

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2015年8月13日(13.08.2015)



(10) 国際公開番号
WO 2015/118865 A1

- (51) 国際特許分類:
G06F 12/00 (2006.01)
- (21) 国際出願番号: PCT/JP2015/000504
- (22) 国際出願日: 2015年2月4日(04.02.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2014-020317 2014年2月5日(05.02.2014) JP
- (71) 出願人: 日本電気株式会社(NEC CORPORATION)
[JP/JP]; 〒1088001 東京都港区芝五丁目7番1号
Tokyo (JP).
- (72) 発明者: 菅 真樹(KAN, Masaki); 〒1088001 東京都
港区芝五丁目7番1号日本電気株式会社内
Tokyo (JP). 鈴木 順(SUZUKI, Jun); 〒1088001 東京
都港区芝五丁目7番1号日本電気株式会社内
Tokyo (JP). 林 佑樹(HAYASHI, Yuki); 〒1088001 東
京都港区芝五丁目7番1号日本電気株式会社内
Tokyo (JP).
- (74) 代理人: 下坂 直樹 (SHIMOSAKA, Naoki); 〒
1088001 東京都港区芝五丁目7番1号日本電気
株式会社内 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,
BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN,
IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR,
LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,
PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,
SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保
護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW,
MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユー
ラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨー
ロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC,
MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR),

[続葉有]

(54) Title: INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING SYSTEM, AND DATA ACCESS METHOD

(54) 発明の名称: 情報処理装置、情報処理システム及びデータアクセス方法

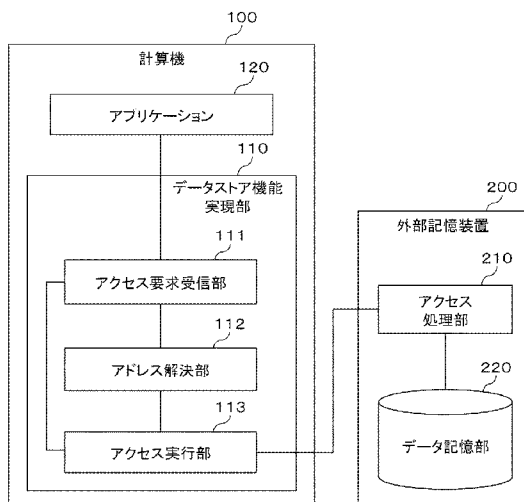
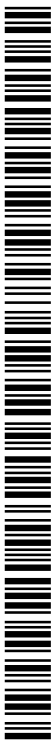


FIG. 2:
 100 Calculator
 110 Data store function actualization unit
 111 Access request reception unit
 112 Address resolution unit
 113 Access execution unit
 120 Application
 200 External storage device
 210 Access processing unit
 220 Data storage unit

(57) Abstract: An information processing device is provided that improves the responsiveness of a storage system and a storage device that is accessed via a network. This information processing device comprises: a means that calculates an address identifying an area, on the basis of an access identifier specifying data to be accessed, said area being in a storage device that stores a first data corresponding to the access identifier; and a means that obtains the first data included in second data being data for the area and read on the basis of the calculated address, on the basis of management information included in the second data, and executes a first data operation on the second data.

(57) 要約: 本発明は、ネットワークを介してアクセスされる記憶装置及び記憶システムの応答性能が向上する情報処理装置を提供する。その情報処理装置は、アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスをそのアクセス識別子に基づいて算出する手段と、算出されたそのアドレスに基づいて読み出される、その領域のデータである、第2のデータに含まれる管理情報に基づいて、その第2のデータに含まれるその第1のデータを取得し、その第2のデータに対してその第1のデータの操作を実行する手段と、を備える。



WO 2015/118865 A1

OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG). 添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

情報処理装置、情報処理システム及びデータアクセス方法

技術分野

[0001] 本発明は、記憶装置及び記憶システムに対するデータアクセスの技術に関する。

背景技術

[0002] 記憶装置及び記憶システムに対するデータアクセスの制御に関するさまざまな技術が知られている。

[0003] 例えば、単一あるいは複数の計算機によって構成されるデータストアシステム（例えば、データベースシステムやファイルシステム、キャッシュシステム）がある。近年では、分散ストレージシステムが、そのようなシステムに、しばしば適用されている。その分散ストレージシステムは、ネットワークを介して接続された複数の汎用的な計算機を含む。

[0004] そして、その分散ストレージシステムは、それらの計算機に搭載された記憶装置を用いて、データの格納及びデータの提供を行う。その記憶装置は、例えばHDD（Hard Disk Drive）、主記憶（例えば、DRAM：Dynamic Random Access Memory）などである。

[0005] 上述のような分散ストレージシステムでは、いずれの計算機にデータを配置し、いずれの計算機によってデータを処理するのかを、ソフトウェアまたは特別なハードウェアが決定する。このようなアーキテクチャは、Shared nothing Architectureと呼ばれる。

[0006] SAN（Storage Area Network）は、例えば複数のサーバ間でFC（Fibre Channel）などのネットワークを介して、記憶装置を共有する。例えば、データストアシステムは、SANにより共有される記憶装置を用いて実現される。

- [0007] SANにおいて、複数の計算機間でデータを共有してシステムを実現するには、Shared Everythingアーキテクチャに基づくソフトウェアを用いる必要がある。例えば、ファイルシステムの場合、そのソフトウェアは、SANファイルシステムなどである。また、データベースシステムの場合、そのソフトウェアは、Oracle（登録商標） RAC（Real Application Clusters）（登録商標）などがある。
- [0008] Shared Everythingアーキテクチャは、通常、FCやiSCSI（internet Small Computer System Interface）を用いて実現される。FCやiSCSIは、通信遅延が大きい。そのため、記憶装置としても応答性能に優れた記憶装置が使われるケースは少なく、HDDなどの応答性能に劣る記憶装置が主に使われる。
- [0009] 一方、HDDは、シーケンシャルなアクセス性能に優れる。そのため、データベースなどのソフトウェアは、例えば、Write Ahead Logなどの方法を用いて更新情報だけをシーケンシャルに書き出すことによって、その共有される記憶装置の低い性能をカバーする。
- [0010] 近年では、高速かつ汎用的なPCI-e（Peripheral Component Interconnect-Express）インタフェースで、計算機にSSD（Solid State Drive）などの高速な記憶装置を接続する構成が、用いられる。このような構成は、高速な記憶装置に低遅延でアクセスすることを可能にする。従って、このような構成は、SAN上のストレージに対するキャッシュなどの用途に用いられる。
- [0011] このような構成のPCI-eデバイスをExpEther（登録商標）などにより複数ホスト間で共有する技術を用いることにより、Shared Everythingアーキテクチャを実現することができる。また、このような構成を取ることで、上述のSAN上のストレージと比較して、低遅延でのストレージ共有の実現が可能になる。

