基幹系システム向け仮想化技術「Virtage」の開発（その2）
機器透過性

氏　名[†1] 井上 裕功
所　属[†1] (株)日立製作所 ｴﾝﾀｰﾌﾟﾗｲｽﾞｻｰﾊﾞ事業部
氏　名[†2] 上野 仁
所　属[†2] 同 ｴﾝﾀｰﾌﾟﾗｲｽﾞｻｰﾊﾞ事業部
氏　名[†3] 戸塚 崇夫
所　属[†3] 同 ｴﾝﾀｰﾌﾟﾗｲｽﾞｻｰﾊﾞ事業部

1 はじめに

企業の基幹系システムをサーバ仮想化環境で実現するためには、物理サーバと同様な運用を可能とする機器透過性が必要である。本論文では、日立の基幹系システム向け仮想化技術「Virtage」における、機器透過性の実現方法とその利用シーンについて示す。

2 機器透過性

機器透過性とは、仮想化制御方式における、以下の二つの特徴的な性質のことをいう。

(1) ゲストOS(仮想計算機上のOS)が発行するI/Oコマンドと、そのレスポンスが物理サーバの場合と等価であること。
(2) ゲストOSが利用するディスクのフォーマットが物理サーバの場合と等価であること。

「Virtage」がこの機器透過性を採用した目的について、それぞれの性質の利点と共に示す。

まず(1)により、クラスタ制御ソフトなどが発行する、単純なread/writeでない制御系コマンドのアクセスに対し、正しい応答を返すことができる。これにより、仮想サーバを連携したクラスタシステムを、特別な制約なしに構築することが可能となる。

(2)の目的は二つある。一つはバックアップ環境を構築する際に、ファイルシステムを直接アクセスできるため、仮想化を利用しないバックアップサーバからのアクセスが可能になる点である。これによりLANを介さずに、ゲストOSが使用しているディスクの差分バックアップをファイル単位でとる、いわゆるLANフリーバックアップが可能となる。

もう一つの目的は、仮想ファイルシステムでキャッシュされると都合の悪いソフトを利用できる点である。ソフトウェア方式の仮想ファイルシステムでは、ファイルシステムのキャッシュ機構を用意することが多いが、その場合、物理ディスクへの書き込みのタイミングをゲストOSへ知らせるのが難しい。(2)の性質により書き込みを確実に行うことができるため、これによりデータベースソフトなどの利用が可能になる。

3 仮想化制御アーキテクチャの比較

この章ではI/O仮想化方式について、代表的な二つの方式を比較する形で説明し、機器透過性実現にはそのうちのパススルー方式が必要となることを述べる。

(1) パススルー方式

パススルー方式とは、ゲストOSのDMA転送アドレスを、H/W機構によりホスト物理アドレスに変換し、ダイレクトにセットすることを特徴とする仮想化方式である。「Virtage」ではこのH/W機構として独自のアシスト機構を開発し、これを実現した。第2章で述べた機器透過性の特徴は、このパススルー方式により実現される。

(2) ハイパバイザエミュレーション方式

ハイパバイザエミュレーション方式とは、ゲストOSのDMA転送アドレス変換を、I/O起動時にハイパバイザがトラップしてエミュレーションすることによって行うことを特徴とする、仮想化方式である。この方式ではゲストOSに対するエミュレータと物理デバイスに対するドライバ、及びアドレス変換制御が必要となる。そのため、ゲストOSに対するエミュレータと特定の

Development of Server Virtualization Feature "Virtage" (2), I/O Path Through System
[†1] Hironori Inoue, Enterprise Server Division, Hitachi, Ltd.
[†2] Hitoshi Ueno, Enterprise Server Division, Hitachi, Ltd.
[†3] Takao Totsuka, Enterprise Server Division, Hitachi, Ltd.

