

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4259036号
(P4259036)

(45) 発行日 平成21年4月30日(2009.4.30)

(24) 登録日 平成21年2月20日(2009.2.20)

(51) Int.Cl.

F 1

G06F 12/08	(2006.01)	G06F 12/08	557
G06F 13/10	(2006.01)	G06F 12/08	551Z
G06F 13/00	(2006.01)	G06F 13/10	340A
G06F 12/00	(2006.01)	G06F 13/00	540B
G06F 3/06	(2006.01)	G06F 12/00	514E

請求項の数 5 (全 15 頁) 最終頁に続く

(21) 出願番号

特願2002-137919 (P2002-137919)

(22) 出願日

平成14年5月14日 (2002.5.14)

(65) 公開番号

特開2003-330794 (P2003-330794A)

(43) 公開日

平成15年11月21日 (2003.11.21)

審査請求日

平成17年4月1日 (2005.4.1)

(73) 特許権者 000005108

株式会社日立製作所

東京都千代田区丸の内一丁目6番6号

(74) 代理人 100100310

弁理士 井上 学

(72) 発明者 小原 清弘

東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72) 発明者 渡邊 龍

東京都江東区新砂一丁目6番27号 株式
会社日立製作所公共システム事業部内

(72) 発明者 末野 良信

東京都江東区新砂一丁目6番27号 株式
会社日立製作所公共システム事業部内

最終頁に続く

(54) 【発明の名称】 ネットワークデータのキャッシング方法およびシステム

(57) 【特許請求の範囲】

【請求項 1】

ネットワークI/Fと、二次記憶装置と、該二次記憶装置への接続I/Fと、OSと前記二次記憶装置のデバイスドライバとが格納される記憶手段と、前記デバイスドライバおよびOSを動作させるプロセッサとを有するサーバシステムであって、

前記OSは、前記プロセッサに、所定のデータの読み出し要求を行わせるためのものであり、

前記デバイスドライバは、前記プロセッサに、前記読み出し要求のあった前記所定のデータが、前記二次記憶装置に存在するか否かを判断させるためのものであり、

前記デバイスドライバは、前記プロセッサに、前記所定のデータが前記二次記憶装置に存在しない場合に、前記ネットワークI/Fに対して前記所定のデータの読み出しを要求させるためのものであり、

前記デバイスドライバは、前記プロセッサに、前記所定のデータが前記二次記憶装置に存在する場合に、前記二次記憶装置への接続I/Fに対して前記所定のデータの読み出しを要求させるためのものである、

ことを特徴とするサーバシステム。

【請求項 2】

請求項1に記載のサーバシステムにおいて、前記OSに基づき前記プロセッサが行う前記二次記憶装置に対する書き込み要求のあったデータの前記二次記憶装置における格納位置と、該データの格納元における格納位置とを対照する管理テーブルを備えることを特徴と

するサーバシステム。

【請求項 3】

クライアント端末が接続されたローカル側サーバと、ネットワークを介して前記ローカル側サーバと接続されるリモート側サーバを備えたクライアント・サーバシステムにおいて、

前記ローカル側サーバは、外部ネットワークへ接続される第1のネットワークI/Fと、第1の二次記憶装置と、該第1の二次記憶装置への接続I/Fと、OSと前記第1の二次記憶装置に対応するデバイスドライバが格納される記憶手段と、前記OSとデバイスドライバとを動作させる第1のプロセッサとを有し、

前記リモート側サーバは、第2の二次記憶装置と、該第2の二次記憶装置を制御する制御部とを有し、前記制御部は、外部ネットワークに接続される第2のネットワークI/Fと、

該第2の二次記憶装置への接続I/Fと、制御部を動作させる第2のプロセッサとを備え、

前記OSは、前記第1のプロセッサに、所定のデータの読み出し要求を行わせるためのものであり、

前記デバイスドライバは、前記第1のプロセッサに、前記読み出し要求のあった前記所定のデータが、前記第1の二次記憶装置に存在するか否かを判断させるためのものである、

前記デバイスドライバは、前記第1のプロセッサに、前記所定のデータが前記第1の二次記憶装置に存在しない場合に、前記リモート側サーバの制御部に対して前記所定のデータの読み出しを要求させるためのものであり、

前記デバイスドライバは、前記第1のプロセッサに、前記所定のデータが前記第1の二次記憶装置に存在する場合に、前記第1の二次記憶装置への接続I/Fに対して前記所定のデータの読み出しを要求させるためのものである、ことを特徴とするクライアント・サーバシステム。

【請求項 4】

請求項3に記載のクライアント・サーバシステムにおいて、

さらに、制御計算機を有し、

前記ローカル側サーバは前記OSに基づき前記第1のプロセッサが行う第1の二次記憶装置に対する書き込み要求のあったデータの、前記第1の二次記憶装置における格納位置と前記第2の二次記憶装置における格納位置とを対照する管理テーブルを備え、

前記管理テーブルは、前記OSに基づき前記第1のプロセッサが行うデータの書き込み命令を受けたことを示す情報が記録された項目を備え、

前記制御計算機は、前記管理テーブルにおける前記情報が記録された項目に対応するデータの読み出しを、前記第一の二次記憶装置に対して所定のタイミングで要求することを特徴とするクライアント・サーバシステム。

【請求項 5】

ネットワークI/Fと、二次記憶装置と、該二次記憶装置への接続I/Fと、前記二次記憶装置のデバイスドライバとOSが格納される記憶手段と、OSと前記デバイスドライバを動作させるプロセッサとを備えたサーバの制御方法において、