- [0012] 特許文献1は、分散型システムの一例を開示する。特許文献1の分散型システムは、ネットワークで接続された複数のノードにより、識別子によって特定されるレコードが分散管理される。そのノードは、レコード記憶手段とインデクス付与手段とレコード取得手段とを備える。
- [0013] そのレコード記憶手段は、そのノードが管理する複数のレコードを識別子の任意の範囲ごとに集合体として記憶する。
- [0014] そのインデクス付与手段は、その集合体に対して、その集合体の範囲に含まれる識別子を用いたインデクスを付与する。
- [0015] そのレコード取得手段は、レコード取得要求に対して、そのインデクスを参照することにより、そのレコード取得要求にて要求されたレコードをそのレコード記憶手段から取得する。
- [0016] 特許文献2は、ストレージシステムの一例を開示する。特許文献2のストレージシステムは、複数のホストと、ボリューム仮想化装置と、複数のストレージと、管理クライアントと、ストレージ管理サーバとを備える。
- [0017] そのホストとそのストレージとは、ボリューム仮想化装置を挟んで、LAN (Local Area Network) 等の通信ネットワークを介して接続される。
- [0018] ボリューム仮想化装置は、そのホストに、それらの複数のストレージをひとつの仮想的なストレージ装置として認識させる。
- [0019] ストレージ管理サーバは、それらの複数のストレージ上のボリューム配置を制御する。

先行技術文献

特許文献

- [0020] 特許文献1：特開2012-168781号公報
特許文献2：特開2013-033515号公報

発明の概要

発明が解決しようとする課題

- [0021] しかしながら、上述した先行技術文献に記載された技術においては、外部記憶装置の応答性能が所望の値よりも低いという問題点がある。
- [0022] その理由は、ネットワークを介してアクセスされる記憶装置（特許文献1のノード及び特許文献2のストレージ装置）は、記憶デバイスへのアクセス時間に加え、ネットワーク上のコマンド及び応答の通信による遅延が影響するからである。
- [0023] 即ち、特許文献1や特許文献2の技術により、データベースやKVS（Key Value Store）を実現した場合、所望のデータを取得するために、アクセス元の計算機、外部記憶装置及び中間装置間などで複数回の通信が必要になる場合があるからである。即ち、これらのネットワークを介してアクセスされる記憶装置は、ホスト（アクセス元）上の記憶手段と比較して、通信遅延の影響によって応答性能が劣るからである。
- [0024] 本発明の目的は、上述した問題点を解決する情報処理装置、情報処理システム、データアクセス方法、及び、そのためのプログラムあるいはそのプログラムを記録したコンピュータ読み取り可能な非一時的記録媒体を提供することにある。

課題を解決するための手段

- [0025] 本発明の一様態における情報処理装置は、アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出するアドレス解決手段と、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行するアクセス実行手段と、を含む。
- [0026] 本発明の一様態におけるデータアクセス方法は、コンピュータが、アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、算出された前記アドレスに基づいて読み出される、前記領域のデ

ータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行する。

[0027] 本発明の一様態におけるコンピュータ読み取り可能な非一時的記録媒体は、アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行する処理をコンピュータに実行させるプログラムを記録する。

[0028] 本発明の一様態における情報処理システムは、アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する記憶装置上と、前記記憶装置の領域を特定するアドレスを前記アクセス識別子に基づいて算出するアドレス解決手段と、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行するアクセス実行手段と、を含む情報処理装置と、を含む。

発明の効果

[0029] 本発明は、ネットワークを介してアクセスされる記憶装置及び記憶システムの応答性能を、向上することが可能になるという効果がある。

図面の簡単な説明

[0030] [図1]本発明の第1の実施形態に係る情報処理システムの構成を示すブロック図である。

[図2]第1の実施形態における計算機及び外部記憶装置の内部構成を示すブロック図である。

[図3]第1の実施形態に係る計算機を実現するコンピュータのハードウェア構

成を示すブロック図である。

[図4]第1の実施形態における計算機及び外部記憶装置の動作を示すフローチャートである。

[図5]第1の実施形態におけるデータ記憶部のアドレス空間の一例を示す図である。

[図6]第1の実施形態におけるページの構造の一例を示す図である。

[図7]第1の実施形態における計算機及び外部記憶装置の動作を示すフローチャートである。

[図8]本発明の第2の実施形態に係る情報処理装置の構成を示すブロック図である。

発明を実施するための形態

[0031] 本発明を実施するための形態について図面を参照して詳細に説明する。なお、各図面及び明細書記載の各実施形態において、同様の構成要素には同様の符号を付与し、適宜説明を省略する。

[0032] <<<第1の実施形態>>>

図1は、本発明の第1の実施形態に係る情報処理システム10の構成の一例を示すブロック図である。

[0033] 図1に示すように、本実施形態に係る情報処理システム10は、計算機（情報処理装置とも呼ばれる）100と外部記憶装置200とネットワーク300とを備える。図1に示す例に係わらず、計算機100及び外部記憶装置200のそれぞれの台数は、1及び複数の任意の台数であってよい。また、外部記憶装置200は、情報処理装置や複数の記憶装置などを含む記憶システムであってよい。

[0034] ===計算機100===

計算機100は、外部記憶装置200に対するデータアクセスを制御し、データストア機能を実現する。計算機100は、演算装置（例えば、CPU（Central Processing Unit））、記憶部、ネットワーク300に接続するためのインターフェース部などを備える、コンピュー

タ装置（情報処理装置とも呼ばれる）である。そのインタフェースは、例えば、ネットワークカードやホストバスアダプタ、ExpEther機能を備えたカードなどである。

[0035] ===外部記憶装置200===

外部記憶装置200は、計算機100と結合するためのインタフェース手段と、記憶デバイスと、記憶デバイスへのアクセス処理を行う手段とを、少なくとも備える。その記憶デバイスは、フラッシュメモリやDRAM、MRAM (Magnetoresistive Random Access Memory)、HDDなどである。

[0036] そのインタフェース手段は、Ethernet（登録商標）やFibre Channel、InfiniBandなどを制御する。例えば、外部記憶装置200は、ExpEtherによって計算機100と結合される場合には、ExpEther機能を備えるカード、PCI-eインタフェースが搭載された記憶デバイスを搭載する。

[0037] ネットワーク300は、計算機100及び外部記憶装置200を相互に接続する。そして、ネットワーク300は、計算機100及び外部記憶装置200間の、データ、制御メッセージ及びその他のメッセージなどを媒介する。ネットワーク300は、例えばEthernet、InfiniBand、またはこれらを利用するTCP/IP (Transmission Control Protocol/Internet Protocol)、RDMA (Remote Direct Memory Access) などの上位プロトコルにより実現される。また、ネットワーク300は、Fibre ChannelやFCoE (Fibre Channel over Ethernet)、ExpEtherなどによって実現されてもよい。ネットワーク300は、これらに限られず、任意の方法で実現されてよい。