デバイスを限定することで開発工数を削減し、標準インタフェースを提供することが多い。

図 3-1に各方式の構造図を、表 3-1に特徴の比較をそれぞれ示す。

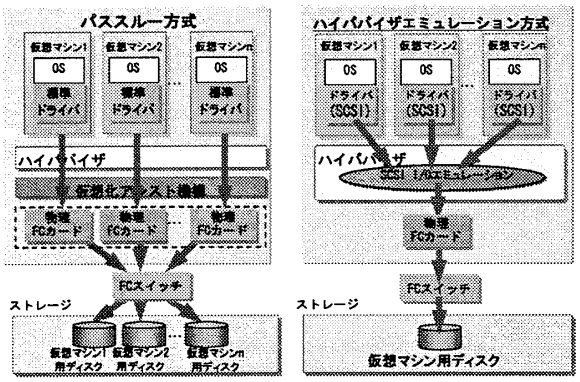

図 3-1 I/O仮想化方式の構造図

表 3-1 I/O仮想化方式の特徴比較

	パススルー方式		ハイパバイザ エミュレーション方式	
I/O性能	○	H/Wによる直接実行	△	ハイパバイザ・エミュレーション
H/W 透過性(1)	○	クラスタソフトなど対応可	×	制御系アクセスへの応答が困難
H/W 透過性(2)	○	LAN フリーバックアップ/DB ソフトなど対応可	×	仮想ファイルシステム対応要
ディスク仮想化	×	仮想ディスクをサポートしない	○	VM の移動性良い

企業の基幹系システム向け仮想化環境では、確実な構成を組むことを重視するため、「Virtage」では H/W を直接アクセスするパススルー方式を採用した。

4 BladeSymphony BS1000 での実現方法とサポート結果

この章では「Virtage」の稼動プラットフォームである日立のブレードサーバ BladeSymphony BS1000 における、機器透過性の具体的な実現方法をアーキテクチャ別に記す。

IPF サーバモジュールでは、先に述べた DMA アドレス変換の仮想化アシスト機構を、自製のチップセットの中に有し、これによって I/O 仮想化のパススルー方式を実現している。

一方 x86 サーバモジュールでは自製 HBA に仮想化アシスト機構の論理を組み込んでおり、これによって I/O のパススルー方式を実現している。

BS1000 の H/W 構成を図 4-1に示す。方式は異なるが、アーキテクチャの種別に関係なく機器透過性を実現することができた。このサポート結果として、第2章で述べたクラスタシステム・LAN フリーバックアップ・データベースなどの、機器透過性により実現される仮想環境を、物理サーバと同様に構築することが可能となっている。

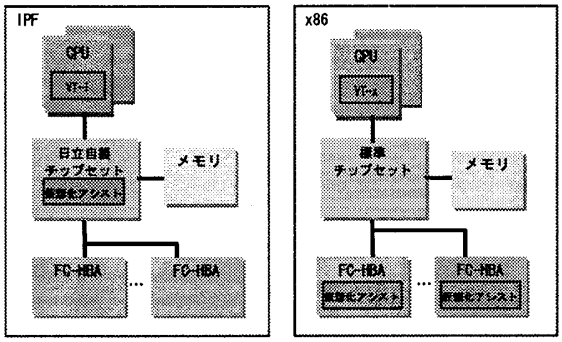

図 4-1 BS1000 の H/W 構造図

5 おわりに

基幹系システム用途に必要となる機器透過性持つ仮想化制御方式を、実現することができた。今後は、以下のような課題について検討していきたい。

(1) パススルー方式の特性を生かした仮想ディスク方式
(2) VT-d などの、今後期待される仮想化サポート技術を利用した、より高性能・高機能の仮想化方式

6 参考文献

[1]田口敏夫ほか：仮想計算機システムの制御効率を向上するための方式と実験結果，情報処理学会論文誌，Vol.20，No.4，pp.281-289(1979).
[2]田口敏夫ほか：実計算機モードと仮想計算機モード間の動的切替え制御方式について，情報処理学会論文誌，Vol.22，No.3，pp.206-215(1981).
[3]梅野英典ほか：仮想計算機システムにおける論理プロセッサをスケジュールする新方式の開発と評価，情報処理学会論文誌，Vol.44，No.3，pp.868-882(2003).

IPF: Itanium Processor Family