(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6421470号
(P6421470)

(45)発行日　平成30年11月14日(2018.11.14)　(24)登録日　平成30年10月26日(2018.10.26)

(51)Int.Cl.			F I		
G06F	9/455	(2006.01)	G06F	9/455	150
G06F	9/50	(2006.01)	G06F	9/50	120Z
G06F	9/44	(2018.01)	G06F	9/44	

請求項の数 7　(全 31 頁)

(21)出願番号　特願2014-124545(P2014-124545)
(22)出願日　平成26年6月17日(2014.6.17)
(65)公開番号　特開2016-4432(P2016-4432A)
(43)公開日　平成28年1月12日(2016.1.12)
審査請求日　平成29年3月9日(2017.3.9)

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100092152
弁理士　服部　毅巖
(72)発明者　長谷川　裕毅
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　治部　将之
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　元藤　雄路
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

最終頁に続く

(54)【発明の名称】仮想マシンマイグレーションプログラム、仮想マシンマイグレーションシステムおよび仮想マシンマイグレーション方法

(57)【特許請求の範囲】
【請求項１】
コンピュータに、
第１のオペレーティングシステムを含む第１の仮想マシンが配置された他のコンピュータから、前記コンピュータ上に起動された第２の仮想マシンに含まれる第２のオペレーティングシステムを実行することで前記第２のオペレーティングシステムにより前記第１のオペレーティングシステムのデータを取得し、
前記第２のオペレーティングシステムを実行することで前記第２のオペレーティングシステムにより、前記第２の仮想マシンを再起動した際に実行されるオペレーティングシステムが、前記第２のオペレーティングシステムから、前記取得した第１のオペレーティングシステムのデータに基づいて実行される前記第１のオペレーティングシステムに変わるように、前記第２の仮想マシンの起動方法を示す起動情報を書き換え、
前記第２の仮想マシンを再起動する、
処理を実行させる仮想マシンマイグレーションプログラム。
【請求項２】
前記第２の仮想マシンには、前記コンピュータが備える記憶装置の中から、第１の記憶領域と前記第２のオペレーティングシステムのデータを記憶する第２の記憶領域と前記起動情報を記憶する第３の記憶領域とが割り当てられ、
前記第１のオペレーティングシステムのデータは前記第１の記憶領域に書き込まれ、
前記起動情報は、前記第２の仮想マシンの再起動の際に前記第１の記憶領域に記憶され

たデータが読み込まれるように書き換えられる、
請求項１記載の仮想マシンマイグレーションプログラム。
【請求項３】
前記第２の仮想マシンへの前記第２の記憶領域の割り当ては、前記第２の仮想マシンで前記第１のオペレーティングシステムが実行された後に解除される、
請求項２記載の仮想マシンマイグレーションプログラム。
【請求項４】
第１のオペレーティングシステムを含む第１の仮想マシンが配置され、前記第１のオペレーティングシステムのデータを送信する第１の情報処理装置と、
第２のオペレーティングシステムを含む第２の仮想マシンを起動し、
前記第２のオペレーティングシステムを実行することで前記第２のオペレーティングシステムにより前記第１の情報処理装置から前記第１のオペレーティングシステムのデータを取得し、
前記第２のオペレーティングシステムを実行することで前記第２のオペレーティングシステムにより、前記第２の仮想マシンを再起動した際に実行されるオペレーティングシステムが、前記第２のオペレーティングシステムから、前記取得した第１のオペレーティングシステムのデータに基づいて実行される前記第１のオペレーティングシステムに変わるように、前記第２の仮想マシンの起動方法を示す起動情報を書き換え、
前記第２の仮想マシンを再起動する、第２の情報処理装置と、
を有する仮想マシンマイグレーションシステム。
【請求項５】
前記第１の情報処理装置は、前記第１のオペレーティングシステムを一時的にサスペンド状態に遷移させることで、メモリに記憶されたデータおよびプロセッサ内のレジスタに記憶されたデータの少なくとも一方を抽出する、
請求項４記載の仮想マシンマイグレーションシステム。
【請求項６】
前記第１の情報処理装置は、前記第１の情報処理装置における単位時間当たりのデータ更新量が閾値を超えている場合、前記第１の情報処理装置で実行されている１または２以上のプロセスに対してプロセッサのリソースの割り当てを制限する、
請求項４または５記載の仮想マシンマイグレーションシステム。
【請求項７】
第１の情報処理装置と第２の情報処理装置とを有するシステムが実行する仮想マシンマイグレーション方法であって、
前記第１の情報処理装置に配置された第１のオペレーティングシステムを含む第１の仮想マシンと異なる、第２のオペレーティングシステムを含む第２の仮想マシンを、前記第２の情報処理装置上に起動し、
前記第１のオペレーティングシステムのデータを、前記第１の情報処理装置から、前記第２のオペレーティングシステムを実行することで前記第２のオペレーティングシステムにより前記第２の情報処理装置にコピーし、
前記第２のオペレーティングシステムを実行することで前記第２のオペレーティングシステムにより、前記第２の仮想マシンを再起動した際に実行されるオペレーティングシステムが、前記第２のオペレーティングシステムから、前記取得した第１のオペレーティングシステムのデータに基づいて実行される前記第１のオペレーティングシステムに変わるように、前記第２の仮想マシンの起動方法を示す起動情報を書き換え、
前記第２の仮想マシンを再起動する、
仮想マシンマイグレーション方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は仮想マシンマイグレーションプログラム、仮想マシンマイグレーションシステ

ムおよび仮想マシンマイグレーション方法に関する。

【背景技術】

【０００２】

現在、プロセッサやメモリやハードディスクなどのコンピュータのリソースを保持し、ネットワーク経由でユーザにリソースを使用させる情報処理サービスが存在する。このような情報処理サービスは、クラウドサービスと呼ばれることがある。

【０００３】

データセンタなどクラウドサービスを提供する施設では、仮想化技術を用いて、１つの物理的なコンピュータ（物理マシン）上に複数の仮想的なコンピュータ（仮想マシン）を形成することがある。ハイパーバイザなど仮想マシンを制御する制御ソフトウェアは、複数の仮想マシンそれぞれにリソースを割り当てる。クラウドサービスの施設は、例えば、ユーザからの要求に応じてリソースを確保し、当該ユーザ用の仮想マシンを生成する。ユーザは、生成された仮想マシン上で、オペレーティングシステム（ＯＳ：Operating System）やアプリケーションソフトウェアなどの所望のソフトウェアを実行させることができる。すなわち、ユーザは、当該ユーザ用の仮想マシンのリソースを使用することができる。仮想マシンなど情報基盤（インフラストラクチャ）そのものをユーザに提供するクラウドサービスは、ＩａａＳ（Infrastructure as a Service）と呼ばれることがある。

【０００４】

仮想マシンは、ハイパーバイザなどの制御ソフトウェアによる制御のもと、ある物理マシンから他の物理マシンに移行すること（マイグレーション）ができる場合がある。特に、移行元の物理マシンにおいて仮想マシンのＯＳをシャットダウンせずに、メモリデータやプロセッサの状態データなどを移行先の物理マシンにコピーし、移行先の物理マシンで仮想マシンの処理を引き継ぐことができる場合がある。このようなマイグレーション方法は、ライブマイグレーションと呼ばれることがある。仮想マシンのマイグレーションは、一部の物理マシンでリソース不足が生じたときや一部の物理マシンを停止して保守作業を行うときなどに、同一のクラウドサービスの施設内で行われることがある。

【０００５】

なお、ある物理マシンから他の物理マシンに仮想マシンを移行するとき、仮想マシン上で実行されるプログラムのうち実行頻度の高いプログラムのメモリデータを優先的に物理マシン間で送信する仮想マシンの移行システムが提案されている。また、物理マシン間で処理を移行するとき、移行先の物理マシンに応じてプロセスマイグレーションと仮想マシンマイグレーションとを選択的に実行する分散処理システムが提案されている。メモリデータやレジスタデータの移行は、プロセスマイグレーションではプロセス単位で行われ、仮想マシンマイグレーションでは仮想マシン単位で行われる。

【０００６】

また、物理マシン間でゲストＯＳを移行する情報処理システムが提案されている。この情報処理システムでは、移行元の物理マシンが参照するストレージと移行先の物理マシンが参照するストレージに、同じゲストＯＳのイメージデータを格納しておく。また、移行元の物理マシンは、ゲストＯＳのイメージデータへの書き込みを禁止し、書き込みデータをメモリ上にキャッシュしておく。ゲストＯＳの移行時には、移行元の物理マシンがメモリデータを移行先の物理マシンにコピーし、移行先の物理マシンは移行先のストレージに格納されたイメージデータと受信したメモリデータとを用いてゲストＯＳを起動する。

【０００７】

また、ライブマイグレーションの際に、移行元の物理マシンにおける単位時間当たりのメモリへの書き込み量を、物理マシン間のデータ送信帯域よりも小さくなるように低減する仮想化システムが提案されている。この仮想化システムでは、メモリに対してデータを書き込んだ旨をプログラムに報告するときに発行される割り込みを一時的に保留して遅延させることで、メモリに対する連続的なデータ書き込みについて書き込み速度を低減する。

【先行技術文献】

【特許文献】
【０００８】
【特許文献１】国際公開第２００９／０６９５７３号
【特許文献２】国際公開第２０１０／０３５４８０号
【特許文献３】特開２０１１－２１０１５１号公報
【特許文献４】特開２０１３－２５０９５０号公報
【発明の概要】
【発明が解決しようとする課題】
【０００９】
ところで、ユーザは、互いに離れた複数の情報処理施設（例えば、異なるサービス事業者がもつ複数のデータセンタ）を使い分けることがある。このとき、ある情報処理施設に配置された仮想マシンを他の情報処理施設に移行したいことが考えられる。例えば、ある仮想マシンの負荷が高くなったとき、リソースを豊富にもつ情報処理施設に当該仮想マシンを移行したいことが考えられる。また、例えば、ある仮想マシンを、課金の少ない情報処理施設やセキュリティの高い情報処理施設に移行したいことなども考えられる。
【００１０】
一方で、情報処理施設には、ユーザの仮想マシン間のデータ通信またはユーザの仮想マシンと外部ネットワークとの間のデータ通信に用いられるユーザネットワークとは別に、仮想マシンの制御に用いられる管理ネットワークが設けられることがある。この場合、同一の情報処理施設内でのマイグレーションには、管理ネットワークが使用される。
【００１１】
しかし、セキュリティなどの観点から、複数の情報処理施設の間で管理ネットワークを相互接続する（互いに相手の管理ネットワークにアクセス可能にする）ことは容易でない。特に、異なるサービス事業者がもつデータセンタの間で管理ネットワークを相互接続することは難しい。このため、同一の情報処理施設内でのマイグレーション方法を、異なる情報処理施設の間のマイグレーションに適用することは難しいという問題がある。その結果、異なる情報処理施設の間では仮想マシンを円滑に移行することが難しかった。
【００１２】
１つの側面では、本発明は、異なる情報処理施設の間で仮想マシンを円滑に移行できるようにする仮想マシンマイグレーションプログラム、仮想マシンマイグレーションシステムおよび仮想マシンマイグレーション方法を提供することを目的とする。
【課題を解決するための手段】
【００１３】
１つの態様では、コンピュータに以下の処理を実行させる仮想マシンマイグレーションプログラムが提供される。第１のオペレーティングシステムを含む第１の仮想マシンが配置された他のコンピュータから、コンピュータ上に起動された第２の仮想マシンに含まれる第２のオペレーティングシステムを用いて第１のオペレーティングシステムのデータを取得する。第２の仮想マシンの再起動の際に取得した第１のオペレーティングシステムのデータに基づいて第１のオペレーティングシステムが実行されるように、第２の仮想マシンの起動方法を示す起動情報を書き換える。第２の仮想マシンを再起動する。
【００１４】
また、１つの態様では、第１の情報処理装置と第２の情報処理装置とを有する仮想マシンマイグレーションシステムが提供される。
また、１つの態様では、第１の情報処理装置と第２の情報処理装置とを有するシステムが実行する仮想マシンマイグレーション方法が提供される。
【発明の効果】
【００１５】
１つの側面では、異なる情報処理施設の間で仮想マシンを円滑に移行できる。
【図面の簡単な説明】
【００１６】

【図１】第１の実施の形態の仮想マシンマイグレーションシステムを示す図である。
【図２】第２の実施の形態の情報処理システムを示す図である。
【図３】ホストサーバのハードウェア例を示すブロック図である。
【図４】仮想マシンとハイパーバイザの配置例を示すブロック図である。
【図５】仮想マシンのマイグレーション例を示す第１の図である。
【図６】仮想マシンのマイグレーション例を示す第２の図である。
【図７】仮想マシンのマイグレーション例を示す第３の図である。
【図８】仮想マシンのマイグレーション例を示す第４の図である。
【図９】仮想マシンの機能例を示すブロック図である。
【図１０】状態テーブルの例を示す図である。
【図１１】ディスク更新情報とメモリ更新情報の例を示す図である。
【図１２】プロセステーブルの例を示す図である。
【図１３】リソース制限テーブルと重要プロセスリストの例を示す図である。
【図１４】仮想マシン生成要求の例を示す図である。
【図１５】マイグレーション開始の手順例を示すフローチャートである。
【図１６】ディスクデータ送信の手順例を示すフローチャートである。
【図１７】ディスク監視の手順例を示すフローチャートである。
【図１８】メモリデータ送信の手順例を示すフローチャートである。
【図１９】メモリ監視の手順例を示すフローチャートである。
【図２０】プロセス監視の手順例を示すフローチャートである。
【図２１】プロセス監視の手順例を示すフローチャート（続き）である。
【図２２】マイグレーション完了の手順例を示すフローチャートである。
【図２３】マイグレーション完了の手順例を示すフローチャート（続き）である。
【発明を実施するための形態】
【００１７】
以下、本実施の形態を図面を参照して説明する。
［第１の実施の形態］
図１は、第１の実施の形態の仮想マシンマイグレーションシステムを示す図である。
【００１８】
第１の実施の形態の仮想マシンマイグレーションシステムは、情報処理装置１，２を有する。情報処理装置１，２は、１または２以上の仮想マシンを配置することができるコンピュータであり、例えば、ユーザに対してＩａａＳなどのクラウドサービスを提供するサーバコンピュータである。情報処理装置１，２は、異なる情報処理施設（例えば、異なるデータセンタ）に設置される。情報処理装置１が属する情報処理施設と情報処理装置２が属する情報処理施設とは、異なるサービス事業者によって管理されてもよい。情報処理装置１，２は、インターネットなどの広域ネットワークを介して互いに通信できる。
【００１９】
情報処理装置１には、仮想マシン１ａが配置される。仮想マシン１ａには、情報処理装置１が備えるプロセッサやメモリなどの情報処理のリソースが割り当てられる。仮想マシン１ａは、オペレーティングシステム１ｂを含む。オペレーティングシステム１ｂは、仮想マシン１ａに割り当てられたリソースを用いて実行される。
【００２０】
上記のプロセッサは、ＣＰＵ（Central Processing Unit）やＤＳＰ（Digital Signal Processor）であってもよい。また、プロセッサは、ＡＳＩＣ（Application Specific Integrated Circuit）やＦＰＧＡ（Field Programmable Gate Array）などの特定用途の電子回路を含んでもよい。プロセッサは、例えば、メモリに記憶されたプログラムを実行する。複数のプロセッサの集合（マルチプロセッサ）を「プロセッサ」と呼ぶこともある。メモリは、例えば、ＲＡＭ（Random Access Memory）などの半導体メモリである。
【００２１】
また、情報処理装置１は、記憶部１ｃを有する。記憶部１ｃは、例えば、ＨＤＤ（Hard