[0038] 図2は、計算機100及び外部記憶装置200の内部構成を示すブロック図である。図2に示す各構成要素は、ハードウェア単位の回路でも、コンピ

ユーザ装置の機能単位に分割された構成要素でもよい。ここでは、図2に示す構成要素は、コンピュータ装置の機能単位に分割された構成要素として説明する。

[0039] なお、図2では、図1に示す情報処理システム10のネットワーク300が省略されている。また、図2は、計算機100及び外部記憶装置200のそれぞれを1つずつ含む情報処理システム10の構成を示す。

[0040] ===計算機100の内部構成===

計算機100は、データストア機能実現部110とアプリケーション120とを含む。

[0041] データストア機能実現部110は、例えば、(データベースやKVS(Key Value Store)など)を実現するための、演算装置(後述するCPU701)上で動作するソフトウェア(データストアソフトウェア)である。

[0042] アプリケーション120はデータストアを利用する任意のソフトウェアである。アプリケーション120は、データストア機能実現部110が動作する計算機100とは別の計算機で動作してもよい。ここで、そのデータストアは、データストア機能実現部110により実現される。

[0043] データストア機能実現部110は、アクセス要求受信部111とアドレス解決部112とアクセス実行部113とを含む。

[0044] ===アクセス要求受信部111===

アクセス要求受信部111は、アプリケーション120からのデータアクセス要求命令を受け付ける。なお、アクセス要求受信部111は、アドレス解決部112の一部として含まれてもよい。

[0045] そのデータアクセス要求命令は、データストアソフトウェアの機能(即ち、データストア機能実現部110が提供するデータストアの機能)によって異なる。例えば、そのデータストアがデータベースの場合、そのデータアクセス要求命令は、SQL(Structured Query Language)で指定されるようなデータ操作命令である。また、そのデータストア

アがKVSの場合、そのデータアクセス要求命令は、Keyに対応するvalueを取得したり、登録・更新したりするような処理命令である。

[0046] ===アドレス解決部112===

アドレス解決部112は、アクセス要求受信部111が受け付けたデータアクセス要求命令を解釈し、そのデータアクセス要求命令に対応するデータ（第1のデータ）を記憶する、ページ（領域とも呼ばれる）を特定するアドレスを算出する。そのページは、外部記憶装置200のデータ記憶部220上の部分的な領域である。アドレス解決部112は、そのアドレスをアクセス識別子に基づいて算出する。そのアクセス識別子は、データストア機能実現部110が提供するデータストアの機能において、その第1のデータを特定する情報である。アドレス解決部112は、そのアクセス識別子を、そのデータアクセス要求命令を解釈して取得する。

[0047] その「ページを特定するアドレス」は、外部記憶装置200のデータ記憶部220（後述）の物理アドレスであってよい。また、その「ページを特定するアドレス」は、外部記憶装置200のアクセス処理部210（後述）においてデータ記憶部220（後述）の物理アドレスに変換可能な、論理アドレスであってもよい。その「ページを特定するアドレス」を算出する動作の詳細については後述する。

[0048] ===アクセス実行部113===

アクセス実行部113は、アドレス解決部112によって算出されたアドレスに対して、データアクセスリクエスト（例えば、read要求やWrite要求など）を発行する。

[0049] 即ち、アクセス実行部113は、アドレス解決部112によって算出されたアドレスを指定して、そのページのデータ（第2のデータ）を読み出す要求（read要求）、及びそのページにデータを書き込む要求（Write要求）を発行する。

[0050] また、アクセス実行部113は、読み出された第2のデータに含まれる管理情報に基づいて、読み出された第2のデータに含まれる第1のデータを取

得する。更に、アクセス実行部 113 は、読み出された第 2 のデータに含まれる管理情報に基づいて、読み出された第 2 のデータに対し、第 1 のデータの操作（追加、削除及び更新）を実行する。管理情報の詳細は、後述する。

[0051] ===外部記憶装置 200 の内部構成===

外部記憶装置 200 は、アクセス処理部 210 とデータ記憶部 220 とを含む。

[0052] ===アクセス処理部 210===

アクセス処理部 210 は、計算機 100 のアクセス実行部 113 からデータアクセスリクエストを受け取り、そのデータアクセスリクエストに含まれるアドレスに基づいて、データ記憶部 220 に記憶されるデータを取得あるいは操作し、計算機 100 へ応答する。

[0053] また、アクセス処理部 210 は、データ記憶部 220 の記憶媒体特性に基づいた制御などを行う機能を備える。アクセス処理部 210 は、一般的には何らかの集積回路や FPGA (Field Programmable Gate Array) 上のロジックとして実現されている。具体的には、アクセス処理部 210 は、フラッシュメモリのコントローラや、DRAM コントローラなどである。

[0054] ===データ記憶部 220===

データ記憶部 220 は実際の記憶媒体であり、フラッシュメモリや DRAM、HDD、またはそれらの組み合わせなどから構成される。

[0055] データ記憶部 220 は、アクセス識別子に対応する第 1 のデータを含む第 2 のデータを、そのアクセス識別子に対応するアドレスで特定されるページに記憶する。

[0056] 以上が、計算機 100 及び外部記憶装置 200 それぞれの機能単位の各構成要素についての説明である。

[0057] 次に、計算機 100 のハードウェア単位の構成要素について説明する。

[0058] 図 3 は、本実施形態における計算機 100 を実現するコンピュータ 700 のハードウェア構成を示す図である。

- [0059] 図3に示すように、コンピュータ700は、CPU（プロセッサとも呼ばれる）701、記憶部702、記憶装置703、入力部704、出力部705及び通信部706を含む。更に、コンピュータ700は、外部から供給される記録媒体（または記憶媒体）707を含む。例えば、記録媒体707は、情報を非一時的に記憶する不揮発性記録媒体（非一時的記録媒体）である。また、記録媒体707は、情報を信号として保持する、一時的記録媒体であってもよい。
- [0060] CPU701は、オペレーティングシステム（不図示）を動作させて、コンピュータ700の全体の動作を制御する。例えば、CPU701は、記憶装置703に装着された記録媒体707から、そのプログラムやデータを読み込み、読み込んだそのプログラムやそのデータを記憶部702に書き込む。ここで、そのプログラムは、例えば、後述の図4及び図7に示すフローチャートにおける計算機100の動作を、コンピュータ700に実行させるためのプログラムである。
- [0061] そして、CPU701は、その読み込んだプログラムに従って、またその読み込んだデータに基づいて、図2に示すアクセス要求受信部111、アドレス解決部112及びアクセス実行部113として各種の処理を実行する。
- [0062] なお、CPU701は、通信網（不図示）に接続されている外部コンピュータ（不図示）から、記憶部702にそのプログラムやそのデータをダウンロードしてもよい。
- [0063] 記憶部702は、そのプログラムやそのデータを記憶する。記憶部702は、外部記憶装置200から受信するデータ及び外部記憶装置200へ送信するデータを記憶する手段を含んでよい。
- [0064] 記憶装置703は、例えば、光ディスク、フレキシブルディスク、磁気光ディスク、外付けハードディスク及び半導体メモリである。記憶装置703は、そのプログラムをコンピュータ読み取り可能に記憶する。また、記憶装置703は、そのデータを記憶してもよい。記憶装置703は、外部記憶装置200から受信するデータ及び外部記憶装置200へ送信するデータを記

