仮想化に最適なプラットフォーム PRIMERGY BX900ブレードサーバ

## New Blade Server PRIMERGY BX900 Suitable for Server Virtualization

### あらまし

現在、IT業界においては、サーバ統合、省電力化を推進する重要な技術として仮想化機能が広まりつつある。仮想化には、サーバの仮想化のみならず、ネットワークの仮想化やストレージの仮想化などITシステムの主な構成要素が含まれている。

近年サーバ分野で急速に普及しつつあるブレードサーバは、これらITシステム構成要素を統合した製品であり、仮想化のメリットを最も発揮しやすい。

本稿では、最新のテクノロジーを採用した新ブレードサーバPRIMERGY BX900における、仮想化の先進的取組みについて、それを支える高速伝送技術や独自開発高速スイッチ技術などとともに紹介する。

### Abstract

Recently, the use of virtualization technology has been increasing in the IT industry to realize server consolidation and energy saving. Virtualization includes not only server virtualization but also network virtualization and storage virtualization. A blade server is a new integrated IT system that includes these IT components, and it is the best way to realize the benefits of virtualization. The new blade server PRIMERGY BX900 is based on Fujitsu's new proprietary technologies, such as very high-speed transfer technology and high-speed Ethernet switch technology. With these advanced basic technologies, PRIMERGY BX900 provides customers with the great benefits of server virtualization, network virtualization and I/O virtualization.



松原正和（まつばら まさかず）  
IAサーバビジネス統括部 所属  
現在、PRIMERGYの製品企画業務に従事。



中川幸洋（なかがわ ゆきひろ）  
ITシステム研究所サーバテクノロジー  
研究部 所属  
現在、高性能インタコネクト関連の研究に従事。



牧 雄治郎（まき ゆうじろう）  
第一サーバ事業部ブレードサーバ技術部 所属  
現在、ブレードサーバの開発業務に従事。

## まえがき

従来、ブレードサーバはその構造に由来する高密度性から、企業に散在するサーバの集約を目的に利用されてきた。近年では、テクノロジーの発展に伴い、その適用範囲をサーバ集約のみならず、基幹システムやSaaS (Software as a Service), IaaS (Infrastructure as a Service) などの次世代ビジネスへと拡大する傾向にあり、データセンタへの導入が進んでいる。

一方で、データセンタにおけるブレードサーバに対する課題として、運用の効率化が挙げられる。具体的には、ブレードサーバ導入/運用負担の軽減、運用コストの削減である。また、近年のCPU高性能化に伴い、Hypervisorという仮想化技術を用い1台の物理サーバ上で複数の仮想サーバを運用する形態が一般化しつつあり、物理/仮想サーバ両面からの効率化が求められる。

これらの課題を解決するため、富士通はハードウェア、ソフトウェアの仮想化技術の強化に取り組んだ。その結果、データセンタでの課題を解決する新ブレードサーバPRIMERGY BX900を開発した。

本稿では、PRIMERGY BX900のシステム概要、仮想化に向けたネットワークおよびI/O仮想化を実現する技術について紹介する。

## システム概要

本章では、各種仮想化機能を有するPRIMERGY

BX900のシステム概要について説明する。

PRIMERGY BX900は、サーバブレード、ストレージブレード、スイッチブレード、それらを収納するシャーシから構成される (図-1)。

### ● サーバブレード

サーバブレード (PRIMERGY BX920 S1) は、高さ約5 U、幅約1 Uの細長い形状の筐体の中に、CPU、チップセット、メモリ、HDDやI/Oを実装した基板を配置し、サーバの役割を担う。

CPUおよびチップセットにはIntel社のXeonプラットフォームを採用し、仮想化環境への最適化を図った。Xeonプロセッサ5500番台のインテルバーチャライゼーションテクノロジーは、ハードウェアレベルで仮想化をアシストし、従来のソフトウェアによる仮想化と比較すると、処理能力が向上し、物理リソースネイティブに近いパフォーマンス特性を仮想サーバが引き出せる。さらに、BX920 S1は、ダイレクトI/Oインテルバーチャライゼーションテクノロジーにも対応しており、仮想マシンへのI/Oデバイスの直接割当てをサポートすることで、従来の仮想マシン運用時の問題であったI/O性能の劣化を抑制できる。

内蔵HDDとして高信頼なSAS HDDを最大2台搭載し、RAIDを構成可能である。メモリは9スロットを実装し、仮想化環境で要求される高メモリ容量に対応した。I/Oは標準で1ギガLANを4ポート備え、さらに拡張ボード (サーバブレード1台あたり最大2枚搭載可) により、1ギガLANを8ポートもしくは