Disk Drive）やフラッシュメモリなど不揮発性の記憶装置である。記憶部１ｃは、記憶領域１ｄを含む。記憶領域１ｄは、仮想マシン１ａに割り当てられたリソースであり、オペレーティングシステム１ｂのデータを記憶する。オペレーティングシステム１ｂのデータには、オペレーティングシステム１ｂの処理を定義したプログラムや、設定ファイルなどのプログラム以外のデータが含まれてもよい。情報処理装置２は、記憶部２ｃを有する。記憶部２ｃは、例えば、ＨＤＤやフラッシュメモリなど不揮発性の記憶装置である。

【００２２】

ここで、情報処理装置１から情報処理装置２への仮想マシン１ａのマイグレーションを考える。同一の情報処理施設の中でのマイグレーションでは、ユーザネットワークとは別個に設けられた管理ネットワークを介して、オペレーティングシステムのデータが転送されることがある。ユーザネットワークは、ユーザの仮想マシン間またはユーザの仮想マシンと情報処理施設の外との間のデータ通信に用いられるネットワークである。管理ネットワークは、仮想マシンの制御に用いられるネットワークである。

【００２３】

しかし、異なる情報処理施設の間では、セキュリティなどの観点から、管理ネットワークへの直接アクセス（相互接続）が制限されることが想定される。このため、情報処理装置１，２の間のマイグレーションは、管理ネットワークを用いて行うことが難しい場合がある。そこで、第１の実施の形態では、管理ネットワークを用いずに（例えば、ユーザネットワークを用いて）仮想マシン１ａのマイグレーションを可能にする。

【００２４】

仮想マシン１ａのマイグレーションでは、情報処理装置２は、仮想マシン２ａを起動する。仮想マシン２ａは、オペレーティングシステム２ｂを含む。オペレーティングシステム２ｂは、マイグレーション中に一時的に用いられるソフトウェアであり、仮想マシン２ａに対して割り当てられた情報処理装置２のプロセッサやメモリを用いて実行される。

【００２５】

また、仮想マシン２ａには、記憶部２ｃに含まれる記憶領域２ｄ，２ｅ，２ｆが割り当てられる。記憶領域２ｄは、空の記憶領域である。記憶領域２ｅは、オペレーティングシステム２ｂのデータを記憶する。記憶領域２ｆは、起動情報を記憶する。起動情報は、仮想マシン２ａの起動方法を示す。例えば、仮想マシン２ａが最初に起動されるときの起動情報は、起動時に記憶領域２ｅに記憶されたデータを読み込むことを示している。これにより、仮想マシン２ａの起動時にオペレーティングシステム２ｂが実行される（Ｓ１）。

【００２６】

仮想マシン２ａが起動されると、情報処理装置１から情報処理装置２にオペレーティングシステム１ｂのデータがコピーされる。データ送信では、例えば、広域ネットワークおよび各情報処理施設内のユーザネットワークが用いられ、各情報処理施設内の管理ネットワークは迂回される。情報処理装置１は、記憶領域１ｄに記憶されたデータを情報処理装置２に送信する。情報処理装置２は、仮想マシン２ａのオペレーティングシステム２ｂを用いて、情報処理装置１からオペレーティングシステム１ｂのデータを取得する。取得されたオペレーティングシステム１ｂのデータは、記憶領域２ｄに書き込まれる（Ｓ２）。

【００２７】

オペレーティングシステム１ｂのデータのコピーが完了すると、情報処理装置２は、オペレーティングシステム２ｂを用いて、記憶領域２ｆに記憶された起動情報を書き換える（Ｓ３）。書き換えられた起動情報は、仮想マシン２ａの再起動の際に、記憶領域２ｅに記憶されたデータに代えて記憶領域２ｄに記憶されたデータを読み込むことを示している。これにより、仮想マシン２ａの再起動時に、オペレーティングシステム２ｂに代えてオペレーティングシステム１ｂが実行されることになる。

【００２８】

起動情報が書き換えられると、情報処理装置２は、仮想マシン２ａを再起動する。例えば、情報処理装置２は、仮想マシン２ａのオペレーティングシステム２ｂをシャットダウンさせる。情報処理装置２は、記憶領域２ｆに記憶された起動情報を読み込み、読み込ん

だ起動情報に従って記憶領域２ｄに記憶されたデータを読み込む。これにより、仮想マシン２ａではオペレーティングシステム１ｂが実行される（Ｓ４）。このとき、仮想マシン１ａを停止させ、情報処理装置１から仮想マシン１ａを削除してもよい。また、記憶領域２ｅに記憶されたオペレーティングシステム２ｂのデータを削除してもよい。移行元の仮想マシン１ａの処理は、移行先の仮想マシン２ａに引き継がれる。よって、仮想マシン１ａは、実質的に情報処理装置１から情報処理装置２に移行したことになる。

【００２９】

第１の実施の形態の仮想マシンマイグレーションシステムによれば、移行先の情報処理装置２で仮想マシン２ａが起動し、仮想マシン２ａ上でオペレーティングシステム２ｂが一時的に実行される。オペレーティングシステム１ｂのデータのコピーは、仮想マシン１ａと仮想マシン２ａとの間の通信として行うことができる。よって、情報処理施設外部からの管理ネットワークへの直接アクセスが制限されていても、管理ネットワークを迂回しユーザネットワークを用いてオペレーティングシステム１ｂのデータのコピーできる。

【００３０】

そして、起動情報を書き換えることで、仮想マシン２ａが再起動すると、オペレーティングシステム２ｂに代えてオペレーティングシステム１ｂが仮想マシン２ａ上で実行される。これにより、異なる情報処理施設の間（例えば、異なるサービス事業者のデータセンタの間）でも、仮想マシン１ａのマイグレーションを円滑に行うことができる。なお、上記のマイグレーションの制御は、例えば、オペレーティングシステム１ｂ，２ｂに組み込まれるプログラムモジュールとして実装することが可能である。

【００３１】

［第２の実施の形態］

図２は、第２の実施の形態の情報処理システムを示す図である。

第２の実施の形態の情報処理システムは、データセンタ１０，２０、インターネット３０およびクライアント３１を含む。データセンタ１０，２０は、クラウドサービスを提供する情報処理施設である。データセンタ１０とデータセンタ２０とは、異なるサービス事業者によって管理されている。クライアント３１のユーザは、契約により、データセンタ１０，２０両方のクラウドサービスを利用できる。クライアント３１は、広域ネットワークであるインターネット３０を介して、データセンタ１０，２０にアクセスする。

【００３２】

データセンタ１０は、ユーザネットワーク１１、管理ネットワーク１２、管理サーバ１３およびホストサーバ１００，１００ａを有する。データセンタ２０も、データセンタ１０と対応するように、ユーザネットワーク２１、管理ネットワーク２２、管理サーバ２３およびホストサーバ２００，２００ａを有する。以下では、代表してデータセンタ１０について説明し、データセンタ２０については説明を省略する。

【００３３】

ユーザネットワーク１１は、ホストサーバ１００，１００ａに配置されたユーザ用の仮想マシンがデータ通信に使用するローカルネットワークである。ユーザネットワーク１１は、例えば、スイッチなどの通信装置を含む有線通信ネットワークである。ユーザネットワーク１１は、データセンタ１０の外部にあるインターネット３０と接続されている。データセンタ２０に属するコンピュータやクライアント３１は、インターネット３０およびユーザネットワーク１１を介して、ユーザ用の仮想マシンと通信できる。

【００３４】

管理ネットワーク１２は、ホストサーバ１００，１００ａに配置された管理用の仮想マシンや管理サーバ１３が、ユーザ用の仮想マシンの制御に使用するローカルネットワークである。管理ネットワーク１２は、例えば、スイッチなどの通信装置を含む有線通信ネットワークである。仮想マシンの制御には、ホストサーバ１００，１００ａ上に新たな仮想マシンを生成することや、データセンタ１０内のホストサーバ１００，１００ａ間で仮想マシンを移行することなどが含まれる。例えば、管理サーバ１３が、管理ネットワーク１２を介してホストサーバ１００に、新たな仮想マシンの生成を指示する。データセンタ２

０に属するコンピュータやクライアント３１は、管理ネットワーク１２には直接アクセスしない。

【００３５】

管理サーバ１３は、データセンタ１０の外部に対してクラウドサービスの管理インタフェースを提供するサーバコンピュータである。管理サーバ１３は、インターネット３０と接続されている。データセンタ２０内のコンピュータやクライアント３１は、インターネット３０を介して管理サーバ１３にアクセスすることができる。管理サーバ１３は、仮想マシンの生成・起動・停止・削除などのコマンドを受信することがある。

【００３６】

例えば、仮想マシンの生成が要求されると、管理サーバ１３は、データセンタ１０内のホストサーバのうち要求量以上の空きリソースをもつホストサーバを選択し、選択したホストサーバのリソースを新たな仮想マシンに対して割り当てる。仮想マシンの起動が要求されると、管理サーバ１３は、指定された仮想マシンを起動し、当該仮想マシン上でオペレーティングシステムを実行させる。仮想マシンの停止が要求されると、管理サーバ１３は、指定された仮想マシン上のオペレーティングシステムをシャットダウンし、当該仮想マシンを停止させる。仮想マシンの削除が要求されると、管理サーバ１３は、指定された仮想マシンに割り当てていたリソースを解放する。

【００３７】

ホストサーバ１００，１００ａは、複数の仮想マシンを動作させる仮想化環境をもつサーバコンピュータである。ホストサーバ１００，１００ａそれぞれには、１つの管理用の仮想マシンと１または２以上のユーザ用の仮想マシンとが配置される。管理用の仮想マシンはホストＯＳを実行し、ユーザ用の仮想マシンはゲストＯＳを実行する。

【００３８】

また、ホストサーバ１００，１００ａそれぞれは、当該ホストサーバが備えるプロセッサやメモリなどのリソースを複数の仮想マシンに割り振るハイパーバイザを実行している。ハイパーバイザは、管理サーバ１３からの指示に応じて、仮想マシンの生成・削除やデータセンタ１０内での仮想マシンのマイグレーションを行う。また、ハイパーバイザは、仮想マシンの通信について、ユーザネットワーク１１と管理ネットワーク１２とを使い分ける。管理用の仮想マシンの通信については、主に管理ネットワーク１２が使用される。ユーザ用の仮想マシンの通信については、主にユーザネットワーク１１が使用される。

【００３９】

クライアント３１は、ユーザが操作する端末装置としてのクライアントコンピュータである。クライアント３１は、データセンタ１０のクラウドサービスを利用するにあたり、インターネット３０を介して管理サーバ１３にアクセスする。例えば、データセンタ１０に新たな仮想マシンを配置する場合、クライアント３１は、管理サーバ１３に仮想マシンの生成を要求する。生成された仮想マシンには、インターネット３０およびユーザネットワーク１１を介してアクセスすることができる。同様に、クライアント３１は、データセンタ２０のクラウドサービスを利用するにあたり、管理サーバ２３にアクセスする。

【００４０】

第２の実施の形態では、データセンタ１０とデータセンタ２０との間で仮想マシンのマイグレーションを行うことができる。異なるデータセンタの間のマイグレーションにおけるデータ転送は、管理ネットワーク１２，２２を用いず、ユーザネットワーク１１，２１を用いて仮想マシン間のデータ通信として行われる。データセンタ１０からデータセンタ２０に仮想マシンを移行する場合、クライアント３１は、ユーザネットワーク１１を介して移行元の仮想マシンにマイグレーションを指示する。また、データセンタ２０からデータセンタ１０に仮想マシンを移行する場合、クライアント３１は、ユーザネットワーク２１を介して移行元の仮想マシンにマイグレーションを指示する。

【００４１】

図３は、ホストサーバのハードウェア例を示すブロック図である。

ホストサーバ１００は、ＣＰＵ１０１、ＲＡＭ１０２、ＨＤＤ１０３、画像信号処理部

１０４、入力信号処理部１０５、媒体リーダ１０６および通信インタフェース１０７，１０８を有する。上記のユニットは、それぞれバス１０９に接続されている。なお、ＨＤＤ１０３は、第１の実施の形態の記憶部１ｃの一例である。

【００４２】

ＣＰＵ１０１は、プログラムの命令を実行する演算回路を含むプロセッサである。ＣＰＵ１０１は、ＨＤＤ１０３に記憶されたプログラムやデータの少なくとも一部をＲＡＭ１０２にロードし、プログラムを実行する。なお、ＣＰＵ１０１は複数のプロセッサコアを備えてもよく、ホストサーバ１００は複数のプロセッサを備えてもよく、以下で説明する処理を複数のプロセッサまたはプロセッサコアを用いて並列に実行してもよい。また、複数のプロセッサの集合（マルチプロセッサ）を「プロセッサ」と呼んでもよい。

【００４３】

ＲＡＭ１０２は、ＣＰＵ１０１が実行するプログラムやＣＰＵ１０１が演算に用いるデータを一時的に記憶する揮発性の半導体メモリである。なお、ホストサーバ１００は、ＲＡＭ以外の種類のメモリを備えてもよく、複数個のメモリを備えてもよい。

【００４４】

ＨＤＤ１０３は、ＯＳやミドルウェアやアプリケーションソフトウェアなどのソフトウェアのプログラム、および、データを記憶する不揮発性の記憶装置である。プログラムには、仮想マシンのマイグレーションを制御するプログラムが含まれる。なお、ホストサーバ１００は、フラッシュメモリやＳＳＤ（Solid State Drive）などの他の種類の記憶装置を備えてもよく、複数の不揮発性の記憶装置を備えてもよい。