前記OSに基づき、前記プロセッサが、所定のデータの読み出し要求を行うステップと、

前記デバイスドライバに基づき、前記プロセッサが、前記読み出し要求のあった前記所定のデータが、前記二次記憶装置に存在するか否かを判断するステップと、

前記デバイスドライバに基づき、前記プロセッサが、前記所定のデータが前記二次記憶装置に存在しない場合に、前記ネットワークI/Fに対して前記所定のデータの読み出しを要求するステップと、

前記デバイスドライバに基づき、前記プロセッサが、前記所定のデータが前記二次記憶装置に存在する場合に、前記二次記憶装置への接続I/Fに対して前記所定のデータの読み出しを要求するステップと

を含むことを特徴とするサーバの制御方法。

10

20

30

40

50

【発明の詳細な説明】**【0001】****【発明の属する技術分野】**

本発明はネットワークを経由した遠隔データアクセスに関する。特にデータのキャッシングおよびその制御を実現するシステムに関する。

【0002】**【従来の技術】**

近年、ASP (Application Service Provider) の形式で、サービスを提供する企業が現れている。ASPの利点は、データベース等のデータは全てデータセンタ側にあるため、顧客ユーザのシステム規模を小さくでき、顧客にとって最新のアプリケーションを低成本で利用できる点がある。サービス提供企業側も、複数の顧客のシステムを統合でき、設備投資も少なくて済む。さらにデータバックアップ作業等もデータセンタ側で行うため、顧客側でのデータの管理コストも低減できる等の利点もある。これらの利点から、最近では様々なサービスがASP形式で提供されており、医療情報分野においても同様である。

10

【0003】

また同様なサービスをディスク上のデータに対して行うことを可能にする技術として、iSCSIと呼ばれる技術が近年注目されている。iSCSIは、TCP/IPネットワークを経由して、ディスク入出力コマンドであるSCSIコマンドを受け渡す技術である。この技術の規格は、インターネット上の規格の制定団体であるIETF (The Internet Engineering Task Force) のワーキンググループによって議論されており、現在ドラフトレベルの規格が公表されている。またこのプロトコルに対応したストレージ製品として、米国IBM社の、IBM Total Storage IP Storage 200iがある。

20

【0004】

この技術を利用すると、データセンタに置かれ。顧客とLANや通信回線で接続されたディスクを、顧客配下のローカルディスクと同様に見せることができる。このため、ASPでの利点と同様に、クライント側に大量の記憶装置を置く必要が無くなり、データのバックアップ等の管理もデータセンタ側に委託でき、管理コストが低減できる。また、全て顧客の下で動作していたシステムに対し、既存のアプリケーションプログラムの変更無しに、ディスク等の記憶装置をデータセンタに移行し、管理コスト低減の利益を享受する事も可能になる。このようなストレージ管理サービスを提供する形態は、SSP (Storage Service Provider) と呼ばれている。

30

【0005】

一方ASPやSSPの問題点としては、操作毎にクライアントとサーバ間で通信が発生し、通信遅延のため動作が遅くなる点が挙げられる。特に医療情報分野では、CTやMRI等の画像の大容量データを必要とする場合が多く、通信による遅延が顕著になる。一般に通信遅延を隠蔽するには、通信データのキャッシングが有効である。しかし、キャッシングによる通信遅延の隠蔽は、キャッシングされたデータが再利用される場合に限定され、一回だけしか参照されないデータに対しては有効でない。

40

【0006】

一回だけしか参照されないデータに対しても通信遅延を隠蔽する方法としては、先読みキャッシュの方法が知られている。例えば特許第3074535号では、医療情報システムに対し、スケジューリングによる先読みキャッシュにより、データアクセスの遅延を隠蔽する方法が開示されている。

【0007】

しかし先読みキャッシュによる遅延の隠蔽方式を、既存のアプリケーションに対して適用するのは難しい。これは先読みを行うプログラムと既存のアプリケーションは別のプログラムであり、キャッシングされたデータが既存のアプリケーションとは別のアプリケーションプログラム中に保存されることに起因する。このため、既存のアプリケーションではデ

50

ータアクセスに際し、アクセスしたいデータが先読みされている否かを判断し、先読みされている場合は先読みプログラムからデータを受け取り、先読みされていない場合はOSに要求するように変更する必要がある。

【0008】

【発明が解決しようとする課題】

先読みキャッシュによる遅延の隠蔽では、キャッシュされたデータが既存のアプリケーションとは別のアプリケーションプログラム中に保存される。このためこの方式を、既存のアプリケーションに対して適用するのは難しい。

【0009】

本発明の目的は、ネットワークを経由してデータの送受信を行うiSCSIのような方式に対し、ネットワークによるデータアクセス遅延を隠蔽するため、キャッシュ動作を行う手段を提供する事にある。また、本発明の別の目的は、キャッシュされているデータとされていないデータの区別をアプリケーションプログラムから隠蔽する手段を提供することにある。これにより、すでにアプリケーションプログラムが動作している環境に対し、先読みプログラムによるキャッシングを導入した場合において、既存のプログラムは従来通りOSに対してデータの要求を行う方式のままで、キャッシュされたデータを得ることが可能になる。