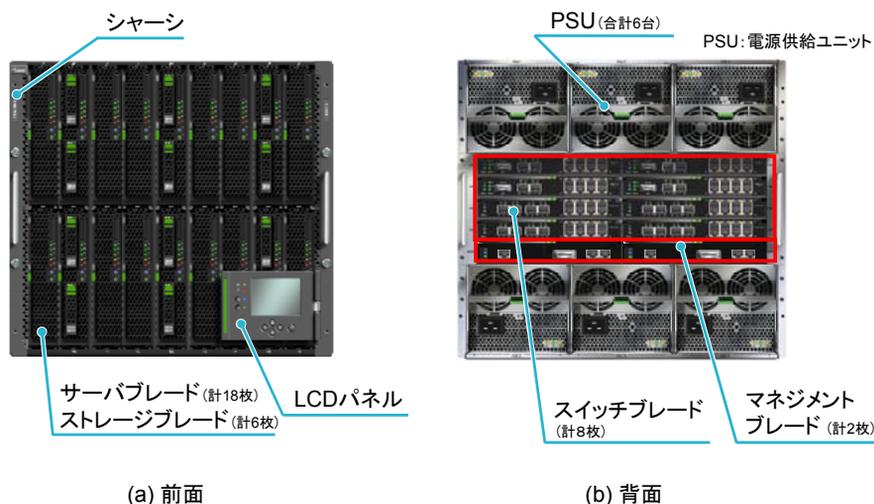


図-1 PRIMERGY BX900 ブレードサーバ  
Fig.1-PRIMERGY BX900 blade server.

10ギガLAN，8ギガファイバチャネルを4ポート拡張可能である。

## ● ストレージブレード

ストレージブレードは，サーバブレードと1対1で接続され，ストレージ大容量化およびデータのバックアップを実現する。ストレージブレードSX940 S1は2.5インチHDDを4台内蔵（RAID対応）し，BX920 S1のストレージ容量を拡張する。また，ストレージブレードSX910 S1はLTOなどのバックアップデバイスを搭載し，BX920 S1との組み合わせにより，バックアップサーバを形成する。

## ● スイッチブレード

スイッチブレードは，サーバブレードのI/Oをシャーシ外部のネットワーク/ファブリックに接続する役割を持つ。PRIMERGY BX900では3種類のスイッチブレードが当初用意される。1ギガスイッチブレード（ダウンリンクポート：1 Gbps，アップリンクポート：1 Gbpsもしくは10 Gbps），フル10ギガスイッチブレード（富士通独自開発のスイッチLSI MB86C69内蔵，アップリンクポート・ダウンリンクポートともに10 Gbps），および8ギガファイバチャネルスイッチブレードである。詳細は次章で説明するが，各スイッチブレードには，通常のスイッチとしてのモードと接続簡易化機能であるIBP（Intelligent Blade Panel）（LAN）とアクセスゲートウェイ（FC）に対応するモードがあり，管理者が選択することができる。これにより，ネットワーク管理者のみならず，サーバ管理者でもネットワーク構成を容易に構築できる手段を提供する。

## ● シャーシ

BX900 S1シャーシは，中央にミッドプレーンと呼ばれるPCB（Print Circuit Board）を配し，前方からサーバブレード，ストレージブレードを，背面から電源，スイッチブレードおよびマネジメントブレード（MMB）を収納する。ミッドプレーン上には高速シグナル伝送路が敷設されており，この伝送路によりサーバブレードとスイッチブレードを接続するケーブルレス構造となっている。伝送路は1本あたり片方向10 Gbpsの帯域幅を持ち，将来のインタコネクタ技術にも対応できる能力を有する。さらに，ミッドプレーンは電源モジュールから，すべてのモジュールに電力を分配する役割を担い，各モジュールは電源モジュールを共有する。電源モ

ジュールの冗長構成は自由に変更可能であり，ブレードシステムの構成により電源の稼働数を変動させることで，電源負荷をコントロールし電力効率が高い状態を維持し，省エネルギー化を図る。マネジメントブレードは，全モジュールのハードウェアの異常監視，設定，管理を行う。マネジメントブレードの操作は，従来のWeb/CLIに加え，BX900では新たにシャーシ備付のLCDパネルからも行える。これにより，データセンタでのオンサイト管理を可能とした。