【００４５】

画像信号処理部１０４は、ＣＰＵ１０１からの命令に従って、ホストサーバ１００に接続されたディスプレイ１１１に画像を出力する。ディスプレイ１１１としては、ＣＲＴ（Cathode Ray Tube）ディスプレイ、液晶ディスプレイ（ＬＣＤ：Liquid Crystal Display）、プラズマディスプレイ（ＰＤＰ：Plasma Display Panel）、有機ＥＬ（ＯＥＬ：Organic Electro-Luminescence）ディスプレイなどを用いることができる。

【００４６】

入力信号処理部１０５は、ホストサーバ１００に接続された入力デバイス１１２から入力信号を取得し、ＣＰＵ１０１に出力する。入力デバイス１１２としては、マウスやタッチパネルやタッチパッドやトラックボールなどのポインティングデバイス、キーボード、リモートコントローラ、ボタンスイッチなどを用いることができる。また、ホストサーバ１００に、複数の種類の入力デバイスが接続されていてもよい。

【００４７】

媒体リーダ１０６は、記録媒体１１３に記録されたプログラムやデータを読み取る読み取り装置である。記録媒体１１３として、例えば、フレキシブルディスク（ＦＤ：Flexible Disk）やＨＤＤなどの磁気ディスク、ＣＤ（Compact Disc）やＤＶＤ（Digital Versatile Disc）などの光ディスク、光磁気ディスク（ＭＯ：Magneto-Optical disk）、半導体メモリなどを使用できる。媒体リーダ１０６は、例えば、記録媒体１１３から読み取ったプログラムやデータをＲＡＭ１０２またはＨＤＤ１０３に格納する。

【００４８】

通信インタフェース１０７は、ユーザネットワーク１１に接続され、ユーザネットワーク１１を介してデータセンタ１０内外のコンピュータと通信を行う。通信インタフェース１０８は、管理ネットワーク１２に接続され、管理ネットワーク１２を介してデータセンタ１０内のコンピュータと通信を行う。通信インタフェース１０７は、スイッチなどの通信装置とケーブルで接続される有線通信インタフェースでもよいし、基地局またはアクセスポイントと無線リンクで接続される無線通信インタフェースでもよい。

【００４９】

なお、ホストサーバ１００は、媒体リーダ１０６を備えていなくてもよく、ユーザが操作する端末装置から制御可能である場合には画像信号処理部１０４や入力信号処理部１０５を備えていなくてもよい。また、ディスプレイ１１１や入力デバイス１１２が、ホスト

サーバ１００の筐体と一体に形成されていてもよい。管理サーバ１３，２３、クライアント３１およびホストサーバ１００ａ，２００，２００ａも、ホストサーバ１００と同様のハードウェア構成によって実現することができる。

【００５０】

図４は、仮想マシンとハイパーバイザの配置例を示すブロック図である。

ホストサーバ１００は、仮想マシン１２１，１２３，１２５を有する。仮想マシン１２１，１２３は、クライアント３１のユーザからの要求に応じて生成されたユーザ用の仮想マシンである。仮想マシン１２１は、ゲストＯＳ１２２を実行する。仮想マシン１２３は、ゲストＯＳ１２４を実行する。ゲストＯＳ１２２，１２４上では、ユーザから指示されたアプリケーションソフトウェアが実行される。仮想マシン１２５は、ホストサーバ１００の管理に用いられる管理用の仮想マシンである。仮想マシン１２５は、ホストＯＳ１２６を実行する。ホストＯＳ１２６上では、管理ソフトウェアが実行される。

【００５１】

また、ホストサーバ１００は、ハイパーバイザ１２７を実行する。ハイパーバイザ１２７は、ＣＰＵ１０１の処理時間やＲＡＭ１０２の記憶領域などのリソースを仮想マシン１２１，１２３，１２５に割り振り、仮想マシン１２１，１２３，１２５の動作を制御する。ハイパーバイザ１２７は、仮想スイッチ１２８，１２９を有する。

【００５２】

仮想スイッチ１２８は、通信インタフェース１０８の通信帯域を管理し、仮想マシン１２５と管理ネットワーク１２との間の通信を実現する。例えば、仮想スイッチ１２８は、通信インタフェース１０８から届いたパケットをホストＯＳ１２６に届け、ホストＯＳ１２６が出力したパケットを通信インタフェース１０８から送信する。

【００５３】

仮想スイッチ１２９は、通信インタフェース１０７の通信帯域を管理し、仮想マシン１２１，１２３とユーザネットワーク１１との間の通信や、仮想マシン１２１，１２３の間の通信を実現する。例えば、仮想スイッチ１２９は、通信インタフェース１０７から届いたパケットを、宛先アドレスに応じてゲストＯＳ１２２，１２４に振り分ける。また、仮想スイッチ１２９は、ゲストＯＳ１２２，１２４が出力したパケットを、宛先アドレスに応じて通信インタフェース１０７から送信するかまたは他のゲストＯＳに届ける。

【００５４】

次に、データセンタ１０，２０の間の仮想マシンのマイグレーションを説明する。

図５は、仮想マシンのマイグレーション例を示す第１の図である。

以降の第２の実施の形態の説明では、データセンタ１０に属するホストサーバ１００からデータセンタ２０に属するホストサーバ２００へのマイグレーションを考える。

【００５５】

前述の通り、ホストサーバ１００は、ＣＰＵ１０１、ＲＡＭ１０２およびＨＤＤ１０３などのリソースを有する。これらのリソースを用いて、ホストサーバ１００ではゲストＯＳ１２２を含む仮想マシン１２１が動作する。ＣＰＵ１０１が有するレジスタ１０１ａには、仮想マシン１２１に関するデータが格納され得る。ＲＡＭ１０２の中の仮想マシン１２１に対して割り当てられた記憶領域には、仮想マシン１２１に関するデータが一時的に格納される。この割り当てられた記憶領域は、仮想メモリと言うこともできる。

【００５６】

ＨＤＤ１０３の中の記憶領域のうち、パーティション１３１，１３４が仮想マシン１２１に対して割り当てられる。パーティション１３１は、ゲストＯＳ１２２に関するシステムデータを記憶する。システムデータには、ゲストＯＳ１２２のプログラムや設定ファイルなどが含まれる。パーティション１３４は、仮想マシン１２１が使用するデータであってシステムデータ以外のものを記憶する。パーティション１３４は、ユーザが任意のデータを保存するために使用することが可能である。パーティション１３１，１３４は、仮想マシン１２１にとって仮想ディスクと言うこともできる。

【００５７】

パーティション１３１は、起動情報領域１３２および退避領域１３３を含む。起動情報領域１３２は、起動情報を記憶する。起動情報は、仮想マシン１２１の起動時に実行されるオペレーティングシステムのデータが記憶されたパーティションを示す。ここでは、起動情報はパーティション１３１を指し示している。よって、仮想マシン１２１の起動時にはパーティション１３１からＲＡＭ１０２にシステムデータがロードされることになる。

【００５８】

退避領域１３３は、レジスタ１０１ａやＲＡＭ１０２に記憶された仮想マシン１２１に関するデータを退避する記憶領域である。例えば、仮想マシン１２１がサスペンド状態に遷移するなど処理を中断するときに、中断時点のレジスタ１０１ａやＲＡＭ１０２やその他の周辺デバイスのデータを含む中断ファイルが退避領域１３３に書き込まれる。仮想マシン１２１が中断した処理を再開するとき、中断ファイルを用いて状態が再現される。

【００５９】

また、ホストサーバ２００は、ＣＰＵ２０１、ＲＡＭ２０２およびＨＤＤ２０３などのリソースを有する。これらのリソースを用いて、ホストサーバ２００ではゲストＯＳ２２２を含む仮想マシン２２１が動作する。仮想マシン２２１は、仮想マシン１２１の処理を引き継ぐ移行先の仮想マシンである。ゲストＯＳ２２２は、仮想マシン１２１のマイグレーションのために一時的に実行されるオペレーティングシステムである。

【００６０】

ＣＰＵ２０１が有するレジスタ２０１ａには、仮想マシン２２１に関するデータが格納され得る。ＲＡＭ２０２の中の仮想マシン２２１に対して割り当てられた記憶領域には、仮想マシン２２１に関するデータが一時的に格納される。

【００６１】

ＨＤＤ２０３の中の記憶領域のうち、パーティション２３１，２３４，２３５が仮想マシン２２１に対して割り当てられる。パーティション２３１は、ホストサーバ１００から移行されるゲストＯＳ１２２に関するシステムデータを格納するために用意した記憶領域である。パーティション２３４は、ホストサーバ１００から移行されるシステムデータ以外のデータを格納するために用意した記憶領域である。パーティション２３１はパーティション１３１に対応し、パーティション２３４はパーティション１３４に対応する。パーティション２３５は、ゲストＯＳ２２２に関するシステムデータを記憶する。

【００６２】

パーティション２３１は、起動情報領域２３２および退避領域２３３を含む。起動情報領域２３２は、起動情報領域１３２と同様に起動情報を記憶する。ただし、仮想マシン２２１が生成された時点における起動情報領域２３２の起動情報は、パーティション２３５を指し示している。よって、仮想マシン２２１が最初に起動するときは、パーティション２３５からゲストＯＳ２２２のシステムデータがロードされることになる。

【００６３】

マイグレーションを開始する場合、ホストサーバ２００上に仮想マシン２２１が生成される。仮想マシン２２１に対しては、ＨＤＤ２０３のパーティション２３１，２３４，２３５が割り当てられる。パーティション２３１の大きさはパーティション１３１と同じでよく、パーティション２３４の大きさはパーティション１３４と同じでよい。パーティション２３１の中の起動情報領域２３２以外の領域とパーティション２３４は空でよい。

【００６４】

ホストサーバ２００は、ゲストＯＳ２２２のシステムデータをパーティション２３５に格納する。ゲストＯＳ２２２のシステムデータは、例えば、データセンタ１０が備える所定の記憶装置に用意されている。ホストサーバ２００は、所定の記憶装置からパーティション２３５にシステムデータをコピーする。所定の記憶装置は、何れかのホストサーバが備える不揮発性の記憶装置でもよいし、管理ネットワーク１２に接続されたストレージ装置でもよい。また、ホストサーバ２００は、起動パーティションとしてパーティション２３５を指定した起動情報を起動情報領域２３２に書き込む。

【００６５】

そして、ホストサーバ２００は、仮想マシン２２１を起動する。このとき、ホストサーバ２００は、起動情報領域２３２に記憶された起動情報をＲＡＭ２０２にロードし、起動情報に従って、パーティション２３５からＲＡＭ２０２にシステムデータをロードする。これにより、仮想マシン２２１上でゲストＯＳ２２２が実行される。

【００６６】

図６は、仮想マシンのマイグレーション例を示す第２の図である。

ホストサーバ２００に仮想マシン２２１が起動すると、仮想マシン１２１のゲストＯＳ１２２は、パーティション１３１に記憶されたシステムデータやパーティション１３４に記憶されたデータをホストサーバ２００に送信する。仮想マシン２２１のゲストＯＳ２２２は、受信したパーティション１３１のシステムデータをパーティション２３１に格納し、受信したパーティション１３４のデータをパーティション２３４に格納する。

【００６７】

また、仮想マシン１２１のゲストＯＳ１２２は、ＲＡＭ１０２に記憶された仮想マシン１２１のデータをホストサーバ２００に送信する。仮想マシン２２１のゲストＯＳ２２２は、受信したＲＡＭ１０２のデータをＨＤＤ２０３の退避領域２３３に格納する。パーティション１３１，１３４およびＲＡＭ１０２のデータの送信は、仮想マシン１２１と仮想マシン２２１との間の通信として行われる。よって、上記のデータ送信は、管理ネットワーク１２，２２を使用せず、ユーザネットワーク１１，２１を使用して行われる。

【００６８】

その後、仮想マシン１２１のゲストＯＳ１２２は、ホストサーバ２００に送信済のデータがホストサーバ１００において更新されたか監視する。更新された場合、仮想マシン１２１のゲストＯＳ１２２は、前回送信してからの差分データをホストサーバ２００に送信する。仮想マシン２２１のゲストＯＳ２２２は、受信した差分データをパーティション２３１，２３４の中の該当する記憶領域に上書き保存する。差分データの送信は、ホストサーバ１００における前回差分データ送信時からのデータ更新量が閾値以下に低下するまで、あるいはデータ更新量の低下が限界に達する状態になるまで、継続する。

【００６９】

図７は、仮想マシンのマイグレーション例を示す第３の図である。

ホストサーバ１００における単位時間当たりのデータ更新量が低下すると、仮想マシン１２１のゲストＯＳ１２２は、仮想マシン１２１を停止する前の最終データをホストサーバ２００に送信する。最終データには、パーティション１３１，１３４の更新データやＲＡＭ１０２の更新データの他に、レジスタ１０１ａに記憶されたＣＰＵ１０１の状態データや周辺デバイスの状態を示すデバイスデータなどが含まれる。レジスタ１０１ａのデータは、ゲストＯＳ１２２が一時的にサスペンド状態に遷移することで、退避領域１３３に格納される中断ファイルから抽出することができる。

【００７０】

仮想マシン２２１のゲストＯＳ２２２は、受信した最終データに含まれるパーティション１３１，１３４の更新データを、パーティション２３１，２３４に上書き保存する。また、仮想マシン２２１のゲストＯＳ２２２は、受信した最終データに含まれるＲＡＭ１０２の更新データを退避領域２３３に上書き保存し、受信した最終データに含まれるＣＰＵ１０１の状態データなどを退避領域２３３に格納する。

【００７１】

そして、仮想マシン２２１のゲストＯＳ２２２は、最新のＲＡＭ１０２のデータとＣＰＵ１０１の状態データから、仮想マシン１２１の状態を再現するための中断ファイルを生成し退避領域２３３に格納する。また、仮想マシン２２１のゲストＯＳ２２２は、起動情報領域２３２に記憶された起動情報を、起動時にパーティション２３５のシステムデータに代えてパーティション２３１のシステムデータがロードされるように書き換える。このとき、仮想マシン１２１のゲストＯＳ１２２はシャットダウンしてよい。