10

【0010】

本発明のもう一つの目的は、前記キャッシュ動作の制御を、キャッシュを行っているシステム以外のシステムから制御する手段を提供することにある。

20

【0011】

【課題を解決するための手段】

上記データアクセス遅延の隠蔽は、クライアントシステム内のデバイスドライバレベルでデータをキャッシングする手段、キャッシュしたデータを外部記憶装置に格納する手段により達成される。デバイスドライバとは、ハードウェアの違いを吸収し、OSにデバイスアクセスのための共有の手順を提供するためのソフトウェアである。ここでのデバイスとは、ディスクインターフェースやネットワークインターフェースを意味する。

【0012】

また上記のキャッシュの制御は、クライアントの外から、キャッシュフラッシュ等のキャッシュ制御の命令をクライアントに送信する手段、クライアント内において受信したキャッシュ制御命令を実行する手段により達成される。

30

【0013】

【発明の実施の形態】

次に図1～図5を用いて本発明の第一の実施例を説明する。最初に、通信プロトコルにiSCSIを用いた場合の一般的な手順を示し、次に本発明であるクライアントシステムのデバイスドライバでのキャッシングについて述べる。

【0014】

図2に、クライアントシステム1とiSCSI対応ディスクシステム11の一構成例を示す。クライアントシステム1のハードウェア構成は、WindowsやUNIX等のOSが動作する、汎用的なコンピュータシステムと同様である。すなわち、プロセッサ2、主記憶3、ネットワークインターフェース4、ディスクインターフェース5および6を持ち、内部バス7で相互接続された構成である。またディスクインターフェースは5および6は、それぞれディスク装置8および9と接続されている。

40

【0015】

ディスク装置8には、OS、アプリケーションプログラム、デバイスドライバ等のプログラムが記録されている。これらのプログラムはクライアントシステム1の電源投入後、必要に応じてプロセッサ2およびディスクインターフェース5により、主記憶3上に読み込まれ実行される。

【0016】

ネットワークインターフェース4は、LANやWAN10を介してiSCSI対応ディスク

50

システム11と接続されている。LANやWANの中でルータやハブ等の中継装置を用いた場合も本特許は適用可能であるが、実施例の動作には影響しないので省略する。

【0017】

i SCSI対応ディスクシステム11は、i SCSI対応ディスク制御装置12を持ち、ディスク装置13に接続されている。複数のディスク装置を持つ場合も本特許は適用可能であるが、実施例の動作には影響しないので、ディスク制御装置が一つの場合について説明する。

【0018】

図3に、クライアントシステム1のソフトウェア階層とハードウェア階層を示す。ここでのハードウェア階層の各要素は、図2の各要素に対応している。また、上下で接している階層間は互いにデータ交換可能である事を示す。10

【0019】

まず、一般的なSCSIディスク入出力動作について述べる。アプリケーション層14からディスクにデータを書き込む場合、ディスク入力要求は最初にOS層15に伝えられる。OSは、ディスク入力を実行するのに必要な情報を、ハードウェア機器に依存しない形式で、SCSI用デバイスドライバ16に渡す。SCSI用デバイスドライバ16は、OSから受け取った情報を、ディスクインターフェース5独自の入出力仕様に変換する。そして、SCSI用デバイスドライバ16が、ディスクインターフェース5に情報を書き込むことにより、ディスクインターフェース5が起動される。続いてデータを伴ったSCSIコマンドの書き込み要求がディスク装置8に送られ、データの書き込みが実行される。20

【0020】

アプリケーション層からのデータの読み出しも、ディスク出力要求が前記ディスク入力要求の場合と同様の手順で、ディスク装置8に送られる。そしてディスク装置8からの出力データは、前記とは逆の手順でアプリケーション層に送られ一連のデータ読み出し操作が完了する。

【0021】

続いて一般的なi SCSI対応ディスク制御装置12に対する入出力動作について述べる。アプリケーション層14からディスクにデータを書き込む場合、ディスク入力要求は最初にOS層15に伝えられる。OSは、ディスク入力を実行するのに必要な情報を、ハードウェア機器に依存しない形式で、i SCSI用デバイスドライバ17に渡す。30

【0022】

i SCSI用デバイスドライバの、OS側とのインターフェースのパラメタ等の構成は、通常のSCSI用デバイスドライバと同じになっている。このため、OSがi SCSIデバイスドライバに渡す情報は、SCSIデバイスドライバの場合と同じである。また同様の理由から、OSから見て、i SCSIプロトコルでアクセスするディスク装置とSCSIプロトコルでアクセスするディスク装置は、同じ種類のデバイスに見える。

【0023】

i SCSI用デバイスドライバ17は、OSから受け取った情報を、i SCSI仕様のコマンド形式に変換する。そして、ネットワークインターフェース4およびLAN/WAN10を使用し、i SCSI対応ディスクシステム11にi SCSIコマンドを送ることにより、データの書き込みが実行される。40

【0024】

LAN/WANを経由して交換されるi SCSIコマンドの詳細は、インターネット上の規格の制定団体であるIETF(The Internet Engineering Task Force)のワーキンググループによって議論されており、現在ドラフトレベルの規格が公表されている。

【0025】

アプリケーション層からのデータの読み出しも、前記ディスク入力要求の場合と同様の手順で、i SCSI対応ディスクシステム11に送られる。そして、i SCSI対応ディスクシステム11からの出力データは、前記とは逆の手順でアプリケーション層に送られ50