## 仮想化に向けたネットワーク

### ● 技術背景と求められる要件

サーバの仮想化，CPUのマルチコア化の進展により，1台の物理サーバ上でより多くのOSやアプリケーションが実行されるため，高性能・高機能な10ギガイーサネットワークのインタフェースとスイッチが要求されている。また，データセンタ化と仮想化の進展により，DCE（Data Center Ethernet）などの堅ろう性やLAN/SANを統合したI/O統合，ネットワークを簡単かつ安全に設定運用するための機能などが要求されている。次節以降では，BX900の特に10ギガネットワークがこれらの要件に対応して，どのような機能や性能を備えているかについて述べる。

### ● 10ギガネットワーク

10ギガネットワークは，10ギガスイッチブレードと10ギガメザニンカード，そして，それらをつなぐミッドプレーンから構成される。ここでは，10ギガネットワークを実現するための核となる高性能，高密度，省電力な10ギガスイッチブレードについて述べる。

10ギガスイッチブレードは，以下の特長を持つ。

#### (1) 業界最速のスイッチ性能

業界最大の520 Gbpsのスループットと業界最短の300 nsのスイッチ遅延を実現した。520 Gbpsのスループットは，ダウンリンク18ポート，アップリンク8ポートの計26ポートすべてを収容可能とし，さらに300 nsの業界最小のスイッチ遅延により，ストレージ，高速トランザクション，HPC（High Performance Computing）などへの展開を可能とした。

(2) 業界最小サイズ

ミッドプレーンを介したスイッチブレードとメザニンカード間の10ギガイーサネット伝送は、従来4本の信号で実現していたが、今回、それを1本の信号、すなわち、10ギガシリアル伝送（IEEE 10GBASE-KR）を用いることにより、業界最小のスイッチブレードを実現した。最大8枚の10ギガスイッチブレードをBX900に搭載することを可能とした。

(3) 業界最小の消費電力

スイッチブレードを構成するスイッチLSIとして、自社開発の26ポート10ギガスイッチLSI MB86C69を採用した。MB86C69は、ミッドプレーンを10ギガシリアル伝送可能な高速I/Oマクロ<sup>(1)</sup>を搭載することにより、外付けSERDES（SERializer/DESerializer）チップを不要とし、他社に対して約半分となる、業界最小の消費電力30Wを実現した。

なお、この高速I/Oマクロは、光モジュールSFP+も直接ドライブ可能である。

(4) データセンタ向け機能

10ギガスイッチブレードは、データセンタ向けに、豊富な認証、QoS、冗長、アクセス制御機能を実現した。さらに、現在、標準化が進んでいるDCE機能をいち早く取り込み、ロスのないイーサネットを実現し、I/O統合を実現するFCoE（Fiber Channel over Ethernet）への対応を可能としている。

10ギガスイッチブレードの諸元を表-1に、写真を図-2に示す。

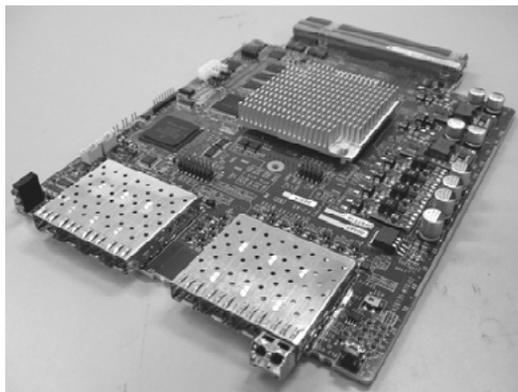


図-2 10ギガスイッチブレードSBAX2  
Fig.2-10Gbit Ethernet switch blade “SBAX2”.

10ギガスイッチブレードは、スイッチLSI MB86C69とスイッチソフトを実行するマイクロプロセッサで構成される。スイッチソフトは、SR-Sシリーズに搭載されている高度なセキュリティ機能を持つスイッチソフトをベースに開発し、スイッチブレード向けに最適化した。MB86C69は、20ポートのスイッチLSI MB8AA3020<sup>(2)</sup>をブレードサーバ向けに強化し、ポート数、バンド幅、データセンタ向け機能、高速インタフェース能力を大幅に向上させた。MB86C69のチップを図-3に示す。