【００７２】

図８は、仮想マシンのマイグレーション例を示す第４の図である。

ホストサーバ２００で中断ファイルが生成されて起動情報が書き換えられると、仮想マシン２２１のゲストＯＳ２２２は、仮想マシン２２１を再起動するコマンドを発行する。すると、ゲストＯＳ２２２がシャットダウンし、仮想マシン２２１は起動情報領域２３２に記憶された起動情報をロードする。起動情報はパーティション２３１を指し示しているため、仮想マシン２２１はパーティション２３１からＲＡＭ２０２にシステムデータをロードする。ロードされるシステムデータは、ゲストＯＳ１２２のシステムデータである。

【００７３】

仮想マシン２２１上でゲストＯＳ１２２が起動すると、仮想マシン２２１のゲストＯＳ１２２は、退避領域２３３に記憶された中断ファイルを用いて、仮想マシン１２１が停止したときの仮想マシン１２１の状態を再現する。仮想マシン２２１のゲストＯＳ１２２は、中断ファイルに含まれるＣＰＵ１０１の状態データをＣＰＵ２０１のレジスタ２０１ａにロードする。また、仮想マシン２２１のゲストＯＳ１２２は、中断ファイルに含まれるＲＡＭ１０２のデータをＲＡＭ２０２にロードする。これにより、仮想マシン１２１が停止したときの仮想マシン１２１の処理を、仮想マシン２２１が引き継ぐことができる。

【００７４】

次に、仮想マシン１２１，２２１が有するマイグレーション機能について説明する。

図９は、仮想マシンの機能例を示すブロック図である。

仮想マシン１２１で実行されるゲストＯＳ１２２は、マイグレーション制御部１４１、更新監視部１４２およびカーネル１４３を有する。マイグレーション制御部１４１と更新監視部１４２は、例えば、ゲストＯＳ１２２に対して組み込まれたプログラムモジュールとして実装できる。上記のプログラムモジュールは、データセンタ１０によって予め組み込まれていてもよいし、仮想マシン１２１のユーザが組み込んでもよい。

【００７５】

例えば、データセンタ１０において仮想マシン１２１を生成したときに、マイグレーション制御部１４１と更新監視部１４２を有するゲストＯＳ１２２が自動的に使用されるようにしてもよい。また、仮想マシン１２１のユーザは、データセンタ１０によって用意されたオペレーティングシステムに対して上記のプログラムモジュールを追加してもよい。また、仮想マシン１２１のユーザは、仮想マシン１２１の生成後に、仮想マシン１２１が実行するオペレーティングシステムをゲストＯＳ１２２に変更してもよい。

【００７６】

マイグレーション制御部１４１は、仮想マシン１２１のマイグレーションを制御する。マイグレーション制御部１４１は、クライアント３１からユーザネットワーク１１を介してマイグレーション開始の指示を受信する。すると、マイグレーション制御部１４１は、ユーザネットワーク１１を介してデータセンタ２０の管理サーバ２３にアクセスし、データセンタ２０に移行先の仮想マシン２２１を生成させる。そして、マイグレーション制御部１４１は、仮想マシン１２１に関するＨＤＤ１０３のデータやＲＡＭ１０２のデータなどを、ユーザネットワーク１１，２１を介して仮想マシン２２１に送信する。

【００７７】

また、マイグレーション制御部１４１は、仮想マシン２２１へのデータ送信を継続している間、仮想マシン１２１で実行中のプロセスを監視する。仮想マシン１２１のプロセスによるデータ更新頻度が高い場合、マイグレーション制御部１４１は、データ更新頻度が下がるように各プロセスのＣＰＵ使用率を制限する。データ送信が完了して移行先の仮想マシン２２１が再起動すると、ゲストＯＳ１２２は仮想マシン２２１上で実行されることになる。この場合、マイグレーション制御部１４１は、マイグレーションの後処理（例えば、移行元の仮想マシン１２１の削除など）を仮想マシン２２１において実行する。

【００７８】

更新監視部１４２は、マイグレーション制御部１４１からの指示に応じて、ＨＤＤ１０３のデータやＲＡＭ１０２のデータの更新を監視する。ＨＤＤ１０３やＲＡＭ１０２へのデータの書き込みは、例えば、カーネル１４３を利用して検出できる。更新監視部１４２は、更新されたデータを示す情報をマイグレーション制御部１４１に提供する。

【００７９】
カーネル１４３は、ゲストＯＳ１２２の中核となる機能を実装したプログラムモジュールである。カーネル１４３は、ＣＰＵ１０１やＲＡＭ１０２などのハードウェアへのアクセスを制御し、ゲストＯＳ１２２内のカーネル１４３以外のプログラムモジュールに対してハードウェアアクセスに関するインタフェースを提供する。例えば、カーネル１４３は、プロセスの情報や、ＲＡＭ１０２やＨＤＤ１０３へのアクセスの情報を収集する。また、カーネル１４３は、各プロセスに対してＣＰＵ１０１の処理時間を割り当てる。
【００８０】
仮想マシン１２１は、制御情報記憶部１４４を有する。制御情報記憶部１４４は、例えば、仮想マシン１２１に割り当てられたＲＡＭ１０２の記憶領域またはＨＤＤ１０３の記憶領域として実現できる。制御情報記憶部１４４は、マイグレーション制御部１４１および更新監視部１４２がマイグレーション処理に使用する制御情報を記憶する。
【００８１】
仮想マシン２２１で実行されるゲストＯＳ２２２は、マイグレーション制御部２４１を有する。マイグレーション制御部２４１は、例えば、ゲストＯＳ２２２に対して組み込まれたプログラムモジュールとして実装できる。このプログラムモジュールは、データセンタ２０によって予め組み込まれていてもよい。例えば、マイグレーションの移行先として仮想マシン２２１が生成されたときに、マイグレーション制御部２４１を有するゲストＯＳ２２２がデータセンタ２０によって自動的に選択されるようにしてもよい。
【００８２】
マイグレーション制御部２４１は、移行元のマイグレーション制御部１４１と連携して仮想マシン１２１のマイグレーションを制御する。マイグレーション制御部２４１は、マイグレーション制御部１４１から仮想マシン１２１に関するデータを受信し、受信したデータをＨＤＤ２０３に保存する。全てのデータを受信し終えると、マイグレーション制御部２４１は、仮想マシン２２１を再起動したときに、受信したデータに基づいてゲストＯＳ１２２が実行されるように、仮想マシン２２１の起動情報を書き換える。
【００８３】
また、マイグレーション制御部２４１は、ユーザネットワーク２１を介してデータセンタ１０の管理サーバ１３にアクセスし、仮想マシン１２１を停止させる。そして、マイグレーション制御部２４１は、仮想マシン２２１を再起動する。再起動後の仮想マシン２２１では、ゲストＯＳ２２２に代えてゲストＯＳ１２２が実行されることになる。よって、移行先の仮想マシン２２１におけるマイグレーション処理は、マイグレーション制御部１４１がマイグレーション制御部２４１から引き継ぐことになる。
【００８４】
仮想マシン２２１は、制御情報記憶部２４２を有する。制御情報記憶部２４２は、例えば、仮想マシン２２１に割り当てられたＲＡＭ２０２の記憶領域またはＨＤＤ２０３の記憶領域として実現できる。制御情報記憶部２４２は、マイグレーション制御部２４１がマイグレーション処理に使用する制御情報を記憶する。なお、ゲストＯＳ１２２のシステムデータが仮想マシン１２１から仮想マシン２２１に移行すると、制御情報記憶部１４４に記憶された制御情報も仮想マシン１２１から仮想マシン２２１に移行することになる。
【００８５】
図１０は、状態テーブルの例を示す図である。
制御情報記憶部１４４は、状態テーブル１５１を記憶する。状態テーブル１５１は、パラメータ名およびフラグの項目を有する。パラメータ名の項目には、パラメータの名称として「移行済」、「ディスク更新低下」および「メモリ更新低下」が登録される。フラグの項目には、「１」（Ｔｒｕｅ）または「０」（Ｆａｌｓｅ）が登録される。
【００８６】
移行済パラメータ＝１は、仮想マシン１２１のデータを仮想マシン２２１にコピーし終えたことを示し、移行済パラメータ＝０は、データをコピーし終えていないことを示す。ディスク更新低下パラメータ＝１は、未送信であるＨＤＤ１０３の差分データの量が閾値

以下であることを示し、ディスク更新低下パラメータ＝０は、閾値より大きいことを示す。メモリ更新低下パラメータ＝１は、未送信であるＲＡＭ１０２の差分データの量が閾値以下であることを示し、メモリ更新低下パラメータ＝０は、閾値より大きいことを示す。全てのパラメータについてフラグの初期値は「０」である。

【００８７】

図１１は、ディスク更新情報とメモリ更新情報の例を示す図である。

制御情報記憶部１４４は、ディスク更新情報１５２およびメモリ更新情報１５３を記憶する。ディスク更新情報１５２は、更新監視部１４２がＨＤＤ１０３のデータの更新を監視するときに使用するものである。メモリ更新情報１５３は、更新監視部１４２がＲＡＭ１０２のデータの更新を監視するときに使用するものである。

【００８８】

ディスク更新情報１５２は、仮想マシン１２１に割り当てられたＨＤＤ１０３の記憶領域への書き込みの有無を、セクタ単位で表現したビットマップである。セクタは、例えば、ＨＤＤ１０３の記憶領域を細分化した固定長の領域である。あるセクタへデータが書き込まれ当該データが仮想マシン２２１に未送信である場合、当該セクタに対応するビットが「１」に設定される。それ以外のビットは「０」に設定される。

【００８９】

メモリ更新情報１５３は、仮想マシン１２１に割り当てられたＲＡＭ１０２の記憶領域への書き込みの有無を、ページ単位で表現したビットマップである。ページは、例えば、ＲＡＭ１０２の記憶領域を細分化した固定長の領域である。あるページへデータが書き込まれ当該データが仮想マシン２２１に未送信である場合、当該ページに対応するビットが「１」に設定される。それ以外のビットは「０」に設定される。

【００９０】

図１２は、プロセステーブルの例を示す図である。

制御情報記憶部１４４は、プロセステーブル１５４を記憶する。プロセステーブル１５４は、マイグレーション制御部１４１が各プロセスのＣＰＵ使用率を制限するときに使用するものである。プロセステーブル１５４は、プロセスＩＤ、プロセス名、開始状態、状態、ＣＰＵ使用率、ディスク更新量およびメモリ更新量の項目を有する。

【００９１】

プロセスＩＤの項目には、カーネル１４３によって付与されたプロセスの識別番号が登録される。プロセス名の項目には、プロセスの名称（例えば、当該プロセスの処理を定義したプログラムの名称など）が登録される。開始状態の項目には、仮想マシン１２１のマイグレーションの開始時点における各プロセスの状態が登録される。開始状態としては、「実行中」または「一時停止」が挙げられる。状態の項目には、現在の各プロセスの状態が登録される。状態としては、「実行中」、「一時停止」または「終了」が挙げられる。

【００９２】

ＣＰＵ使用率の項目には、カーネル１４３が各プロセスに対して現在割り当てているＣＰＵ時間（リソース量）の指標値が登録される。ＣＰＵ時間は、例えば、１つのプロセッサまたはプロセッサコアの演算能力を１００としたときの百分率で表現できる。ディスク更新量の項目には、各プロセスによる単位時間当たりのＨＤＤ１０３への書き込み量が登録される。メモリ更新量の項目には、各プロセスによる単位時間当たりのＲＡＭ１０２への書き込み量が登録される。プロセステーブル１５４に登録される情報は、例えば、カーネル１４３のインタフェースを用いて取得することができる。

【００９３】

図１３は、リソース制限テーブルと重要プロセスリストの例を示す図である。

制御情報記憶部１４４は、リソース制限テーブル１５５および重要プロセスリスト１５６を記憶する。リソース制限テーブル１５５および重要プロセスリスト１５６は、マイグレーション制御部１４１が各プロセスのＣＰＵ使用率を制限するときに使用するものである。リソース制限テーブル１５５および重要プロセスリスト１５６は、例えば、仮想マシン１２１のユーザによって作成されて制御情報記憶部１４４に格納される。

【００９４】
リソース制限テーブル１５５は、プロセス名およびリソース上限の項目を有する。プロセス名の項目には、プロセステーブル１５４と同様のプロセスの名称が登録される。リソース上限の項目には、各プロセスについてＣＰＵ使用率を最も厳しく制限した場合のＣＰＵ使用率の上限が登録される。例えば、図１３の例ではプロセスＢのリソース上限が２０％に設定されている一方、図１２の例ではプロセスＢの現在のＣＰＵ使用率は２５％である。よって、ＣＰＵ使用率を制限する場合、プロセスＢに対するＣＰＵ時間の割り当てを減らすことにより、プロセスＢのＣＰＵ使用率が２０％以下に低下し得る。
【００９５】
重要プロセスリスト１５６には、１または２以上のプロセス名が登録される。重要プロセスリスト１５６が示すプロセスは、仮想マシン１２１のユーザにとって重要なプロセスであり、ＣＰＵ時間の割り当てを減らさない（ＣＰＵ使用率を制限する対象から除外する）プロセスを示す。例えば、重要プロセスリスト１５６には、プロセスＡ，Ｂ，Ｃ，Ｄ，Ｅのうちプロセス Ａ，Ｃが登録される。この場合、プロセスＢ，Ｄ，ＥのＣＰＵ使用率は制限される可能性がある一方、プロセスＡ，ＣのＣＰＵ使用率は制限されない。
【００９６】
図１４は、仮想マシン生成要求の例を示す図である。
仮想マシン生成要求１５７は、仮想マシン１２１のマイグレーション制御部１４１からデータセンタ２０の管理サーバ２３に対して送信される。仮想マシン生成要求１５７は、移行先の仮想マシン２２１の生成の要求を示す。仮想マシン生成要求１５７は、パラメータ名と設定値の組を複数含む。パラメータには、ＣＰＵ数、メモリ容量、システムディスク容量、他データディスク容量、ネットワーク設定および初期ＯＳが含まれる。
【００９７】
「ＣＰＵ数」は、仮想マシン２２１に割り当てるＣＰＵの数を示す。通常、ＣＰＵ数は仮想マシン１２１と同じでよい。「メモリ容量」は、仮想マシン２２１に割り当てるＲＡＭ２０２の記憶領域の大きさを示す。通常、メモリ容量は仮想マシン１２１と同じでよい。「システムディスク容量」は、仮想マシン２２１に割り当てるＨＤＤ２０３のパーティション２３１，２３５の大きさを示す。通常、システムディスク容量は、仮想マシン１２１のパーティション１３１の大きさに、ゲストＯＳ２２２のシステムデータを記憶するパーティション２３５の大きさを加えたものである。なお、システムディスクのフォーマットは、仮想マシン１２１のパーティション１３１と同じでよい。
【００９８】
「他データディスク容量」は、仮想マシン２２１に割り当てるＨＤＤ２０３のパーティション２３４の大きさを示す。通常、他データディスク容量は、仮想マシン１２１のパーティション１３４と同じでよい。なお、他データディスクのフォーマットは、仮想マシン１２１のパーティション１３４と同じでよい。「ネットワーク設定」は、仮想マシン２２１が使用可能なネットワークを示す。通常、ネットワーク設定は仮想マシン１２１と同じでよい。ただし、仮想マシン２２１は仮想マシン１２１と通信するため、仮想マシン１２１が使用するネットワークにインターネット３０が含まれていない場合、インターネット３０を追加登録する（この場合、仮想マシン１２１の設定も変更する）。
【００９９】
「初期ＯＳ」は、仮想マシン２２１を最初に起動したときに実行されるオペレーティングシステムの種類を示す。初期ＯＳとしてマイグレーション用ＯＳを指定した場合、マイグレーション制御部２４１を有するゲストＯＳ２２２のシステムデータがパーティション２３５に格納される。仮想マシン２２１の起動時には、ゲストＯＳ２２２が実行される。
【０１００】
次に、仮想マシン１２１，２２１が行うマイグレーション処理について説明する。
図１５は、マイグレーション開始の手順例を示すフローチャートである。
クライアント３１は、インターネット３０およびユーザネットワーク１１を介して、仮想マシン１２１に対してマイグレーション開始指示を送信する（Ｓ１１０）。