連のデータ読み出し操作が完了する。

【0026】

次に本特許であるデバイスドライバにおけるキャッシュ動作について説明する。

【0027】

キャッシュデータは、キャッシュ用ディスク装置9に格納する。これは、OS、アプリケーションプログラム、デバイスドライバ等のプログラムが記録されているディスク装置8とは別のディスク装置である(図2)。読み出したデータがキャッシュ上にあるか否かの判定は、クライントシステム1の主記憶3上に置いた、キャッシュ管理テーブル18を用いる。図4にキャッシュ管理テーブルの構成を示す。

【0028】

一般にディスク装置内の記録領域は、ブロックと呼ばれる一定サイズに区切られ、各々にブロック番号と呼ばれる識別番号が割り当てられている。このため入出力は、データの位置を示すブロック番号と、データの長さを示すブロック数を指定して行われる。

10

【0029】

キャッシュ管理テーブル18は、上記の入出力方法に対応し、キャッシュ用ディスクの識別番号19、キャッシュ用ディスクのブロック番号20、キャッシュ用ディスクのブロック数21、i SCSIディスクの識別番号22、i SCSIディスクのブロック番号23、i SCSIディスクのブロック数24、通し番号25、の組を保持する。

【0030】

キャッシュ管理テーブル18の生成は、クライントシステム1の初期化事に行われる。初期化された直後のテーブルは空である。これは何もキャッシュされていない事を意味する。データがキャッシュされる毎に一個のエントリが生成される。

20

【0031】

図4においては、i SCSIディスク中の、ブロック番号0から10ブロック分のデータが、キャッシュ用ディスク中の、ブロック番号0から10ブロック分の領域に格納されている状態を示している。また、i SCSIディスク中の、ブロック番号88から8ブロック分のデータは、キャッシュ用ディスク中の、ブロック番号10から8ブロック分の領域に格納されている。キャッシュ用ディスクの識別番号とi SCSIディスクの識別番号は、各々に複数のディスク装置を使用している場合、ディスク装置を識別する為に用いる。本実施例では各々一個のディスクを使用しているため、識別番号は0が入っている。

30

【0032】

図1(a)にi SCSI対応ディスクシステムに対し、データの読み出し(リード)を行った場合の、i SCSI用デバイスドライバの動作のフローチャートを示す。OSからi SCSI用デバイスドライバへの要求が、データの読み出し(リード)動作の場合、i SCSI用デバイスドライバは、対象のデータがキャッシュ用ディスク9上にあるか否かを、上記のキャッシュ管理テーブル18を検索して判断する(101)。対象のデータがキャッシュ用ディスク上9にある場合、ディスクインターフェース6を使用して、データをキャッシュ用ディスク装置9から読み出し、OSに返す(102)。

【0033】

対象のデータがキャッシュ用ディスク9上にない場合は、従来のi SCSI用デバイスドライバと同様に、OSから受け取った情報を、i SCSI仕様のコマンド形式に変換し、ネットワークインターフェース4およびLAN/WAN10を使用し、i SCSI対応ディスクシステム11にi SCSIコマンドを送る(103)。

40

【0034】

i SCSI対応ディスクシステム11からデータが到着すると、ネットワークインターフェース4からパケット受信の割り込みが入る(105)(図1(b))。これにより、再びi SCSI用デバイスドライバが起動され、パケット中のリードデータ部分をキャッシュディスクに格納すると共に、その情報をキャッシュ管理テーブル中の新しいエントリとして追加する(106)。そして、到着したパケット中のi SCSIコマンド部分を通常のSCSIコマンドの情報に変換しOSに返す(107)。以上の一連の動作により、対象

50

のデータがキャッシュ用ディスク9上にない場合のリード動作が完了する。

【0035】

図5にiSCSI対応ディスクシステムに対し、データの書き込み(ライト)を行った場合の、iSCSI用デバイスドライバの動作のフローチャートを示す。OSからiSCSI用デバイスドライバへの要求が、データの書き込み(ライト)動作の場合、iSCSI用デバイスドライバは、対象のデータがキャッシュ用ディスク9上にあるか否かを、キャッシュ管理テーブル18を検索して判断する(109)。対象のデータがキャッシュ用ディスク9上にある場合、ディスクインターフェース6を使用して、データをキャッシュ用ディスク装置9に書き込みデータを更新する(110)。対象のデータがキャッシュ用ディスク9上に無い場合、データをキャッシュディスクに格納すると共に、その情報をキャッシュ管理テーブル中の新しいエントリとして追加する(111)。
10

【0036】

そしていずれの場合も、OSから受け取った情報を、iSCSI仕様のコマンド形式に変換する。そして、ネットワークインターフェース4およびLAN/WAN10を使用し、iSCSI対応ディスクシステム11にiSCSIコマンドを送ることにより、データの書き込みが実行される(112)。

【0037】

以上により、ネットワークを経由してデータの送受信を行う方式に対し、ネットワークによるデータアクセス遅延を隠蔽できる。またキャッシュされているデータとされていないデータの区別をアプリケーションプログラムから隠蔽でき、先読みプログラム等によるキャッシングを導入した場合、既存のプログラムは従来通りOSに対してデータの要求を行う方式のままで、キャッシュされたデータを得ることが可能になる。
20

【0038】

上記の実施例では、アプリケーション層から要求のあった書き込みデータは、直ちにiSCSI対応ディスクシステム11に書き込まれる動作を説明しているが、LAN/WAN回線が塞がっていたり、指定された時間だけ接続される場合などに対応するため、任意の時間後に書き込む動作も可能である。