憶する手段を含んでよい。

- [0065] 入力部704は、オペレータによる操作の入力や外部からの情報の入力を受け付ける。入力操作に用いられるデバイスは、例えば、マウスやキーボード、内蔵のキーボタン及びタッチパネルなどである。
- [0066] 出力部705は、例えばディスプレイで実現される。出力部705は、例えばGUI (GRAPHICAL User Interface) による入力要求や出力などの確認をするために用いられる。
- [0067] 通信部706は、ネットワーク300とのインタフェースを実現する。通信部706は、アクセス実行部113の一部として含まれる。
- [0068] 以上説明したように、図2に示す計算機100の機能単位のブロックは、図3に示すハードウェア構成のコンピュータ700によって実現される。但し、コンピュータ700が備える各部の実現手段は、上記に限定されない。すなわち、コンピュータ700は、物理的に結合した1つの装置により実現されてもよいし、物理的に分離した2つ以上の装置を有線または無線で接続し、これら複数の装置により実現されてもよい。
- [0069] なお、上述のプログラムのコードを記録した記録媒体707が、コンピュータ700に供給される場合、CPU701は、記録媒体707に格納されたそのプログラムのコードを読み出して実行してもよい。あるいは、CPU701は、記録媒体707に格納されたそのプログラムのコードを、記憶部702、記憶装置703またはその両方に格納してもよい。すなわち、本実施形態は、コンピュータ700 (CPU701) が実行するそのプログラム (ソフトウェア) を、一時的にまたは非一時的に、記憶する記録媒体707の実施形態を含む。なお、情報を非一時的に記憶する記憶媒体は、不揮発性記憶媒体とも呼ばれる。
- [0070] 以上が、本実施形態における計算機100を実現するコンピュータ700の、ハードウェア単位の各構成要素についての説明である。
- [0071] 次に本実施形態の動作について、図面を参照して詳細に説明する。
- [0072] 図4及び後述する図7は、本実施形態の計算機100及び外部記憶装置2

00の動作を示すフローチャートである。なお、これらのフローチャートによる計算機100の処理は、前述したCPU701によるプログラム制御に基づいて、実行されてもよい。また、処理のステップ名については、S101のように、記号で記載する。

[0073] ここでは、データストアは、Key Value Storeであるとする。図4は、1つのKeyに対応する第1のデータを読み込むためのデータアクセス要求命令（Readリクエスト）を処理する例を示す。また、図7は、1つのKeyに対応する何らかの第1のデータを登録・更新するためのデータアクセス要求命令（Updateリクエストやputリクエスト）を処理する例を示す。なお、リレーショナルデータベースに対するSQL処理のような場合についても、その処理は本質的に同様であり、多少の変形で実現できることは明らかである。

[0074] なお、ここでは説明が複雑になることを回避するために、エラー処理などへの対応を記述しない。本実施形態を実施する際には、図4及び図7に示すフローチャートに対して、物理的・論理的障害やユーザ・アプリケーションの利用ミスに対する例外的な処理が追加されてよい。

[0075] また、複数の計算機100が外部記憶装置200を共有し、複数の計算機100間で同一のレコードを更新する場合には、本実施形態に排他制御処理が導入されてよい。また、1つの計算機100内でも高いスループット性能を発揮するために同時並行的に処理する場合があり、そのような場合にも、本実施形態に排他制御処理が導入されてよい。

[0076] === Readリクエスト処理 ===

図4に示すように、データストアソフトウェアを利用するアプリケーション120は、データアクセス要求命令（ここではReadリクエスト）をデータストア機能実現部110へ発行する（ステップS101）。

[0077] 具体的には、アプリケーション120は、データストア機能実現部110が提供するAPI（Application Programming Interface）を呼び出すことで、そのデータアクセス要求命令を発行

してよい。

[0078] また、アプリケーション120は、http (hypertext transfer protocol) やJSON (JavaScript (登録商標) Object Notation) などの任意のプロトコルで通信することで、そのデータアクセス要求命令を発行してよい。この場合、アクセス要求受信部111は、そのプロトコルに対応するサーバとして動作してよい。

[0079] これらの例に係わらず、そのデータアクセス要求命令は、任意の方法でアプリケーション120からデータストア機能実現部110へ発行されてよい。

[0080] 次に、データストア機能実現部110のアクセス要求受信部111は、発行されたそのデータアクセス要求命令を受け付ける (ステップS102)。

[0081] 次に、アドレス解決部112は、アクセス要求受信部111が受け付けたそのデータアクセス要求命令に記載されているアクセス対象の第1のデータを特定する識別子 (以後、アクセス識別子と呼ぶ) を特定する (ステップS103)。

[0082] ここでは、そのアクセス識別子は、Key Value StoreにおけるKeyである。Key Value Storeの場合、例えば、データストア機能実現部110が提供するAPIとして、get命令が用意されている。そのget命令は、Keyに対応するレコードを取得する場合の、「get (Key1)」のような「Key1」を引数とするget命令である。この場合、この引数情報「Key1」が、アクセス対象データ (Value) を示すアクセス識別子である。なお、アドレス解決部112は、上記に限らず、get命令のさまざまな亜種に対応してよい。

[0083] また、データストアがリレーショナルデータベースであって、データアクセス要求命令がSQL命令であるような場合、アドレス解決部112は、そのSQL命令を解釈して、アクセス対象データやデータベース内での実行命令に変換する機構を含んでよい。例えば、その機構は、クエリパーサやクエ

リオプティマイザの一部などであってよい。

[0084] このリレーショナルデータベースの場合、アクセス識別子は、例えば、テーブルを示す情報（例えば、SELECT文で指定されるテーブル名）とレコードID（例えば、SELECT文で指定されるIDフィールドの値）である。なお、アクセス識別子は、上述の例に係わらず、リレーショナルデータベースの実装に依存するものであってよい。

[0085] 次に、アドレス解決部112は、特定されたアクセス識別子から外部記憶装置200のデータ記憶部220のアドレスを算出する（ステップS104）。

ここで、そのアドレスの算出方法について、図面を参照して詳細に説明する。

[0086] 図5は、データ記憶部220のアドレス空間221の一例を示す図である。一般的に、論理あるいは物理アドレスを指定することで、データ記憶部220に含まれる記憶媒体のアクセス先を特定することができる。

[0087] 図5に示すアドレス空間221は、例えば、データ記憶部220の全体あるいは一部の記憶領域である。なお、一部の記憶領域である理由は、例えば、複数のサービスで分割して使ったり、システム実現のための管理情報を格納する領域用に確保したりするためである。このアドレス空間221が、第1のデータ（即ち、第1のデータを含む第2のデータ）の格納先として、利用される。