【０１０１】
マイグレーション開始指示を受信すると、仮想マシン１２１上のマイグレーション制御部１４１は、ユーザネットワーク１１およびインターネット３０を介して、データセンタ２０の管理サーバ２３に仮想マシン生成要求１５７を送信する（Ｓ１１１）。このとき、マイグレーション制御部１４１は、仮想マシン１２１のリソース割り当て状況から、仮想マシン生成要求１５７の各パラメータの設定値を算出する。原則として、仮想マシン２２１の構成は、パーティション２３５を除いて仮想マシン１２１と同じでよい。
【０１０２】
仮想マシン生成要求１５７を受信すると、管理サーバ２３は、仮想マシン生成要求１５７で指定された量のリソースを仮想マシン２２１に割り当てることで、ホストサーバ２００上に仮想マシン２２１を生成する（Ｓ１１２）。管理サーバ２３は、仮想マシン２２１に対して付与されたアドレス（例えば、ＩＰ（Internet Protocol）アドレス）を、インターネット３０およびユーザネットワーク１１を介して仮想マシン１２１に通知する（Ｓ１１３）。マイグレーション制御部１４１は、制御情報記憶部１４４に状態テーブル１５１を生成し、全てのフラグを「０」に初期化する（Ｓ１１４）。
【０１０３】
マイグレーション制御部１４１は、ユーザネットワーク１１およびインターネット３０を介して管理サーバ２３に、ステップＳ１１２で生成された仮想マシンを指定した仮想マシン起動要求を送信する（Ｓ１１５）。仮想マシン起動要求を受信すると、管理サーバ２３は、指定された仮想マシン２２１が配置されたホストサーバ２００を特定し、管理ネットワーク１２を介して、ホストサーバ２００のハイパーバイザに対して仮想マシン起動指示を送信する（Ｓ１１６）。ホストサーバ２００のハイパーバイザは、仮想マシン２２１を起動する。このとき、起動情報領域２３２に記憶された起動情報に従って、パーティション２３５に記憶されたシステムデータがロードされてゲストＯＳ２２２が実行される。そして、マイグレーション制御部２４１が起動し、接続待ち状態になる（Ｓ１１７）。
【０１０４】
マイグレーション制御部１４１は、ステップＳ１１３で通知されたアドレスを用いて、ユーザネットワーク１１およびインターネット３０を介して、仮想マシン２２１に対して接続要求を送信する（Ｓ１１８）。マイグレーション制御部２４１は、接続処理を行い、仮想マシン１２１に対して接続完了通知を返信する（Ｓ１１９）。マイグレーション制御部１４１は、以下で説明するディスクデータ送信、メモリデータ送信およびプロセス監視の３つのプロセスを起動する（Ｓ１２０）。この３つのプロセスは並行に実行される。
【０１０５】
図１６は、ディスクデータ送信の手順例を示すフローチャートである。
このディスクデータ送信のプロセスは、上記のステップＳ１２０で起動される。
マイグレーション制御部１４１は、更新監視部１４２にディスク更新情報１５２の初期化を指示する（Ｓ２１０）。更新監視部１４２は、制御情報記憶部１４４に記憶されたディスク更新情報１５２の全てのビットを「０」に初期化する（Ｓ２１１）。更新監視部１４２は、ビットマップ初期化完了をマイグレーション制御部１４１に通知する。マイグレーション制御部１４１は、ディスク書き込みの監視指示を更新監視部１４２に通知する（Ｓ２１２）。更新監視部１４２は、ディスク監視のプロセスを起動することで、ディスク書き込みの監視を開始する（Ｓ２１３）。ディスク監視のプロセスは、図１６のプロセスなど他のプロセスと並行に実行される（Ｓ２１４）。ディスク監視の処理は後述する。
【０１０６】
更新監視部１４２は、ディスク監視開始をマイグレーション制御部１４１に通知する。マイグレーション制御部１４１は、ユーザネットワーク１１、インターネット３０およびユーザネットワーク２１を介して、ＨＤＤ１０３のパーティション１３１，１３４に記憶されたデータを、セクタ単位で仮想マシン２２１に送信する（Ｓ２１５）。このとき、マイグレーション制御部１４１は、ディスク更新情報１５２のビットが「１」になっているセクタのデータを送信しなくてもよい。データの送信時には、セクタアドレスが付加され

る。マイグレーション制御部２４１は、受信したデータを、パーティション２３１，２３４の中のセクタアドレスが示す記憶領域に保存する（Ｓ２１６）。

【０１０７】

マイグレーション制御部２４１は、データ受信完了を仮想マシン１２１に返信する。パーティション１３１，１３４全体について初回データ送信が終わると、マイグレーション制御部１４１は、更新監視部１４２にディスク更新を問い合わせる（Ｓ２１７）。更新監視部１４２は、ディスク更新情報１５２のビットが「１」になっているセクタ、すなわち、データ書き込みのあったセクタを特定し、特定したセクタのセクタアドレスをマイグレーション制御部１４１に通知する（Ｓ２１８）。

【０１０８】

マイグレーション制御部１４１は、データ書き込みのあったセクタ数またはセクタ数に応じたデータ量（ディスク更新量）が閾値以下であるか判断する（Ｓ２１９）。ディスク更新量が閾値以下である場合、状態テーブル１５１のディスク更新低下フラグを「１」に設定し、ステップＳ２２６に処理が進む。ディスク更新量が閾値より大きい場合、ディスク更新低下フラグを「０」に設定し、ステップＳ２２０に処理が進む。

【０１０９】

ディスク更新量が閾値より大きい場合、マイグレーション制御部１４１は、更新監視部１４２にディスク更新情報１５２の初期化を指示する（Ｓ２２０）。更新監視部１４２は、ディスク更新情報１５２の全てのビットを「０」に初期化する（Ｓ２２１）。更新監視部１４２は、ビットマップ初期化完了をマイグレーション制御部１４１に通知する。

【０１１０】

マイグレーション制御部１４１は、ステップＳ２１８または後述するステップＳ２２５で通知されたセクタアドレスが示すデータ（更新されたデータ）を、ディスク差分データとしてパーティション１３１，１３４から読み出す。マイグレーション制御部１４１は、ディスク差分データにセクタアドレスを付加し、ユーザネットワーク１１、インターネット３０およびユーザネットワーク２１を介して仮想マシン２２１に送信する（Ｓ２２２）。マイグレーション制御部２４１は、受信したデータを、パーティション２３１，２３４の中のセクタアドレスが示す記憶領域に上書き保存する（Ｓ２２３）。

【０１１１】

マイグレーション制御部２４１は、データ受信完了を仮想マシン１２１に返信する。マイグレーション制御部１４１は、更新監視部１４２にディスク更新を問い合わせる（Ｓ２２４）。更新監視部１４２は、ディスク更新情報１５２のビットが「１」になっているセクタ、すなわち、前回の確認以後にデータ書き込みのあったセクタを特定し、特定したセクタのセクタアドレスをマイグレーション制御部１４１に通知する（Ｓ２２５）。

【０１１２】

マイグレーション制御部１４１は、ディスク更新量が閾値以下である場合、状態テーブル１５１のディスク更新低下フラグを「１」に設定し、ディスク更新量が閾値より大きい場合、ディスク更新低下フラグを「０」に設定する。そして、マイグレーション制御部１４１は、状態テーブル１５１のディスク更新低下フラグとメモリ更新低下フラグの両方が「１」であるか判断する（Ｓ２２６）。２つのフラグが「１」である場合、ディスクデータ送信のプロセスが終了し、後述するマイグレーション完了のプロセスが起動する。少なくとも一方のフラグが「０」である場合、ステップＳ２２０に処理が進む。

【０１１３】

図１７は、ディスク監視の手順例を示すフローチャートである。

このディスク監視のプロセスは、上記のステップＳ２１４で起動される。

更新監視部１４２は、ＨＤＤ１０３に対する書き込みが発生したときに当該書き込みが検出されるよう、カーネル１４３のインタフェースを利用してフック設定を行う（Ｓ２３０）。更新監視部１４２は、ＨＤＤ１０３への書き込みが検出されると、書き込み先のセクタを特定する（Ｓ２３１）。更新監視部１４２は、ディスク更新情報１５２の中のステップＳ２３１で特定したセクタに対応するビットを「１」に設定する（Ｓ２３２）。

【０１１４】
更新監視部１４２は、状態テーブル１５１のディスク更新低下フラグとメモリ更新低下フラグの両方が「１」であるか判断する（Ｓ２３３）。２つのフラグが「１」である場合、ディスク監視のプロセスが終了する。少なくとも一方のフラグが「０」である場合、ステップＳ２３１に処理が進み、次にＨＤＤ１０３への書き込みが検出されるのを待つ。
【０１１５】
図１８は、メモリデータ送信の手順例を示すフローチャートである。
このメモリデータ送信のプロセスは、上記のステップＳ１２０で起動される。
マイグレーション制御部１４１は、更新監視部１４２にメモリ更新情報１５３の初期化を指示する（Ｓ２４０）。更新監視部１４２は、制御情報記憶部１４４に記憶されたメモリ更新情報１５３の全てのビットを「０」に初期化する（Ｓ２４１）。更新監視部１４２は、ビットマップ初期化完了をマイグレーション制御部１４１に通知する。マイグレーション制御部１４１は、メモリ書き込みの監視指示を更新監視部１４２に通知する（Ｓ２４２）。更新監視部１４２は、メモリ監視のプロセスを起動することで、メモリ書き込みの監視を開始する（Ｓ２４３）。メモリ監視のプロセスは、図１８のプロセスなど他のプロセスと並行に実行される（Ｓ２４４）。メモリ監視の処理は後述する。
【０１１６】
更新監視部１４２は、メモリ監視開始をマイグレーション制御部１４１に通知する。マイグレーション制御部１４１は、ユーザネットワーク１１、インターネット３０およびユーザネットワーク２１を介して、仮想マシン１２１に割り当てられたＲＡＭ１０２の記憶領域に記憶されたデータを、ページ単位で仮想マシン２２１に送信する（Ｓ２４５）。このとき、マイグレーション制御部１４１は、メモリ更新情報１５３のビットが「１」になっているページのデータを送信しなくてもよい。データの送信時には、ページアドレスが付加される。マイグレーション制御部２４１は、受信したデータを、ＲＡＭ２０２の中のページアドレスが示す記憶領域に保存する（Ｓ２４６）。
【０１１７】
マイグレーション制御部２４１は、データ受信完了を仮想マシン１２１に返信する。仮想マシン１２１に割り当てられたＲＡＭ１０２の記憶領域全体について初回データ送信が終わると、マイグレーション制御部１４１は、更新監視部１４２にメモリ更新を問い合わせる（Ｓ２４７）。更新監視部１４２は、メモリ更新情報１５３のビットが「１」になっているページ、すなわち、データ書き込みのあったページを特定し、特定したページのページアドレスをマイグレーション制御部１４１に通知する（Ｓ２４８）。
【０１１８】
マイグレーション制御部１４１は、データ書き込みのあったページ数またはページ数に応じたデータ量（メモリ更新量）が閾値以下であるか判断する（Ｓ２４９）。メモリ更新量が閾値以下である場合、状態テーブル１５１のメモリ更新低下フラグを「１」に設定し、ステップＳ２５６に処理が進む。メモリ更新量が閾値より大きい場合、メモリ更新低下フラグを「０」に設定し、ステップＳ２５０に処理が進む。
【０１１９】
メモリ更新量が閾値より大きい場合、マイグレーション制御部１４１は、更新監視部１４２にメモリ更新情報１５３の初期化を指示する（Ｓ２５０）。更新監視部１４２は、メモリ更新情報１５３の全てのビットを「０」に初期化する（Ｓ２５１）。更新監視部１４２は、ビットマップ初期化完了をマイグレーション制御部１４１に通知する。
【０１２０】
マイグレーション制御部１４１は、ステップＳ２４８または後述するステップＳ２５５で通知されたメモリアドレスが示すデータ（更新されたデータ）を、メモリ差分データとしてＲＡＭ１０２から読み出す。マイグレーション制御部１４１は、メモリ差分データにページアドレスを付加し、ユーザネットワーク１１、インターネット３０およびユーザネットワーク２１を介して仮想マシン２２１に送信する（Ｓ２５２）。マイグレーション制御部２４１は、受信したデータを、ＲＡＭ２０２の中のページアドレスが示す記憶領域に