【0039】

この動作を実現する場合、キャッシュ管理テーブル18に、未送信フラグ領域を付け加える。直ちにiSCSI対応ディスクシステム11に書き込まない場合、iSCSI用デバイスドライバは、キャッシュ管理テーブル18のエントリを作成又は更新する時に、未送信フラグをセットする。そして、タイム等によるLAN/WAN回線の監視の結果、送信可能になった場合、未送信フラグがセットされているデータをまとめて送信することにより、任意の時間後に書き込む動作を実現する。
30

【0040】

前記の実施例では、キャッシュ用ディスク9が一杯になり新たなデータが入れられない場合、最も古いキャッシュデータを削除し、新たなデータをキャッシュする。これはキャッシュの制御で一般的に用いられている、いわゆる先入れ先出しの方式である。キャッシュ管理テーブル18中の通し番号25は、エントリが生成された順番を表し、このエントリの値を用いて、最も古いキャッシュデータを確定することが可能である。すなわち、キャッシュ用ディスク9が一杯になった場合、通し番号25が一番若いエントリを探し、そのエントリ示すキャッシュ用ディスク9上の領域に新たなデータを格納し、エントリを更新する。
40

【0041】

この先入れ先出し方式に加え、iSCSI対応ディスクシステム11中の指定した領域に対するデータはキャッシュ用ディスク9上から削除せず、常にキャッシュ用ディスク9に置いておく方式も可能である。この動作を実現する場合、キャッシュ管理テーブル18に、キャッシュ常駐フラグ領域を付け加える。このキャッシュ常駐フラグがセットされている場合、前記のキャッシュデータ削除の対象から外すことにより、データのキャッシュへの常駐化を実現できる。
50

【0042】

本発明の第二の実施例を説明する。

【0043】

本実施例では、iSCSI対応ディスクシステムとの接続インターフェースに、キャッシング用ディスクへのインターフェースが含まれている場合を説明する。

【0044】

図6に、クライアントシステム1とiSCSI対応ディスクシステム11の第二の構成例を示す。

【0045】

本実施例でも第一の実施例と同様に、クライアントシステム1のハードウェア構成は、WindowsやUNIX等のOSが動作する、汎用的なコンピュータシステムと同様である。すなわち、プロセッサ2、主記憶3、ディスクインターフェース5を持ち、内部バス7で相互接続された構成である。またディスクインターフェース5はディスク装置8と接続されている。

【0046】

ディスク装置8には、OS、アプリケーションプログラム、デバイスドライバ等のプログラムが記録されている。これらのプログラムはクライアントシステム1の電源投入後、必要に応じてプロセッサ2およびディスクインターフェース5により、主記憶3上に読み込まれ実行される。

【0047】

iSCSIディスクインターフェース31は、プロセッサ32、メモリ33、ネットワークインターフェース34、ディスクインターフェース35、クライアントインターフェース36を持ち、内部バス37で相互接続された構成である。またディスクインターフェース35はディスク装置29と接続されている。iSCSIディスクインターフェース31の動作は、プロセッサ32がメモリ33中にある制御プログラムを実行することにより制御される。

【0048】

クライアントインターフェース36は、クライアントシステム1との入出力を制御する回路であり、制御のための各種レジスタや割り込み信号線等は、従来のディスクインターフェース5等と同様な構成を持つ。このため、クライアントシステム1からは、iSCSIディスクインターフェース31は、従来のディスクインターフェース5等と同様に見える。したがって、クライアントシステム1中のデバイスドライバも、従来のディスクインターフェース用のデバイスドライバが使用される。すなわち、iSCSIディスクインターフェース31は、クライアントシステム1中のデバイスドライバと情報の交換を行う。

【0049】

すなわち、本実施例では従来のデバイスドライバを使用するため、デバイスドライバ中でキャッシングの動作は行わない。キャッシングの動作は、iSCSIディスクインターフェース31内で行なう。このため、第一の実施例で示したキャッシング管理テーブル18は、iSCSIディスクインターフェース31中のメモリ33中に格納される。

【0050】

図7(a)にiSCSI対応ディスクシステムに対し、データの読み出し(リード)を行った場合の、iSCSIディスクインターフェース31の動作のフローチャートを示す。

【0051】

デバイスドライバからiSCSIディスクインターフェース31への要求が、データの読み出し(リード)動作の場合、iSCSIディスクインターフェース31は、対象のデータがキャッシング用ディスク29上にあるか否かを、キャッシング管理テーブルを検索して判断する(201)。対象のデータがキャッシング用ディスク上29にある場合、ディスクインターフェース35を使用して、データをキャッシング用ディスク装置29から読み出し、デバイスドライバに返す(202)。

【0052】

対象のデータがキャッシング用ディスク29上にない場合は、デバイスドライバから受け取

10

20

30

40

50

った情報を、iSCSI仕様のコマンド形式に変換し、ネットワークインターフェース34およびLAN/WAN10を使用し、iSCSI対応ディスクシステム11にiSCSIコマンドを送る(203)。

【0053】

iSCSI対応ディスクシステム11からデータが到着すると、ネットワークインターフェース34からパケット受信の割り込みが入る(205)(図7(b))。これにより再び処理が起動され、パケット中のリードデータ部分をキャッシュディスクに格納すると共に、その情報をキャッシュ管理テーブル中の新しいエントリとして追加する(206)。そして、到着したパケット中のiSCSIコマンド部分を通常のSCSIコマンドの情報に変換しデバイスドライバに返す(207)。以上の一連の動作により、対象のデータがキャッシュ用ディスク29上にない場合のリード動作が完了する。10