[0088] 本実施形態では、データ記憶部220のアドレス空間221が任意のサイズでページ（図5に示すページ223）に分割される。そして、アクセス処理部210は、ページ223のそれぞれに対応するID（IDENTIFIER）を用いて、アドレス空間221をアクセスする。

[0089] そのID（以後、ページIDと呼ぶ）は、例えば、各行の左端から右へ及び上の行から下の行へ向かって、開始を「0」として「1」ずつ増加する連続する数値が、ページごとに付与される。例えば、最上行、最左端のページ223のページIDは「0」、最上行、左から5番目のページ223のペー

ジIDは「4」、上から2行目、左から4番目のページ223のページIDは「12」である。

[0090] ページIDに対応する物理アドレスは、アドレス空間221の最初のアドレスと、ページIDと、ページサイズ（例えば、バイト数で示すページ223の容量）とに基づいて、一意に算出可能である。即ち、データ記憶部220のアドレスを算出するために、計算機100は、アドレス空間221の最初のアドレスと、ページIDと、ページサイズの情報を保持すればよい。アドレス空間221としては、連続したページが確保されることが望ましいが、アドレス空間221が不連続なページ223で構成されてもよい。その場合、計算機100は、連続するページ223における先頭のページ223のページIDと開始アドレスとを保持してよい。

[0091] アドレス解決部112は、以下に示すように、アクセス識別子に基づいてアクセス先のページIDを（即ち、ページ223を）特定する。

[0092] 第1に、アドレス解決部112は、ステップS103で特定されたアクセス識別子（例えば、Keyの値）を数値に変換する。例えば、アドレス解決部112は、一般的なハッシュ関数（md5など）を用いて、Keyの値を数値に変換する。なお、アドレス解決部112は、任意の数式に基づく、数値への変換関数やその処理プログラムのソフトウェアを用いてもよい。

[0093] 第2に、アドレス解決部112は、変換して得たその値（ハッシュ値）を、アドレス空間221の総ページ数で除算し、その剰余をページIDとする。なお、その総ページ数は、アドレス空間221の容量をページサイズで除算することでも算出できる。

[0094] 第3に、アドレス解決部112は、算出したそのページIDに基づいて、ページID×ページサイズ+開始アドレスの演算を実行することにより、アクセス識別子（Keyの値）に対応するページ223を示すアドレスを算出する。

[0095] 上述のアドレス解決部112を含むことにより、計算機100は以下の情報を保持しているだけで、アクセス識別子に対応するページ223をアクセ

することができる。その情報は、外部記憶装置200のアドレス空間221の開始アドレスとハッシュ関数と、ページサイズ及びアドレス空間221のサイズを示す情報とである。但し、そのページ223内の容量を使い尽くすような場合はこの限りではない。ページ223内の容量を使い尽くすような場合については、後述する。

[0096] ページサイズ及びアドレス空間221のサイズを示す情報は、ページサイズ及びアドレス空間221のサイズそのものであってよい。また、ページサイズ及びアドレス空間221のサイズを示す情報は、アドレス空間221の総ページ数及びページサイズであってもよい。ページサイズ及びアドレス空間221のサイズを示す情報は、アドレス空間221の総ページ数及びアドレス空間221の総容量であってもよい。

[0097] 本実施形態のKey Value Storeの場合、計算機100は、その情報に基づいて、Keyの値（アクセス識別子）が格納されるページ223をアクセスすることができる。また、これらの情報はシステム開始時から基本的（障害時や、複数の外部記憶装置200で動作したときの外部記憶装置200の追加や削除など）に不変である。従って、例えば複数の計算機100で外部記憶装置200を共有する場合でも、たかだかその情報を計算機100間で共有することで、外部記憶装置200を共有することができる。

[0098] 即ち、外部記憶装置200を共有するための計算機100間で情報のやりとりが不要であり、データストア機能実現部110における処理の高速化を図ることができる。

[0099] 以上が、そのアドレスの算出方法の説明である。

[0100] 図4に戻って、次に、アクセス実行部113は、ステップS104で算出されたアドレスとページサイズのデータ長とを指定して、外部記憶装置200に対してRead要求（データアクセスリクエスト）を送信する（ステップS105）。

[0101] 次に、外部記憶装置200のアクセス処理部210は、このRead要求

を受け付け、データ記憶部220に対してRead処理を実行する（ステップS106）。

[0102] 次に、アクセス処理部210は、このRead処理で得られた処理結果（ここでは、Readデータ、即ち、ページ223のデータ（第2のデータ））をアクセス実行部113に送信する（ステップS107）。

[0103] 次に、アクセス実行部113は、受信したそのReadデータ（ページ223のデータ）から、アドレス解決部112によって抽出されたアクセス識別子によって指定されるデータレコード（第1のデータ）を抽出する。続けてアクセス実行部113は、抽出したそのデータレコードをアクセス要求受信部111に出力する（ステップS108）。

[0104] ここで、そのReadデータから、そのアクセス識別子に対応するそのデータレコード（ここでは、Key Value StoreにおけるValue）を抽出する方法について説明する。

[0105] 図6は、ページ223の構造の一例を示す図である。図6に示すように、ページ223は、データレコード225及び管理情報226を含む。

[0106] データレコード225は、「Key 3」に対応する「Value」のデータである。

[0107] 管理情報226は、ページ223内に格納されている「Key」の情報と、そのKeyに対応する「value」が格納されているページ223内の位置を示すポインタを含む。即ち、管理情報226は、第2のデータに含まれる第1のデータに対応するアクセス識別子と、その第1のデータのページ223内における位置を示すポインタと、を含む。

[0108] 図6では、管理情報226は、ページ223内の末尾側に格納されている。管理情報226は、ページ223の末尾側に限らず、任意のあらかじめ定められた位置に格納されてよい。

[0109] 図6に示すように、ページ223内に「Key 1」、「Key 2」及び「Key 3」のそれぞれに対応するvalueが格納されている場合、管理情報226は、例えば以下の情報を含む。その情報は、「Key 1 : 0 : x x

、Key 2 : a 1 : y y、Key 3 : a 2 : z z」である。なお、「ページ 2 2 3 内に「Key 1」、「Key 2」及び「Key 3」に対応するデータレコード 2 2 5 格納される」ことは、「「Key 1」、「Key 2」及び「Key 3」のそれぞれのハッシュ値に基づいて求められるページIDが同じである」ことを示す。

[0110] 即ち、管理情報 2 2 6 は、Key ごとの「Key の値 : ポインタ : レコードサイズ (bytes)」を含む。

[0111] 例えば、アクセス実行部 1 1 3 が Key 2 の value にアクセスする場合、アクセス実行部 1 1 3 は、まず管理情報 2 2 6 から Key 2 のポインタ「a 1」を取得する。次に、アクセス実行部 1 1 3 は、「a 1」の位置から、yy bytes 分のデータに対して、アクセス（例えば、読み取り）を実行する。