上書き保存する（Ｓ２５３）。
【０１２１】
マイグレーション制御部２４１は、データ受信完了を仮想マシン１２１に返信する。マイグレーション制御部１４１は、更新監視部１４２にメモリ更新を問い合わせる（Ｓ２５４）。更新監視部１４２は、メモリ更新情報１５３のビットが「１」になっているページ、すなわち、前回の確認以後にデータ書き込みのあったページを特定し、特定したページのページアドレスをマイグレーション制御部１４１に通知する（Ｓ２５５）。
【０１２２】
マイグレーション制御部１４１は、メモリ更新量が閾値以下である場合、状態テーブル１５１のメモリ更新低下フラグを「１」に設定し、メモリ更新量が閾値より大きい場合、メモリ更新低下フラグを「０」に設定する。そして、マイグレーション制御部１４１は、状態テーブル１５１のディスク更新低下フラグとメモリ更新低下フラグの両方が「１」であるか判断する（Ｓ２５６）。２つのフラグが「１」である場合、メモリデータ送信のプロセスが終了し、後述するマイグレーション完了のプロセスが起動する。少なくとも一方のフラグが「０」である場合、ステップＳ２５０に処理が進む。
【０１２３】
図１９は、メモリ監視の手順例を示すフローチャートである。
このメモリ監視のプロセスは、上記のステップＳ２４４で起動される。
更新監視部１４２は、ＲＡＭ１０２に対する書き込みが発生したときに当該書き込みが検出されるよう、カーネル１４３のインタフェースを利用してフック設定を行う（Ｓ２６０）。更新監視部１４２は、ＲＡＭ１０２への書き込みが検出されると、書き込み先のページを特定する（Ｓ２６１）。更新監視部１４２は、メモリ更新情報１５３の中のステップＳ２６１で特定したページに対応するビットを「１」に設定する（Ｓ２６２）。
【０１２４】
更新監視部１４２は、状態テーブル１５１のディスク更新低下フラグとメモリ更新低下フラグの両方が「１」であるか判断する（Ｓ２６３）。２つのフラグが「１」である場合、メモリ監視のプロセスが終了する。少なくとも一方のフラグが「０」である場合、ステップＳ２６１に処理が進み、次にＲＡＭ１０２への書き込みが検出されるのを待つ。
【０１２５】
図２０は、プロセス監視の手順例を示すフローチャートである。
このプロセス監視のプロセスは、上記のステップＳ１２０で起動される。
マイグレーション制御部１４１は、ゲストＯＳ１２２上に存在するプロセスのリストを、カーネル１４３のインタフェースを利用して取得する（Ｓ２７０）。プロセスのリストには、プロセスＩＤ、プロセス名および状態（実行中または一時停止中）が含まれる。マイグレーション制御部１４１は、取得したプロセスＩＤとプロセス名をプロセステーブル１５４に登録し、取得した状態を開始状態としてプロセステーブル１５４に登録する。
【０１２６】
マイグレーション制御部１４１は、各プロセスについて、現在の状態、ＣＰＵ使用率、ディスク更新量およびメモリ更新量を、カーネル１４３のインタフェースを利用して取得する（Ｓ２７１）。マイグレーション制御部１４１は、取得した状態、ＣＰＵ使用率、ディスク更新量およびメモリ更新量をプロセステーブル１５４に登録する。
【０１２７】
マイグレーション制御部１４１は、プロセステーブル１５４に登録されたプロセスの中から、ディスク更新に関する以下の条件を満たすプロセスを高々１つ選択する（Ｓ２７２）。条件を満たすプロセスは、重要プロセスリスト１５６に登録されておらず、ＣＰＵ使用率がリソース制限テーブル１５５に登録されたリソース上限を超えているプロセスの中で、単位時間当たりのディスク更新量が最大のものである。
【０１２８】
マイグレーション制御部１４１は、プロセステーブル１５４に登録されたプロセスの中から、メモリ更新に関する以下の条件を満たすプロセスを高々１つ選択する（Ｓ２７３）

。条件を満たすプロセスは、重要プロセスリスト１５６に登録されておらず、ＣＰＵ使用率がリソース制限テーブル１５５に登録されたリソース上限を超えているプロセスの中で、単位時間当たりのメモリ更新量が最大のものである。

【０１２９】

マイグレーション制御部１４１は、ステップＳ２７２，Ｓ２７３で選択したプロセスがある場合、カーネル１４３のインタフェースを利用して、当該プロセスのＣＰＵ使用率の上限をリソース制限テーブル１５５に登録されたリソース上限に設定する（Ｓ２７４）。これにより、選択したプロセスに対するＣＰＵ時間の割り当てが減少し、単位時間当たりのディスク更新量またはメモリ更新量が小さくなると期待できる。

【０１３０】

マイグレーション制御部１４１は、各プロセスについて、現在の状態、ＣＰＵ使用率、ディスク更新量およびメモリ更新量を、カーネル１４３のインタフェースを利用して取得する（Ｓ２７５）。マイグレーション制御部１４１は、取得した状態、ＣＰＵ使用率、ディスク更新量およびメモリ更新量をプロセステーブル１５４に上書き保存する。マイグレーション制御部１４１は、ゲストＯＳ１２２全体の単位時間当たりのディスク更新量とメモリ更新量を、カーネル１４３のインタフェースを利用して取得する（Ｓ２７６）。ただし、マイグレーション制御部１４１は、ステップＳ２７５で取得したプロセス毎のディスク更新量とメモリ更新量から、ゲストＯＳ１２２全体の値を算出してもよい。

【０１３１】

図２１は、プロセス監視の手順例を示すフローチャート（続き）である。

マイグレーション制御部１４１は、ステップＳ２７６で取得した全体ディスク更新量が閾値以下であるか判断する（Ｓ２７７）。全体ディスク更新量が閾値以下の場合はステップＳ２７８に処理が進み、閾値より大きい場合はステップＳ２７９に処理が進む。全体ディスク更新量が閾値以下の場合、マイグレーション制御部１４１は、状態テーブル１５１のディスク更新低下フラグを「１」に設定する（Ｓ２７８）。そして、ステップＳ２８０に処理が進む。全体ディスク更新量が閾値より大きい場合、マイグレーション制御部１４１は、状態テーブル１５１のディスク更新低下フラグを「０」に設定する（Ｓ２７９）。

【０１３２】

マイグレーション制御部１４１は、ステップＳ２７６で取得した全体メモリ更新量が閾値以下であるか判断する（Ｓ２８０）。全体メモリ更新量が閾値以下の場合はステップＳ２８１に処理が進み、閾値より大きい場合はステップＳ２８２に処理が進む。全体メモリ更新量が閾値以下の場合、マイグレーション制御部１４１は、状態テーブル１５１のメモリ更新低下フラグを「１」に設定する（Ｓ２８１）。そして、ステップＳ２８３に処理が進む。全体メモリ更新量が閾値より大きい場合、マイグレーション制御部１４１は、状態テーブル１５１のメモリ更新低下フラグを「０」に設定する（Ｓ２８２）。

【０１３３】

なお、ステップＳ２７７の閾値は、例えば、ユーザまたはデータセンタ１０の管理者によって予め定義される。ステップＳ２７７の閾値は、前述のステップＳ２１９，Ｓ２２６で使用する閾値と同じでもよい。また、ステップＳ２８０の閾値は、例えば、ユーザまたはデータセンタ１０の管理者によって予め定義される。ステップＳ２８０の閾値は、前述のステップＳ２４９，Ｓ２５６で使用する閾値と同じでもよい。

【０１３４】

マイグレーション制御部１４１は、更にＣＰＵ使用率を制限可能なプロセスが存在するか判断する（Ｓ２８３）。更にＣＰＵ使用率を制限可能なプロセスは、重要プロセスリスト１５６に登録されておらず、かつ、現在のＣＰＵ使用率がリソース制限テーブル１５５に登録されたリソース上限を超えているプロセスである。該当するプロセスが存在する場合はステップＳ２８５に処理が進み、存在しない場合はステップＳ２８４に処理が進む。

【０１３５】

マイグレーション制御部１４１は、状態テーブル１５１のディスク更新低下フラグとメモリ更新低下フラグの両方を「１」に設定する（Ｓ２８４）。これは、単位時間当たりの

ディスク更新量またはメモリ更新量が依然として大きいものの、差分データの送信を打ち切って後述するマイグレーション完了のプロセスを開始することを意味する。

【０１３６】

マイグレーション制御部１４１は、ディスク更新低下フラグとメモリ更新低下フラグの両方が「１」であるか判断する（Ｓ２８５）。２つのフラグが「１」である場合は、プロセス監視のプロセスが終了し、少なくとも一方のフラグが「０」である場合はステップＳ２７２に処理が進む。なお、２つのフラグが「１」になると、前述のステップＳ２２６，Ｓ２５６の判断が「ＹＥＳ」になり、ディスクデータ送信とメモリデータ送信のプロセスも終了する。また、後述するマイグレーション完了のプロセスが開始される。

【０１３７】

図２２は、マイグレーション完了の手順例を示すフローチャートである。

マイグレーション制御部１４１は、カーネル１４３に対してサスペンド状態に遷移するよう指示する（Ｓ１２１）。サスペンド状態では、ゲストＯＳ１２２がシャットダウンせずに、ＣＰＵ１０１による仮想マシン１２１のプログラムの実行が停止する。

【０１３８】

サスペンド指示を受けて、カーネル１４３は、現在の仮想マシン１２１の状態を示す中断ファイルを生成し、ＨＤＤ１０３の退避領域１３３に保存する（Ｓ１２２）。中断ファイルには、ＣＰＵ１０１のレジスタ１０１ａに記憶されている状態データ、ＲＡＭ１０２に記憶されているメモリデータ、その他の周辺デバイスの状態を示すデバイスデータなどが含まれる。また、ゲストＯＳ１２２は、ディスク更新情報１５２を参照して、仮想マシン２２１に未送信の差分データを含む記憶領域のスナップショットを作成しておく。また、カーネル１４３は、すぐにサスペンド状態を解除するよう設定しておく。

【０１３９】

カーネル１４３は、サスペンド状態に遷移した後、ステップＳ１２２の設定に従ってサスペンド状態を解除する（Ｓ１２３）。このとき、ステップＳ１２２で作成した中断ファイルを用いて、サスペンド直前のＣＰＵ１０１の状態などが復元される。これにより、ゲストＯＳ１２２内のマイグレーション制御部１４１の処理が再開する。

【０１４０】

マイグレーション制御部１４１は、３つの監視処理を停止させる（Ｓ１２４）。３つの監視処理には、更新監視部１４２によるディスク監視、更新監視部１４２によるメモリ監視、マイグレーション制御部１４１によるプロセス監視が含まれる。

【０１４１】

マイグレーション制御部１４１は、更新監視部１４２にメモリ更新を問い合わせ、更新されたページのページアドレスを取得する。そして、マイグレーション制御部１４１は、退避領域１３３に記憶された中断ファイルから、更新されたページのメモリデータを抽出する。また、マイグレーション制御部１４１は、中断ファイルから、ＣＰＵ１０１の状態データや周辺デバイスのデバイスデータなどを抽出する。また、マイグレーション制御部１４１は、ステップＳ１２２で作成されたスナップショットから、更新されたセクタのディスクデータを抽出する。マイグレーション制御部１４１は、抽出した上記のデータを含む最終データを、ユーザネットワーク１１、インターネット３０およびユーザネットワーク２１を介して仮想マシン２２１に送信する（Ｓ１２５）。

【０１４２】

マイグレーション制御部１４１は、状態テーブル１５１の移行済フラグを「１」に設定する（Ｓ１２６）。マイグレーション制御部２４１は、仮想マシン１２１から受信した最終データに含まれる各種データを、ＨＤＤ２０３の適切な記憶領域に保存する（Ｓ１２７）。最終データに含まれるディスクデータは、セクタアドレスが示すＨＤＤ２０３内の記憶領域に上書き保存される。最終データに含まれるメモリデータ、ＣＰＵ１０１の状態データ、周辺デバイスのデバイスデータなどは、退避領域２３３に保存される。

【０１４３】

マイグレーション制御部２４１は、退避領域２３３に記憶されたデータに基づいて、仮

想マシン１２１の状態を再現するための中断ファイルを生成する（Ｓ１２８）。マイグレーション制御部２４１は、制御情報記憶部２４２に状態テーブル（状態テーブル１５１と同様のもの）を生成し、移行済フラグを「１」に設定する（Ｓ１２９）。

【０１４４】

マイグレーション制御部２４１は、ユーザネットワーク２１およびインターネット３０を介して、データセンタ１０の管理サーバ１３に対して仮想マシン１２１の停止要求を送信する（Ｓ１３０）。停止要求を受信すると、管理サーバ１３は、管理ネットワーク１２を介して仮想マシン１２１に停止指示を送信する（Ｓ１３１）。仮想マシン１２１は、ゲストＯＳ１２２をシャットダウンさせて仮想マシン１２１を停止させる（Ｓ１３２）。

【０１４５】

マイグレーション制御部２４１は、起動情報領域２３２に記憶された起動情報を編集する（Ｓ１３３）。編集された起動情報は、仮想マシン２２１の起動時にパーティション２３１からシステムデータがロードされることを示す。また、マイグレーション制御部２４１は、退避領域２３３に記憶された中断ファイルを用いて途中状態から処理が開始されるように設定しておく。マイグレーション制御部２４１は、ゲストＯＳ２２２のカーネルに再起動を指示する（Ｓ１３４）。ゲストＯＳ２２２がシャットダウンして仮想マシン２２１が再起動すると、起動情報に従ってパーティション２３１からシステムデータがロードされる。これにより、仮想マシン２２１上ではゲストＯＳ１２２が実行される。また、退避領域２３３に記憶された中断ファイルに基づいて、移行元のＣＰＵ１０１やＲＡＭ１０２の状態が、ＣＰＵ２０１やＲＡＭ２０２上に再現される。

【０１４６】

図２３は、マイグレーション完了の手順例を示すフローチャート（続き）である。

仮想マシン２２１上で起動したマイグレーション制御部１４１は、制御情報記憶部２４２に記憶された状態テーブルを参照し、移行済フラグが「１」であるか判断する（Ｓ１３５）。移行済フラグが「１」の場合はステップＳ１３６に処理が進み、移行済フラグが「０」の場合はマイグレーション制御部１４１の処理が終了する。

【０１４７】

移行済フラグが「１」の場合、マイグレーション制御部１４１は、カーネル１４３のインタフェースを利用して、前述のステップＳ２７４で設定したＣＰＵ使用率の上限を解除する（Ｓ１３６）。マイグレーション制御部１４１は、制御情報記憶部１４４に記憶された状態テーブル１５１などの制御情報を削除する（Ｓ１３７）。マイグレーション制御部１４１は、ゲストＯＳ２２２のシステムデータが記憶されたパーティション２３５を、ゲストＯＳ１２２から認識されないように解放する（Ｓ１３８）。

【０１４８】

マイグレーション制御部１４１は、ユーザネットワーク２１およびインターネット３０を介して、データセンタ１０の管理サーバ１３に対して仮想マシン１２１の削除要求を送信する（Ｓ１３９）。削除要求を受信すると、管理サーバ１３は、仮想マシン１２１に対するホストサーバ１００のリソースの割り当てを解除することで、データセンタ１０から仮想マシン１２１を削除する（Ｓ１４０）。これにより、パーティション１３１，１３４が解放されて他の仮想マシンが使用可能になる。