【0054】

図8にiSCSI対応ディスクシステムに対し、データの書き込み(ライト)を行った場合の、iSCSIディスクインターフェース31の動作のフローチャートを示す。

【0055】

デバイスドライバからiSCSIディスクインターフェース31への要求が、データの書き込み(ライト)動作の場合、iSCSIディスクインターフェース31は、対象のデータがキャッシュ用ディスク29上にあるか否かを、キャッシュ管理テーブルを検索して判断する(209)。対象のデータがキャッシュ用ディスク29上にある場合、ディスクインターフェース35を使用して、データをキャッシュ用ディスク装置29に書き込みデータを更新する(210)。対象のデータがキャッシュ用ディスク29上に無い場合、データをキャッシュディスクに格納すると共に、その情報をキャッシュ管理テーブル中の新しいエントリとして追加する(211)。20

【0056】

そしていずれの場合も、デバイスドライバから受け取った情報を、iSCSI仕様のコマンド形式に変換する。そして、ネットワークインターフェース34およびLAN/WAN10を使用し、iSCSI対応ディスクシステム11にiSCSIコマンドを送ることにより、データの書き込みが実行される(212)。

【0057】

以上により、ネットワークを経由してデータの送受信を行う方式に対し、ネットワークによるデータアクセス遅延を隠蔽できる。またキャッシュされているデータとされていないデータの区別をアプリケーションプログラムから隠蔽でき、先読みプログラム等によるキャッシングを導入した場合、既存のプログラムは従来通りOSに対してデータの要求を行う方式のままで、キャッシュされたデータを得ることが可能になる。30

【0058】

本実施例でも、クライアントシステムからのデータを、任意の時間後にiSCSI対応ディスクシステム11に書き込む動作や、任意のデータのキャッシュへの常駐化を、第一の実施例と同様の手順で実現可能である。

【0059】

本発明の第三の実施例を説明する。40

【0060】

本実施例では、キャッシュ動作をクライアントシステム1以外のシステムから制御する方法を説明する。

【0061】

第一および第二の実施例で、クライアントシステムからのデータを、任意の時間後にiSCSI対応ディスクシステム11に書き込む動作や、任意のデータのキャッシュへの常駐化の実施例を述べたが、この動作の詳細を、クライアントシステム1以外のシステムから制御可能である。図9に、クライアントシステム1以外のシステムからキャッシュ動作を制御する場合の構成例を示す。

【0062】

50

本構成例では、第一の実施例の構成に加え、制御計算機30がLAN/WANに接続されている。制御計算機は、LAN/WANを経由して、クライアントシステム1に対しキャッシュ制御コマンドを送信する。

【0063】

キャッシング制御コマンドは、クライアントシステム1とiSCSI対応ディスクシステム11間で交換されるiSCSIパケットとは区別可能な任意のフォーマットで作成する。これにより、iSCSI用デバイスドライバが、ネットワークインターフェースからパケット受信の割り込みで起動された時に、iSCSIパケットかキャッシング制御コマンドかの区別が可能になり、キャッシング制御コマンドの場合には、その内容に応じた動作が可能になる。

10

【0064】

キャッシング制御コマンドによるiSCSI用デバイスドライバの動作の例は次のものがある。

(1)キャッシングフラッシュ：これは、iSCSI対応ディスクシステム11にまだ送信していないデータを全て送信する動作である。この動作は、iSCSI用デバイスドライバが、キャッシング管理テーブル18中の未送信フラグがセットされているデータ全てを、iSCSI対応ディスクシステム11に送信することにより実現される。また、キャッシング管理テーブル18中のiSCSIディスクの識別番号情報を使用することにより、任意のiSCSIディスクに対する未送信データだけを送信させることもできる。

(2)送信停止：前述の、任意の時間後にiSCSI対応ディスクシステム11に書き込む動作と基本的に同じであり、通常ならiSCSI対応ディスクシステムに対し書き込みのため直ちに送信されるデータを、一時的に送信しないでおく動作である。

20

【0065】

この動作は、クライアントシステム1中の主記憶3上に、送信可否情報を保存し、OSからデータの書き込み要求があった場合、前記送信可否情報を参照し、送信不可の場合、送信されるデータをキャッシング用ディスク上に格納すると共に、そのデータに対するキャッシング管理テーブル18中の未送信フラグをセットすることにより実現する。また、送信可否情報に対するiSCSIディスクの識別番号も記録することにより、任意のiSCSIディスクに対してだけ、一時的に送信しないでおく動作もできる。

(3)送信再開：前述の送信停止を解除する動作である。この動作は、クライアントシステム1中の主記憶3上の送信可否情報を、送信可能を示す情報に更新することにより実現する。

30

【0066】

本実施例は、対象とする構成を第一の実施例に基づいて記述したが、第二の実施例の構成に対しても適用可能である。この場合、動作の対象となるのが、iSCSI用デバイスドライバの代わりに、iSCSIディスクインターフェース31中の制御プログラムとなる。また、送信可否情報が格納されるのは、iSCSIディスクインターフェース31中のメモリ33になる。

【0067】

これまで説明してきた全ての実施例では、LAN/WANネットワークを経由して交換されるディスク入出力コマンドをiSCSIプロトコルとして説明したが、デバイスドライバが解釈するプロトコルを変更すれば、iSCSI以外のプロトコルに対しても適用できる。