[0112] 上述の管理情報 2 2 6 は、データレコード 2 2 5 が可変長である場合の管理情報の例を示す。レコードサイズが、システムとして固定（データレコード 2 2 5 が固定長）である場合、レコードサイズの情報は不要であり、管理情報の容量が削減される。

[0113] ステップ S 1 0 8 において、アクセス実行部 1 1 3 は、探索する Key に対応する情報が見つかるまで、場合によっては管理情報 2 2 6 を全部読む必要がある。アクセス実行部 1 1 3 が、1 つ程度のページ 2 2 3 内のデータレコード 2 2 5 に対応する情報を検索するだけの場合、管理情報 2 2 6 のように単純な構成であってよい。また、管理情報の構造は、管理情報 2 2 6 の例に限らず、Key 値をより検索しやすい構造（例えば、Key 値の昇順／降順でソートされた構造やインデックス構造など）であってもよい。

[0114] なお、本実施形態では、アクセス実行部 1 1 3 は、ページ単位のデータ（第 2 のデータ）を外部記憶装置 2 0 0 から受信する。そして、アクセス実行部 1 1 3 は、ページ 2 2 3 から所望のデータレコード（第 1 のデータ、例えば value）2 2 5 を取り出す処理は、計算機 1 0 0 上のメモリ（例えば、図 3 に示す記憶部 7 0 2）を使用して実行する。例えば、アクセス実行部

113は、ページ223のデータを取得し、取得したそのデータを計算機100上の記憶部702にコピーし、記憶部702に格納された管理情報226を走査し、記憶部702に格納されたデータレコード225にアクセスする。

[0115] 以上が、アクセス識別子に対応するデータレコード225を抽出する方法の説明である。

図4に戻って、次に、アクセス要求受信部111は、リクエスト元のアプリケーション120に対して、取得したデータレコード225をReadリクエストの応答として、出力する（ステップS109）。

[0116] ===Write処理===

Write系のデータアクセス要求命令（Updateリクエストやputリクエスト）の場合、計算機100は、まずページ223のデータを取得する。次に、計算機100は、例えば記憶部702上で、そのデータを更新する。次に、計算機100は、その更新したデータについて、ページ単位のWrite処理を行う。

[0117] 従って、図7において、ステップS101からステップS107までの動作は、図4に示すステップS101からステップS107までの動作と同一である。

[0118] 次に、アクセス実行部113は、アクセス対象のデータレコード225を特定するKey（アクセス識別子）が管理情報226内にあるかどうかを確認する。続けて、アクセス実行部113は、その「Key」が存在する場合、図4に示すステップS108における処理と同様に、その「Key」に対応する「Value」へのポインタを取得する。そして、アクセス実行部113は、そのポインタとデータアクセス要求命令とに基づいてそのページ223のデータ（第2のデータの「value」及び管理情報226）を更新する（ステップS121）。

[0119] なお、更新前後でValueのサイズ（レコードサイズ）が異なる場合、アクセス実行部113は、更新後のValueを、更新前のValueとは

別の場所書き込むようにしてよい。また、データレコード225が更新あるいは削除されることで、ページ223の中に利用不可能なサイズの空き領域が多数できてしまう場合が考えられる。そのような場合には、計算機100は、ガベージコレクション相当の処理を実行してよい。

[0120] なお、ステップS121におけるアクセス実行部113の動作は、データストア機能実現部110としての仕様によって異なる。

[0121] 例えば、データストア機能実現部110が、put (Key、 value) 関数で「Key」に対応する「value」を追加する、というAPIを備えているとする。その場合、そのKeyに対応するデータレコード225がデータ記憶部220に既に存在する場合について、例えば、以下の2つの仕様と考えられる。そのひとつは、そのvalueを更新するという仕様である。また、他のひとつは、アプリケーション120に対して、「既にKeyが存在する」と応答し、valueをupdateしない、といった仕様である。これらについてどのように実装するかは、データストア機能実現部110の仕様策定によって決定される。例えば後者の仕様の場合、データストア機能実現部110は、「value」は書き換えられず、アプリケーション120に応答を返す、といった動作をする。

[0122] 次に、アクセス実行部113は、更新されたページ223のデータとステップS104で算出されたアドレスとを指定し、外部記憶装置200に対してWrite要求（データアクセスリクエスト）を送信する（ステップS125）。

[0123] 次に、外部記憶装置200のアクセス処理部210は、このWrite要求を受け付け、データ記憶部220に対してWrite処理を実行する（ステップS126）。

[0124] 次に、アクセス処理部210は、このWrite処理で得られた処理結果（ここでは、Write成功や失敗といった結果情報）をアクセス実行部113に送信する（ステップS127）。

[0125] 次に、アクセス実行部113は、受信したその処理結果をアクセス要求受

信部 1 1 1 に出力する（ステップ S 1 2 8）。

[0126] 次に、アクセス要求受信部 1 1 1 は、リクエスト元のアプリケーション 1 2 0 に対して、取得したその処理結果を *Write* リクエストの応答として、出力する（ステップ S 1 2 9）。

[0127] 以上が、本実施形態の動作の説明である。

[0128] 次に、本実施形態において、システム構成の差異や特異なケースに対応する例を説明する。

[0129] ===部分的な *Write* ===

Write 処理において、部分的な更新で済む場合、ページ単位のアクセスをする必要がない場合がある。例えば、1つの計算機 1 0 0 が外部記憶装置 2 0 0 を専有し、かつその計算機 1 0 0 が内部の排他制御を実行している場合、データの書き込みが必要なページ 2 2 3 内の領域は、更新される部分だけである。従って、管理情報 2 2 6 の更新が不要である場合、その *Write* 処理の対象であるデータだけが、書き換えられればよい。

[0130] この場合、アクセス実行部 1 1 3 は、ステップ S 1 2 1 の処理をスキップし、ステップ S 1 2 5 及びステップ S 1 2 6 において対象のデータレコード 2 2 5 の部分だけを書き換える。アクセス実行部 1 1 3 がこのように動作することで、データストア機能実現部 1 1 0 における *Write* 処理の性能を向上させることができる。

[0131] また、データストア機能実現部 1 1 0 は、複数の計算機 1 0 0 がある1つの外部記憶装置 2 0 0 を共有する場合にも、複数の計算機 1 0 0 間においてページ単位で排他手続きを取り、更新される部分だけについての *write* 処理を実行してもよい。ここで、その更新される部分は、更新対象のデータレコード 2 2 5 である。

[0132] また、外部記憶装置 2 0 0 の機能（例えば、複数の命令を *Atomic* に実行できるような仕組み）が排他制御を実現し、データストア機能実現部 1 1 0 が、上述のような部分的な *write* 処理を実行してよい。