【０１４９】

第２の実施の形態の情報処理システムによれば、移行先のホストサーバ２００で仮想マシン２２１が起動し、仮想マシン２２１上でゲストＯＳ２２２が一時的に実行される。仮想マシン１２１に関するデータのコピーは、仮想マシン１２１と仮想マシン２２１との間の通信として行うことができる。よって、データセンタ１０の外部からの管理ネットワーク１２への直接アクセスやデータセンタ２０の外部からの管理ネットワーク２２への直接アクセスが制限されていても、管理ネットワーク１２，２２を迂回しユーザネットワーク１１，２１を用いて仮想マシン１２１に関するデータをコピーできる。

【０１５０】

そして、仮想マシン２２１の起動情報を書き換えることで、仮想マシン２２１が再起動

すると、ゲストＯＳ２２２に代えてゲストＯＳ１２２が仮想マシン２２１上で実行される。これにより、異なるサービス事業者が管理するデータセンタ１０，２０の間でも、仮想マシン１２１のマイグレーションを円滑に行うことができる。

【０１５１】

また、移行元の仮想マシン１２１においてメモリデータやディスクデータの更新頻度が高い場合、少なくとも一部のプロセスのＣＰＵ使用率を制限することで、メモリデータやディスクデータの更新頻度が低下する。これにより、送信する差分データの量を低減することができ、インターネット３０の通信遅延が大きい場合であっても、仮想マシン１２１のマイグレーションに要する時間を短縮することができる。また、ＣＰＵ使用率を制限しない重要プロセスを定義しておくことで、業務への影響を軽減できる。

【０１５２】

なお、前述のように、第１の実施の形態の情報処理は、情報処理装置１，２にプログラムを実行させることで実現することができる。また、第２の実施の形態の情報処理は、管理サーバ１３，２３、クライアント３１、ホストサーバ１００，１００ａ，２００，２００ａなどにプログラムを実行させることで実現することができる。

【０１５３】

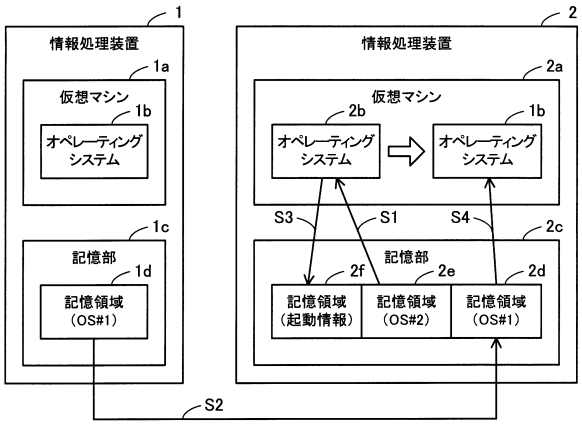
プログラムは、コンピュータ読み取り可能な記録媒体（例えば、記録媒体１１３）に記録しておくことができる。記録媒体としては、例えば、磁気ディスク、光ディスク、光磁気ディスク、半導体メモリなどを使用できる。磁気ディスクには、ＦＤおよびＨＤＤが含まれる。光ディスクには、ＣＤ、ＣＤ－Ｒ（Recordable）／ＲＷ（Rewritable）、ＤＶＤおよびＤＶＤ－Ｒ／ＲＷが含まれる。プログラムは、可搬型の記録媒体に記録されて配布されることがある。その場合、可搬型の記録媒体からＨＤＤなどの他の記録媒体（例えば、ＨＤＤ１０３）にプログラムをコピーして実行してもよい。

【符号の説明】

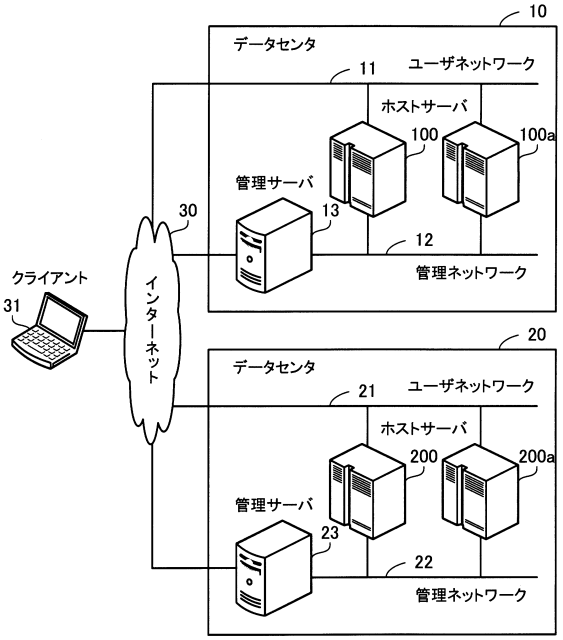
【０１５４】

１，２　情報処理装置
１ａ，２ａ　仮想マシン
１ｂ，２ｂ　オペレーティングシステム
１ｃ，２ｃ　記憶部
１ｄ，２ｄ，２ｅ，２ｆ　記憶領域

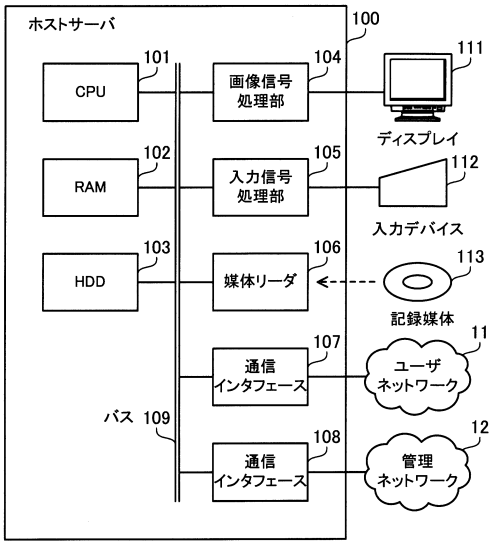
【図１】


1
情報処理装置
1a
仮想マシン
1b
オペレーティングシステム
1c
記憶部
1d
記憶領域（OS#1）
2
情報処理装置
2a
仮想マシン
2b
オペレーティングシステム
1b
オペレーティングシステム
S3
S1
S4
2c
記憶部
2f
2e
2d
記憶領域（起動情報）
記憶領域（OS#2）
記憶領域（OS#1）
S2

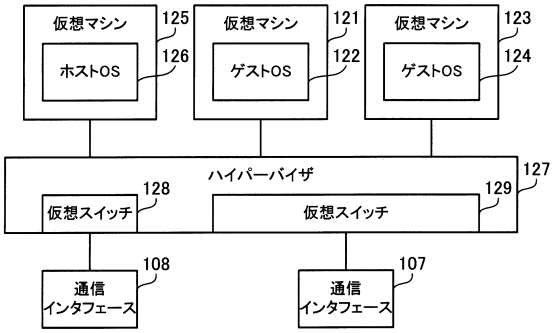
【図２】


10
データセンタ
11
ユーザネットワーク
ホストサーバ
100
100a
管理サーバ
13
12
管理ネットワーク
30
クライアント
31
インターネット
20
データセンタ
21
ユーザネットワーク
ホストサーバ
200
200a
管理サーバ
23
22
管理ネットワーク

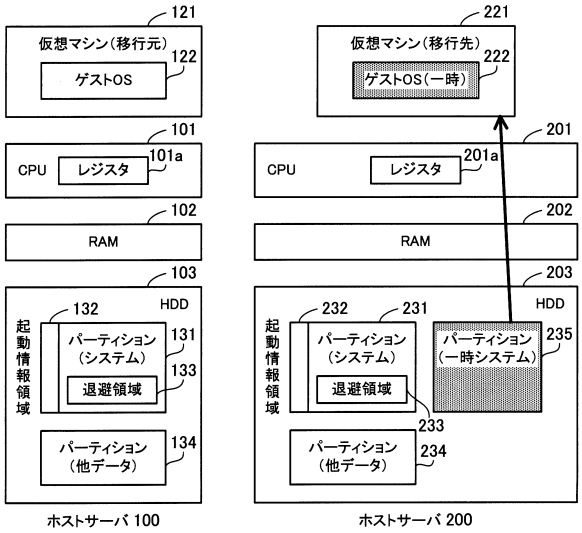
【図３】


ホストサーバ
100
CPU
101
画像信号処理部
104
111
ディスプレイ
RAM
102
入力信号処理部
105
112
入力デバイス
HDD
103
媒体リーダ
106
113
記録媒体
通信インタフェース
107
11
ユーザネットワーク
バス
109
通信インタフェース
108
12
管理ネットワーク

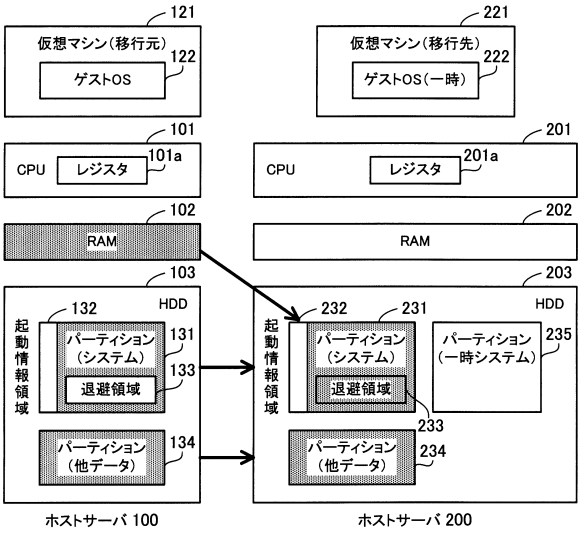
【図４】


仮想マシン
125
ホストOS
126
仮想マシン
121
ゲストOS
122
仮想マシン
123
ゲストOS
124
ハイパーバイザ
127
128
仮想スイッチ
129
仮想スイッチ
108
通信インタフェース
107
通信インタフェース

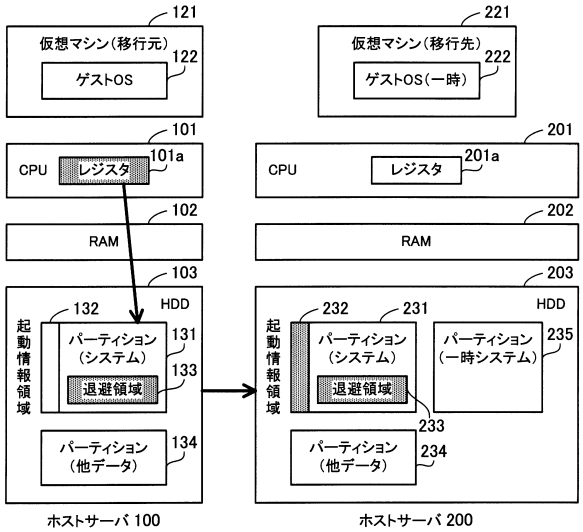
【図５】


121
仮想マシン(移行元)
122
ゲストOS
101
CPU
101a
レジスタ
102
RAM
103
HDD
起動情報領域
132
パーティション(システム)
131
退避領域
133
パーティション(他データ)
134
ホストサーバ 100
221
仮想マシン(移行先)
222
ゲストOS(一時)
201
CPU
201a
レジスタ
202
RAM
203
HDD
起動情報領域
232
231
パーティション(システム)
退避領域
233
パーティション(一時システム)
235
パーティション(他データ)
234
ホストサーバ 200

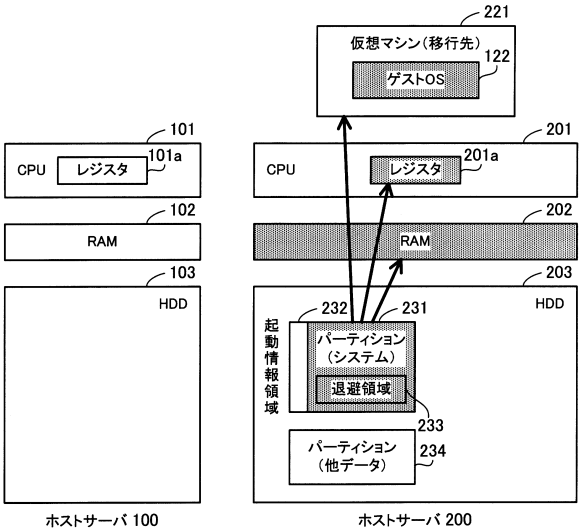
【図６】


121
仮想マシン(移行元)
122
ゲストOS
101
CPU
101a
レジスタ
102
RAM
103
HDD
起動情報領域
132
パーティション(システム)
131
退避領域
133
パーティション(他データ)
134
ホストサーバ 100
221
仮想マシン(移行先)
222
ゲストOS(一時)
201
CPU
201a
レジスタ
202
RAM
203
HDD
起動情報領域
232
231
パーティション(システム)
退避領域
233
パーティション(一時システム)
235
パーティション(他データ)
234
ホストサーバ 200

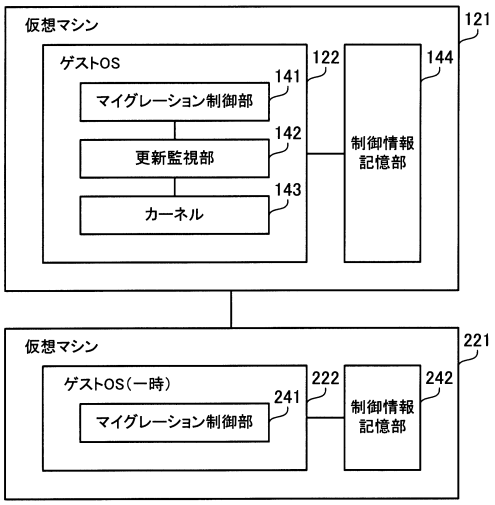
【図７】


121
仮想マシン(移行元)
122
ゲストOS
101
CPU
101a
レジスタ
102
RAM
103
HDD
起動情報領域
132
パーティション(システム)
131
退避領域
133
パーティション(他データ)
134
ホストサーバ 100
221
仮想マシン(移行先)
222
ゲストOS(一時)
201
CPU
201a
レジスタ
202
RAM
203
HDD
起動情報領域
232
231
パーティション(システム)
退避領域
233
パーティション(一時システム)
235
パーティション(他データ)
234
ホストサーバ 200