40

【0068】

さらに、これまで説明してきた全ての実施例では、キャッシング用ディスク9や29は、OS等が格納されるディスク8とは別のディスク装置としていたが、同じディスク装置の場合でも本特許は実施可能である。

【0069】

OS等が格納されるディスクとキャッシング用ディスクを共用する場合、OS等が格納されるディスク上にキャッシング用の領域を予め確保し、そのキャッシング用の領域情報をデバイ

50

ストライバに与えることで実現できる。キャッシング用の領域情報とは、キャッシング用に確保されたディスクの識別番号、キャッシング用に確保された領域の開始ブロック番号とキャッシング用領域のブロック数である。この情報をOSの起動時に主記憶上に置き、i SCSI用デバイスドライバがその情報を読み込むことにより、OS等が格納されるディスクとキャッシング用ディスクの共用が実現できる。

【0070】

【発明の効果】

本発明によれば、ネットワークを経由してデータの送受信を行う方式に対し、ネットワークによるデータアクセス遅延を隠蔽できる。またキャッシングされているデータとされていないデータの区別をアプリケーションプログラムから隠蔽でき、先読みプログラム等によるキャッシングを導入した場合、既存のプログラムは従来通りOSに対してデータの要求を行う方式のままで、キャッシングされたデータを得ることが可能になる効果がある。さらに、前記キャッシング動作の制御を、キャッシングを行っているシステム以外のシステムから制御できるという効果がある。

10

【図面の簡単な説明】

【図1】データの読み出しを行った場合の、i SCSI用デバイスドライバの動作のフローチャートである。

【図2】クライアントシステムとi SCSI対応ディスクシステムの一構成例である。

【図3】クライアントシステム1のソフトウェア階層とハードウェア階層の一構成例である。

20

【図4】キャッシング管理テーブルの一構成例である。

【図5】データの書き込みを行った場合の、i SCSI用デバイスドライバの動作のフローチャートである。

【図6】クライアントシステムとi SCSI対応ディスクシステムの別の構成例である。

【図7】データの読み出しを行った場合の、i SCSIディスクインターフェースの動作のフローチャートである。

【図8】データの書き込みを行った場合の、i SCSIディスクインターフェースの動作のフローチャートである。

【図9】クライアントシステム以外のシステムからキャッシング動作を制御する場合の一構成例である。

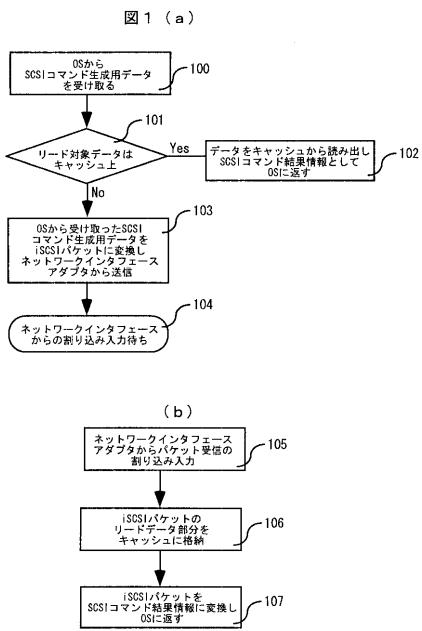
30

【符号の説明】

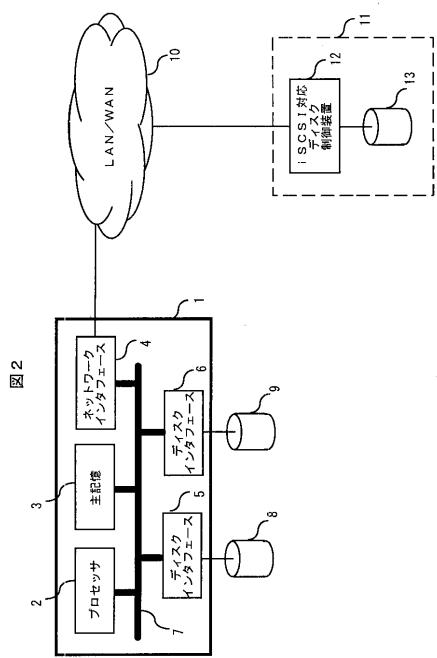
1 …… クライアントシステム、2 …… プロセッサ、3 …… 主記憶、4 …… ネットワークインターフェース、5・6 …… ディスクインターフェース、7 …… 内部バス、8 …… ディスク装置、9 …… キャッシュ用ディスク装置、10 …… LAN / AN、11 …… i SCSI対応ディスクシステム、12 …… i SCSI対応ディスク制御装置、13 …… ディスク装置、14 …… アプリケーション層、15 …… OS層、16 …… SCSI用デバイスドライバ、17 …… i SCSI用デバイスドライバ、18 …… キャッシュ管理テーブル、29 …… キャッシュ用ディスク装置、30 …… 制御計算機、31 …… i SCSIディスクインターフェース、32 …… プロセッサ、33 …… メモリ、34 …… ネットワークインターフェース、35 …… ディスクインターフェース、36 …… クライアントインターフェース、37 …… 内部バス。