表-1 10ギガスイッチブレードの諸元

項目	仕様
ダウンリンク	18ポート (10GBASE-KR)
アップリンク	8ポート (SFP+)
スイッチスループット	520 Gbps
スイッチ遅延	300 ns
消費電力	30 W (最大40 W)
物理サイズ	シングルワイド (幅: 193 mm, 奥行: 268 mm, 高さ: 28 mm)
モード	レイヤ2およびIBP (後述)
VLAN機能	ポート、タグ、プロトコル
冗長機能	STP (STP, MSTP, RSTP), リンクアグリゲーション, バックアップポート, リンクダウンリレー
QoS	IEEE802.1p (COS) /TOS (IP Precedence) /DSCP/ACL (IPv4, IPv6), Strict/DRR
ネットワーク認証	MAC, IEEE802.1X, Web, RADIUS, TACACS+
アクセス制御	レイヤ2~4
マルチキャスト	IGMPスヌープ機能
モニタリング	ポートミラーリング
ネットワーク管理	CLI, Web UI, SNMP/RMON, LLDP, ログイン

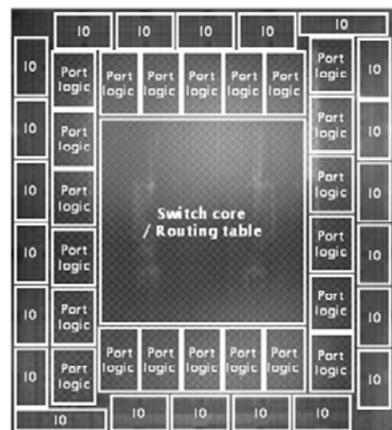


図-3 スイッチLSI MB86C69  
Fig.3-Switch LSI “MB86C69”.

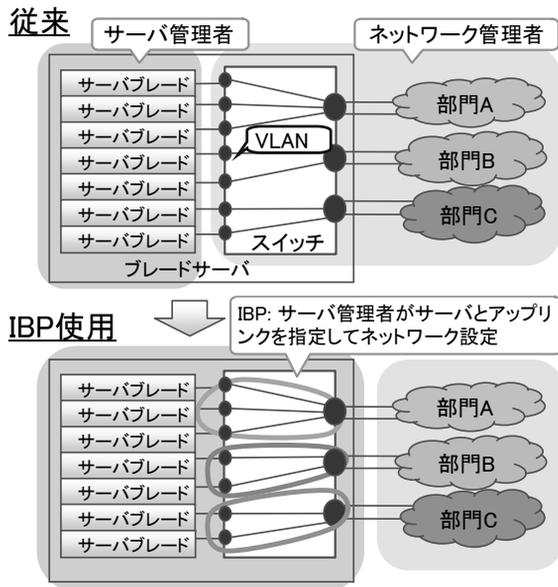


図-4 スイッチブレードの設定の簡便化  
Fig.4-Easy configuration change for switch blade.

● スイッチブレード簡単設定運用機能 (IBP)

IBPは、サーバ管理者にとって難しいと言われていたスイッチブレードの設定・運用を簡単かつ安全に行うために、スイッチブレードに搭載されたソフトウェアである。以下のような特長と機能を有する。

(1) パーティショニング機能

VLANを意識せずに業務や部門別に、サーバブレードとスイッチブレードのアップリンクを対応付けたグループ (パーティション) の形成や、設定ミスによるループの発生を防ぐことができる (図-4)。

(2) サーバ管理との連携

サーバ管理システムServerViewの仮想化ソフトウェアVIOMと連携して設定運用することにより、ブレードサーバの導入や運用を容易にする (図-5)。

I/O仮想化

近年、ライトサイジング化として各拠点に散在させていたサーバ群を集約し、一括したシステム管理を行うためのデータセンタ化が進んでいる。これは、個々のサーバおよびネットワーク性能の向上に伴い、ハードウェアの共有化を図ることで運用設備の削減が可能となり、より使用効率性の高いシステムを構築することを可能としている。その一方で、多機能化したシステム構成の複雑性から、管理コストの上昇が課題とされている。I/O仮想化では、とくに

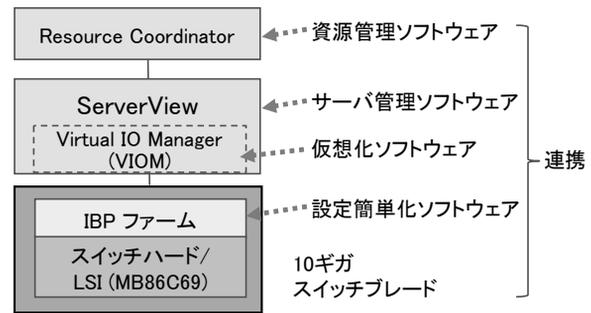


図-5 IBPと仮想化ソフトウェアの連携  
Fig.5-Cooperation between IBP and VIOM.