[0133] ===ページ 2 2 3 からデータがあふれる場合の処理===

例えば、新しいデータ（新しいアクセス識別子（例えば、K e y）に対応するデータ（例えば、V a l u e））が追加されることで、外部記憶装置 200 のデータ記憶部 220 のページ容量が不足してしまうことが考えられる。ここで、ページ容量は、ページサイズで示されるページの容量である。即ち、データレコード 225 のサイズが大きいほど、ページ I D が競合するデータレコード 225 が 1 つのページ 223 に入りきらない、という場合が増加する。

[0134] 従って、アクセス識別子に対応するデータのサイズの大きさを考慮し、ページサイズをそのデータのサイズに適する大きさにしておくことが望ましい。アクセス識別子に対応するデータのサイズは、データストア機能実現部 110 を利用するアプリケーション 120 によって異なる。このため、ページサイズは、利用アプリケーションの適性に合わせて、あらかじめ設定されてよい。

[0135] 利用されるハッシュ関数、レコードサイズに対応するページサイズ、及び総ページ数が適切であるという前提において、各ページ 223 はほぼ均等に使われる。従って、ページ 223 の容量が不足している場合は、全体の記憶容量自体も不足していると考えられる。

[0136] しかし、上述の前提が崩れたり、ある特定のページ 223 に格納されるレコードが偶然により集中したりすると、情報処理システム 10 全体での記憶容量には余裕があるにも拘わらず、特定のページ 223 の容量が不足してしまうことがある。

[0137] この場合、データストア機能実現部 110 は、アクセス識別子から算出されるページ I D で特定されるページ 223 とは異なる、別のページ 223 にデータを格納してよい。

[0138] 例えば、アクセス実行部 113 は、まず、アクセス識別子に基づいて算出されるページ I D で特定されるページ 223（第 1 のページ 223）を取得する。

[0139] 次にアクセス実行部 113 は、取得したその第 1 のページ 223 の、管理

情報 2 2 6 に、そのアクセス識別子の値が含まれているか否かを判定する。

[0140] そのアクセス識別子の値が含まれている場合、アクセス実行部 1 1 3 は、そのアクセス識別子に対応する情報に基づいて、処理を実行する。

[0141] そのアクセス識別子の値が含まれていない場合、アクセス実行部 1 1 3 は、別のページ 2 2 3（第 2 のページ 2 2 3）を取得して処理を実行する。その際、アクセス実行部 1 1 3 は、その第 1 のページ 2 2 3 の管理情報 2 2 6 にアクセス識別子と、データが実際に格納される、その第 2 のページ 2 2 3 のページ ID の情報とを格納する。

[0142] アクセス実行部 1 1 3 がページ容量の不足を検出するタイミングは、アクセス実行部 1 1 3 がそのページ 2 2 3 にデータレコード 2 2 5 を新しく格納したり、既存のデータレコード 2 2 5 を更新したりするタイミングである。ページ容量の不足が検出された時に、あふれたデータレコード 2 2 5 の格納先とされるページ 2 2 3 を選択する方法として、複数の方法が考えられる。

[0143] 例えば、データ記憶部 2 2 0 に、最初のアドレス空間 2 2 1 とは別の記憶領域に予備のページ 2 2 3 を用意しておく。アクセス実行部 1 1 3 は、その予備のページ 2 2 3 にあふれたレコードを追加する。

[0144] また、アクセス実行部 1 1 3 は、別のページ 2 2 3 を任意の方法（たとえばランダム）で選んでもよい。

[0145] また、第 1 のページ 2 2 3 内のどのレコードを第 2 のページ 2 2 3 に格納するかについてもさまざまな手法が考えられる。

[0146] 例えば、アクセス実行部 1 1 3 は、レコードサイズが大きいものを第 2 のページ 2 2 3 に格納する。また、アクセス実行部 1 1 3 は、更新頻度が低いレコード（例えば、管理情報に更新カウント値などを含む）を第 2 のレコードに格納してもよい。

[0147] 特定のアクセス識別子のレコードが存在するか否かの確認は、さまざまな用途で行われる。そのため、アクセス実行部 1 1 3 は、同一ページ 2 2 3 が算出されるアクセス識別子に対応する情報を、最初にアクセス先となるページ 2 2 3（即ち、アクセス識別子から計算されるページ 2 2 3）の管理情報

226に格納してよい。

[0148] このように構成することで、アクセス実行部113は、その特定のアクセス識別子のデータレコード225が存在するか否かについて、1回の外部記憶装置200へのアクセスで判定することができる。

[0149] ===キャッシュ===

アクセス実行部113は、計算機100に搭載されている記憶部702などを用いて、ページ223のデータをキャッシュしてよい。こうすることで、データストア機能実現部110の性能が向上する。1台の計算機100が特定の外部記憶装置200のアドレス空間221を専用で用いる場合、その1台の計算機100は、別の計算機100から外部記憶装置200上のデータを更新されることを考慮しなくてよい。このため、アクセス実行部113は、Readリクエストの処理時に、キャッシュデータの内容に基づいて応答してよい。これにより、往復遅延が削減される。

[0150] Write処理においては、以下の2つの条件を満足する場合、アクセス実行部113は、そのキャッシュ上のページ223にデータを書き込んだ時点で、アプリケーション120に対して応答を返してよい。その第1の条件は、1台の計算機100が特定の外部記憶装置200のアドレス空間221を専用で用いることである。その第2の条件は、データの永続性を担保する手段が、外部記憶装置200に格納することではないことである。

[0151] 但し、外部記憶装置200上のデータが最新データとして利用される場合には、オリジナル及びコピーされた全てのページ223へ、同期的に書き込み実行されることが必要である。このような同期的な書き込み処理によって、ある計算機100の障害時に別の計算機100を追加してデータストアシステムを復旧させることができる。即ち、耐障害性を重視する場合には、上述のように動作させる必要がある。

[0152] また、複数の計算機100で外部記憶装置200を共有させる場合には、別の計算機100からページ223に対する更新処理が発生する可能性がある。そのため、Readキャッシュを利用することができない。そこで、ア

アプリケーション120側やロードバランサなどを用いて、更新処理を動作させる計算機100がアクセス識別子ごとに同一に割り当てられるようにする。この場合、アクセス実行部113は、Readキャッシュを利用してもよい。この場合、アクセス実行部113がReadキャッシュを利用しても、不整合は起きない。

[0153] ===複数の計算機100間の排他制御===

図1に示すように計算機100が複数台である場合、同一ページ223へのWriteアクセスの排他制御を実行する。例えば、計算機100は、別の計算機100との間で2フェーズコミットのような方法を取ることで、その排他制御を実現してよい。

[0154] また、外部記憶装置200上で複数の処理をAtomicに実行し、排他制御が実現されてもよい。この場合、計算機100間での通信は不要である。例えば、外部記憶装置200は、管理情報にアクセス識別子ごとのバージョン番号を格納する。そして、外部記憶装置200は、そのバージョン情報を参照して、そのアクセス識別子に対応するデータレコード225が、既に変更されているか否かを確認する。既に変更されているのでなければ、外部記憶装置200は、そのデータレコード225を更新する。