40

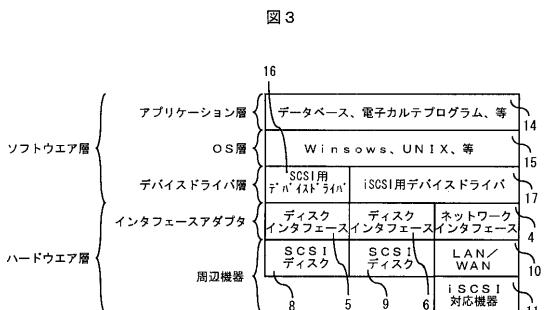
【図1】



【図2】



【図3】

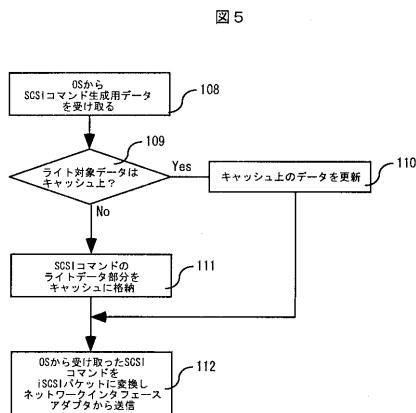


【図4】

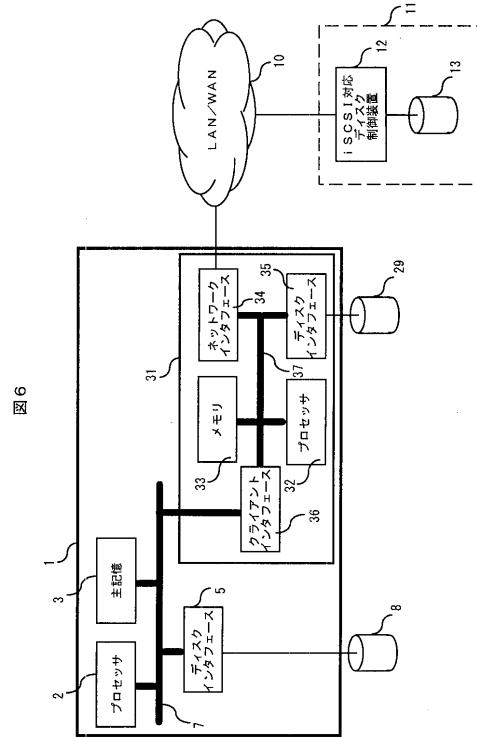
図4

iSCSI用キャッシュ用ディスク用ブロック番号		iSCSI用ディスクの識別番号		iSCSIディスクのブロック数		iSCSIの通し番号	
0	0	0	0	0	10	0	0
0	10	8	0	88	8	1	1
0	18	20	0	20	20	2	2
0	1020	15	0	2020	15	98	98
0	1035	2	0	38	2	99	99

【図5】

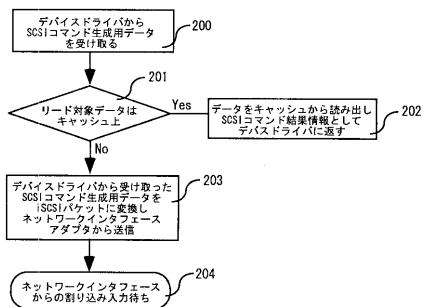


【図6】

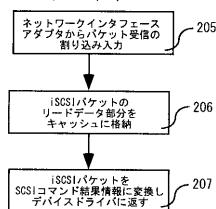


【図7】

図7(a)

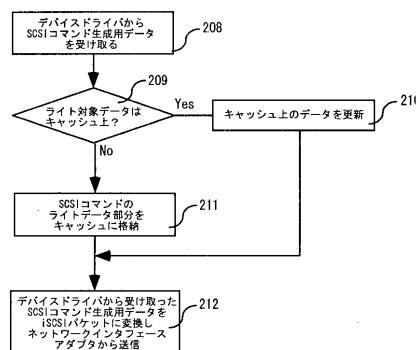


(b)

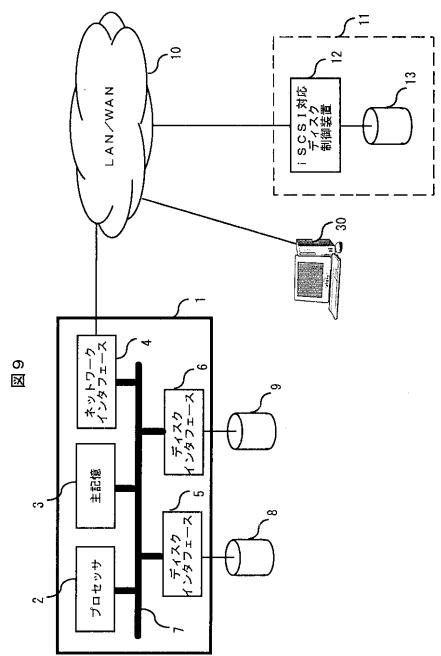


【図8】

図8



【図9】



フロントページの続き

(51)Int.Cl.

F I

G 0 6 F	12/00	5 4 5 A
G 0 6 F	12/00	5 1 4 K
G 0 6 F	3/06	3 0 2 A

審査官 清木 泰

(56)参考文献 特開2000-293316(JP,A)

特表2005-527007(JP,A)

特開2003-162444(JP,A)

特開2002-324000(JP,A)

特開2001-184294(JP,A)

特開平09-167116(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F12/08-12/12

G06F13/00,351-13/14

G06F12/00-12/00,549

G06F 3/06- 3/08