サーバ管理者を悩ませるLAN (イーサネット) とファイバチャネルの物理接続を論理化 (仮想化) することで、より簡単に効率良く運用できる仕組みを提供するものである。PRIMERGYサーバにおいてI/O仮想化を実現するためには、「ServerView Virtual I/O Manager (VIOM)」または「SAN接続仮想化機構」<sup>④</sup>を利用する。

● ServerView Virtual I/O Manager (VIOM)

ブレードサーバでは、その特長でもあるようにLANスイッチおよびファイバチャネルスイッチを内蔵している。これにより、ネットワークへの接続または変更する場合には、接続先のネットワーク装置やストレージ装置などの設定を変更するため、該当装置の管理者およびネットワーク管理者との協調作業を必要とする。つまり、サーバの追加導入や設定変更、保守交換を行う度に、複数の管理者のスケジュール調整や、システム利用者への連絡など、煩雑な作業に追われることになる。VIOMは、LAN接続用にIBP、ファイバチャネル接続用にアクセスゲートウェイと呼ぶ機能を装備したPRIMERGYブレードサーバと連携し、この問題を解決する (図-6)。LANおよびファイバチャネルの管理者は、接続されているホスト (サーバ) 側の固有情報を維持できれば、とくに設定変更の必要がなくなる。そのためVIOMでは、ブレードサーバのLANの固有情報であるMACと、ファイバチャネルの固有情報であるWWN (World Wide Name) を各接続端で維持させることで対処している。

VIOMは、PRIMERGYブレードサーバ内のLANおよびファイバチャネル接続をすべて管理しており、またMACやWWNのI/Oアドレスをあらかじめプール化 (予備I/Oアドレスを確保) しておくことで、

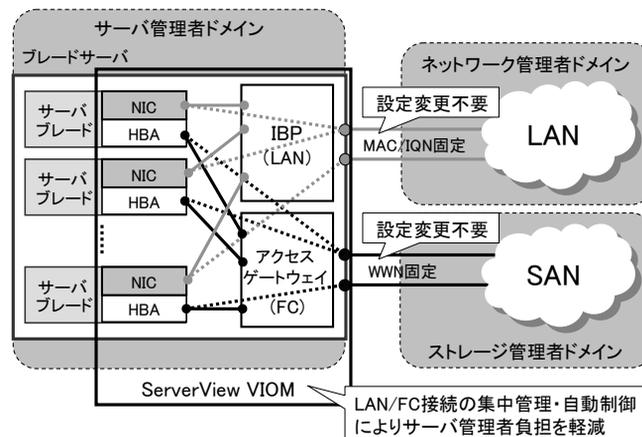


図-6 VIOM概要  
Fig.6-VIOM overview.

サーバ増強によりハードウェアを追加するときにも外部接続の設定を変更する必要が一切ない上に、自動的にI/Oアドレスを割り当て、すぐさまサーバの運用を開始させることができる。

さらに、VIOMではIQN対応も予定しており、iSCSI接続の仮想化も計画している。

#### ● SAN接続仮想化機構

SAN接続仮想化機構 (V-WWN) は、Resource Coordinatorと連携し、PRIMERGYサーバのファイバチャネル接続を仮想化する。V-WWNは、アクセスゲートウェイ機能を必要とせず、サポートされているファイバチャネルカードまたはFC拡張ボードだけの組合せで運用することができるため、ブレードサーバに限定せず、タワー型・ラックマウント型サーバにも応用できる。データセンターで多数のサーバを同時運用する場合などに効果は高まる。

なお、このSAN接続仮想化機構は、「ServerView Resource Coordinator VE I/O仮想化オプション」と名称を変え、さらなる機能拡張を予定している。

## む す び

本稿では、富士通の次世代ブレードサーバであるPRIMERGY BX900の概要、および実装される仮想化技術について紹介した。今後の情報処理分野において中核となる仮想化システムとの親和性を追求した製品として、グローバルなビジネス展開を目指していく。

#### 参考文献

- (1) Y. Hidaka et al. : A 4-Channel 10.3 Gb/s Backplane Transceiver Macro with 35 dB Equalizer and Sign-Based Zero-Forcing Adaptive Control. ISSCC Dig. Tech. Papers, paper 10.5, (February 2009).
- (2) 清水 剛ほか：20ポート10GイーサスイッチLSIとその応用. FUJITSU, Vol.58, No.3, p.246-250 (2007).
- (3) 藤巻秀明ほか：サーバ集約の課題を解決するブレードサーバ PRIMERGY BX620 S4の新技术. FUJITSU, Vol.58, No.5, p.407-412 (2007).