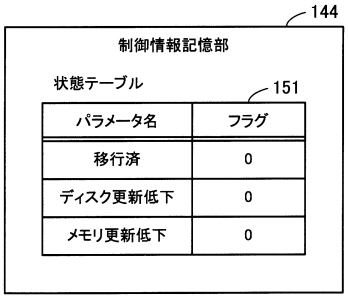
【図８】


101
CPU
101a
レジスタ
102
RAM
103
HDD
ホストサーバ 100
221
仮想マシン(移行先)
122
ゲストOS
201
CPU
201a
レジスタ
202
RAM
203
HDD
起動情報領域
232
231
パーティション(システム)
退避領域
233
パーティション(他データ)
234
ホストサーバ 200

【図９】

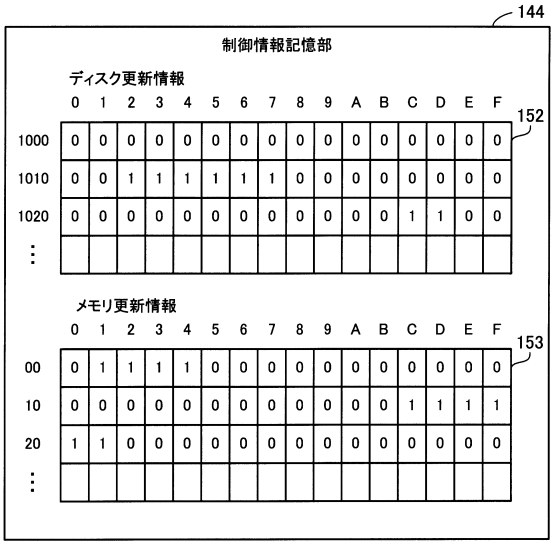


【図１０】

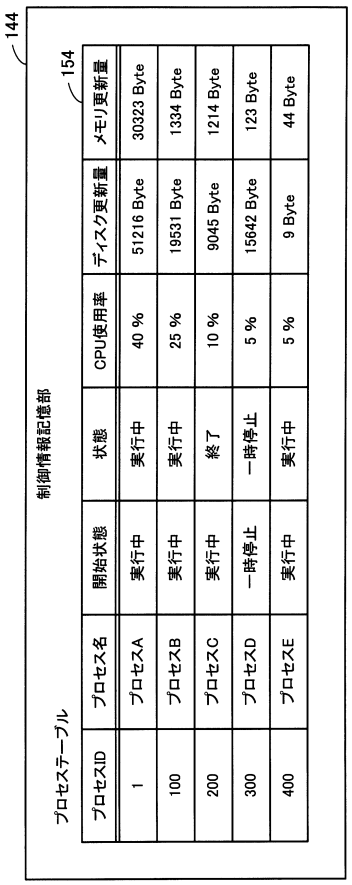


パラメータ名	フラグ
移行済	0
ディスク更新低下	0
メモリ更新低下	0

【図１１】

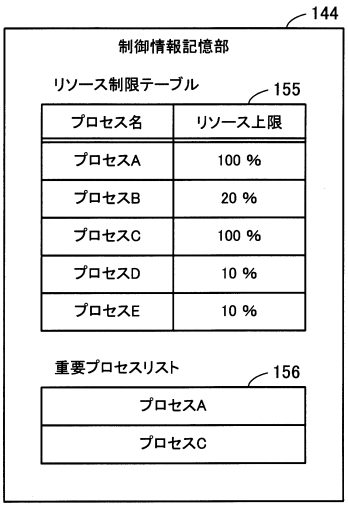


【図１２】



プロセスID	プロセス名	開始状態	状態	CPU使用率	ディスク更新量	メモリ更新量
1	プロセスA	実行中	実行中	40 %	51216 Byte	30323 Byte
100	プロセスB	実行中	実行中	25 %	19531 Byte	1334 Byte
200	プロセスC	実行中	終了	10 %	9045 Byte	1214 Byte
300	プロセスD	一時停止	一時停止	5 %	15642 Byte	123 Byte
400	プロセスE	実行中	実行中	5 %	9 Byte	44 Byte

【図１３】



プロセス名	リソース上限
プロセスA	100 %
プロセスB	20 %
プロセスC	100 %
プロセスD	10 %
プロセスE	10 %

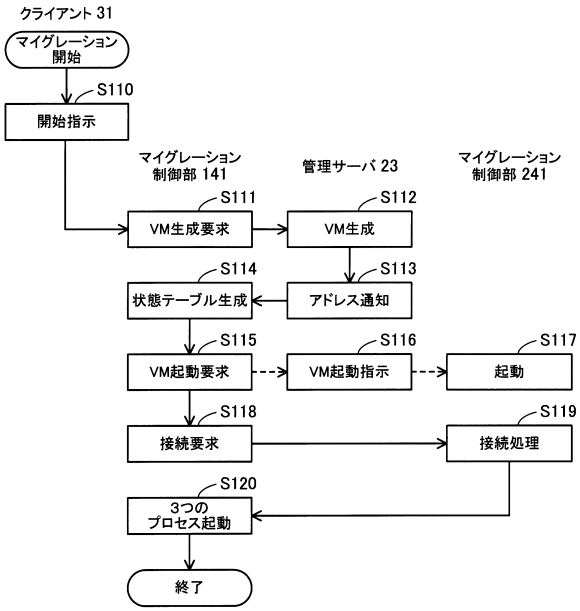
重要プロセスリスト
プロセスA
プロセスC

【図１４】

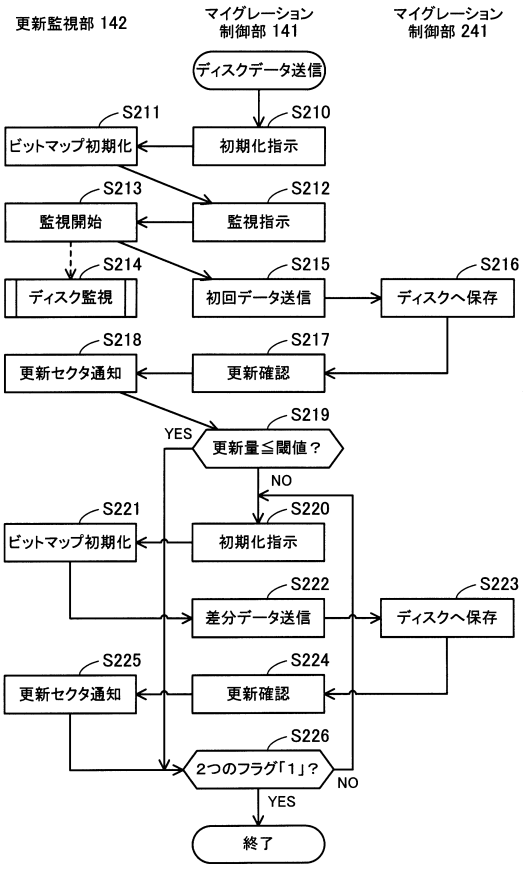
仮想マシン生成要求 157

パラメータ名	設定値
CPU数	2
メモリ容量	2 GByte
システムディスク容量	16 GByte
他データディスク容量	16 GByte
ネットワーク設定	インターネット
初期OS	マイグレーション用

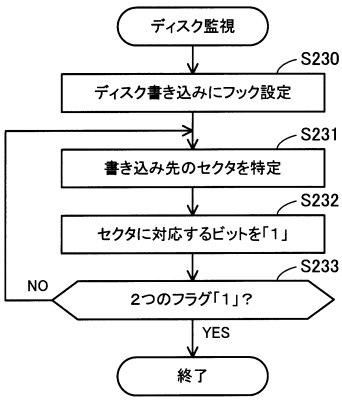
【図１５】



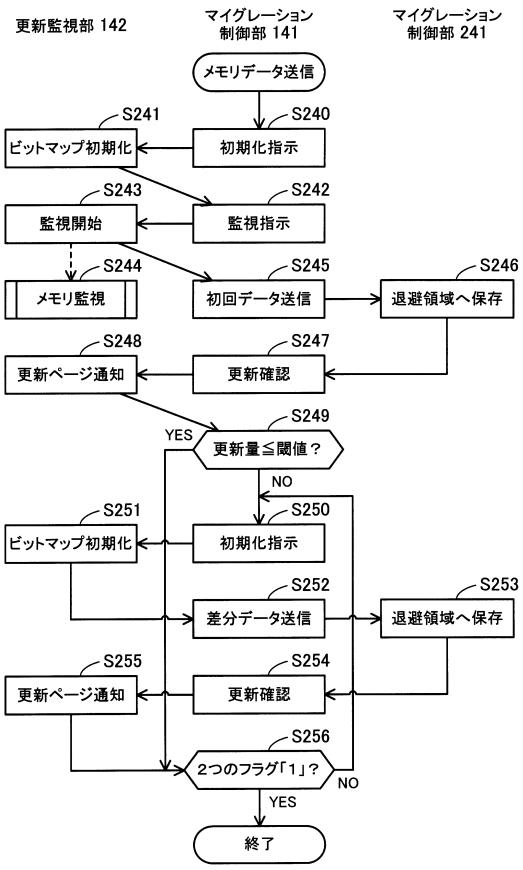
【図１６】



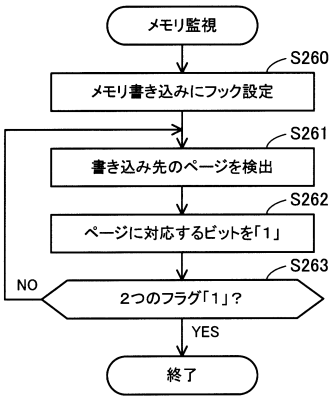
【図１７】



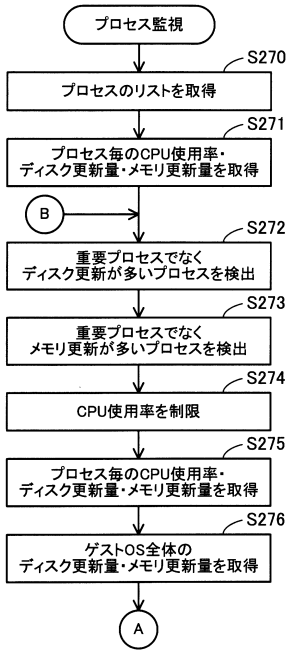
【図１８】



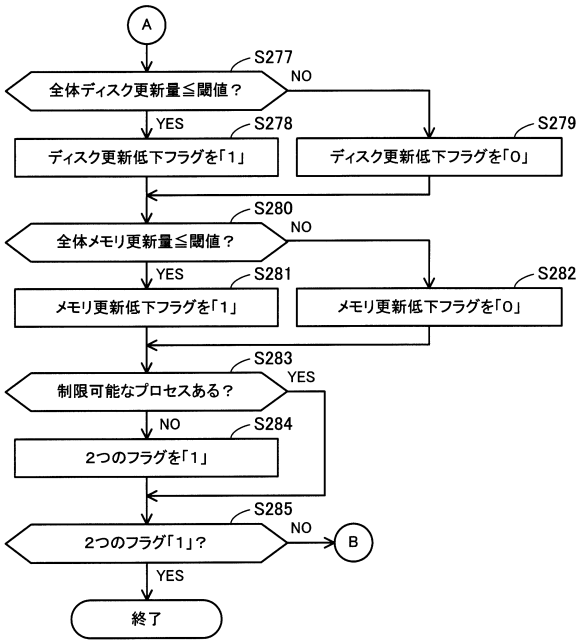
【図１９】



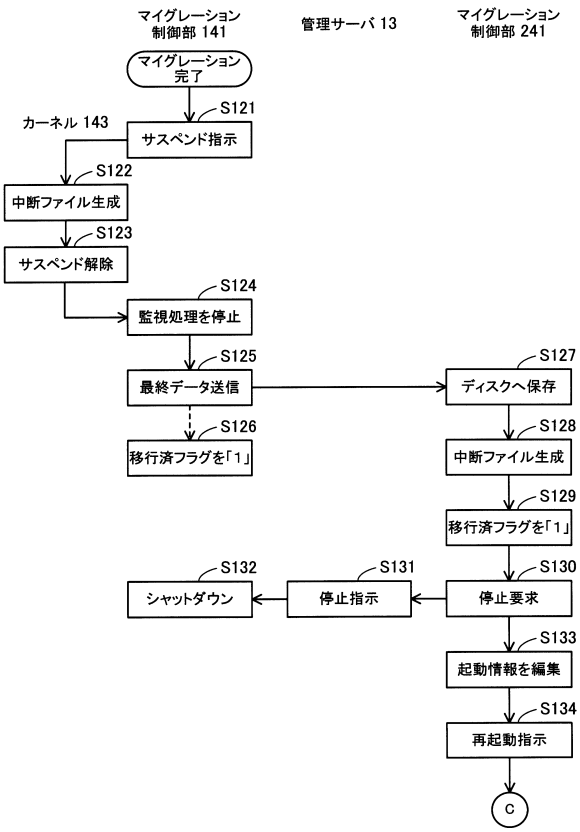
【図２０】



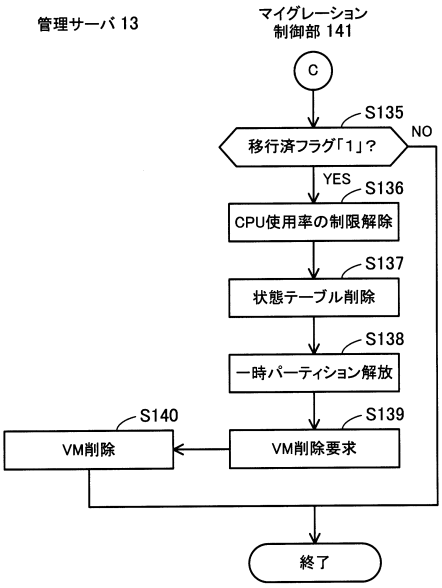
【図２１】



【図２２】



【図２３】


管理サーバ 13
マイグレーション
制御部 141
C
S135
移行済フラグ「1」？
NO
YES
S136
CPU使用率の制限解除
S137
状態テーブル削除
S138
一時パーティション解放
S140
VM削除
S139
VM削除要求
終了

フロントページの続き

(72)発明者　寺嶋　直矢
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　今枝　一英
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

審査官　原　忠

(56)参考文献　特開２０１２－２２１３２１（ＪＰ，Ａ）
国際公開第２００９／０３４７６０（WO，Ａ１）
特開２０１３－２５０９５０（ＪＰ，Ａ）
特開２０１４－１０２７２４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　９／４４
Ｇ０６Ｆ　　９／４５５
Ｇ０６Ｆ　　９／４６　－　９／５４