[0155] ===複数の記憶装置===

複数の外部記憶装置200を用意し、外部記憶装置200へのアクセスを分散させることで、データストアシステムの性能をスケールアップさせることができる。

[0156] この場合、例えば、外部記憶装置200ごとに異なるページIDが付与されてよい。例えば、ページIDの0~1000が1台目の外部記憶装置200に、ページIDの1001~2000が2台目の外部記憶装置200に、付与されてよい。

この情報は、計算機100間で共有される。

[0157] また、アクセス実行部113は、ハッシュ値を計算した時点で、コンシステント・ハッシングのような方法でどの外部記憶装置200を使うかを決定

するようにしてよい。次に、アクセス実行部 113 は、その外部記憶装置 200 の中のどのページ ID を使用するかを決めてもよい。

[0158] ===複製===

信頼性確保のためにページ 223 の複製を生成することができる。

[0159] 例えば、アクセス実行部 113 は、格納先の外部記憶装置 200 をコンシステント・ハッシングのような方法で決定する際に、コンシステント・ハッシングのハッシュリングの隣のノードを必要数選択する。こうして、アクセス実行部 113 は、複数の外部記憶装置 200 を選択する。このような方法は既存技術で用いられているため、詳細な説明を省略する。

[0160] 上述した本実施形態における第 1 の効果は、ネットワーク 300 を介してアクセスされる外部記憶装置 200 及び記憶システムの応答性能を、向上することを可能にする点である。

[0161] その理由は、以下のような構成を含むからである。即ち、第 1 に、アドレス解決部 112 がアクセス識別子に基づいて、データ記憶部 220 上のアドレスを算出する。第 2 に、アクセス実行部 113 がそのアドレスに基づいて読み出したページ 223 に含まれる管理情報 226 を解釈することで、そのアクセス識別子に対応するデータを取得する。

[0162] 換言すると、その理由は、計算機 100 から外部記憶装置 200 へのコマンド回数を削減することにより通信遅延を削減するようにしたからである。

[0163] 上述した本実施形態における第 2 の効果は、複数の計算機 100 間での通信を低減することを可能にする点である。

[0164] その理由は、アドレス解決部 112 が、変更されることが比較的まれな、アドレス空間 221 の開始アドレスとハッシュ関数とページサイズ及びアドレス空間 221 のサイズを示す情報とに基づいてデータ記憶部 220 上のアドレスを算出するからである。

[0165] 上述した本実施形態における第 3 の効果は、データストア機能実現部 110 における Write 処理の性能を向上させることを可能にする点である。

[0166] その理由は、Write 処理においてページ単位のアクセスをする必要が

ない場合、アクセス実行部 113 が対象のデータレコード 225 の部分だけを書き換えるからである。

[0167] 上述した本実施形態における第 4 の効果は、以下の場合であっても、アクセス実行部 113 が、特定のアクセス識別子のデータレコード 225 が存在するか否かについて、1 回の外部記憶装置 200 へのアクセスで判定することを可能にする点である。その場合は、その特定のアクセス識別子から算出されるページ ID で特定されるページ 223 とは異なる、別のページ 223 にデータが格納される場合である。

[0168] その理由は、アクセス実行部 113 が、同一ページ 223 が算出されるアクセス識別子に対応する情報は、最初にアクセス先となるページ 223 の管理情報 226 に格納するからである。

[0169] 上述した本実施形態における第 5 の効果は、データストア機能実現部 110 の性能が、より向上することを可能にする点である。

[0170] その理由は、アクセス実行部 113 は、計算機 100 に搭載されている記憶部 702 などを用いて、ページ 223 のデータをキャッシュするからである。

[0171] 上述した本実施形態における第 6 の効果は、情報処理システム 10 が提供するデータストアの信頼性を向上することを可能にする点である。

[0172] その理由は、アクセス実行部 113 が、格納先の外部記憶装置 200 をコンシステント・ハッシングのような方法で決定する際に、コンシステント・ハッシングのハッシュリングの隣のノードを必要数選択するからである。

[0173] <<<第 2 の実施形態>>>

次に、本発明の第 2 の実施形態について図面を参照して詳細に説明する。以下、本実施形態の説明が不明確にならない範囲で、前述の説明と重複する内容については説明を省略する。

[0174] 図 8 は、本発明の第 2 の実施形態に係る情報処理装置 102 の構成を示すブロック図である。

[0175] 図 8 に示すように、本実施形態における情報処理装置 102 は、アドレス

解決部 1 2 2 とアクセス実行部 1 1 3 とを含む。

- [0176] アドレス解決部 1 2 2 は、記憶装置上のページ 2 2 3 を特定するアドレスを、アクセス対象のデータを特定するアクセス識別子に基づいて算出する。そのページ 2 2 3 は、そのアクセス識別子に対応するデータ（第 1 のデータ）を記憶する。
- [0177] アクセス実行部 1 1 3 は、図 2 に示すアクセス実行部 1 1 3 と同等である。
- [0178] 情報処理装置 1 0 2 のハードウェア構成は、図 3 に示すコンピュータ 7 0 0 と同等である。
- [0179] 上述した本実施形態における効果は、ネットワーク 3 0 0 を介してアクセスされる外部記憶装置 2 0 0 及び記憶システムの応答性能を、向上することを可能にする点である。
- [0180] その理由は、以下のような構成を含むからである。即ち、第 1 に、アドレス解決部 1 2 2 がアクセス識別子に基づいて、データ記憶部 2 2 0 上のアドレスを算出する。第 2 に、アクセス実行部 1 1 3 がそのアドレスに基づいて読み出したページ 2 2 3 に含まれる管理情報 2 2 6 を解釈することで、そのアクセス識別子に対応するデータを取得する。
- [0181] 換言すると、その理由は、計算機 1 0 0 から外部記憶装置 2 0 0 へのコマンド回数を削減することにより通信遅延を削減するようにしたからである。
- [0182] 以上の各実施形態で説明した各構成要素は、必ずしも個々に独立した存在である必要はない。例えば、複数個の任意のその構成要素が 1 個のモジュールとして実現されてよい。また、その構成要素の内の任意のひとつが複数のモジュールで実現されてもよい。また、その構成要素の内の任意のひとつがその構成要素の内の任意の他のひとつであってよい。また、その構成要素の内の任意のひとつの一部と、その構成要素の内の任意の他のひとつの一部とが重複してもよい。
- [0183] 以上説明した各実施形態における各構成要素及び各構成要素を実現するモジュールは、必要に応じ、可能であれば、ハードウェア的に実現されてよい

。また、各構成要素及び各構成要素を実現するモジュールは、コンピュータ及びプログラムで実現されてよい。また、各構成要素及び各構成要素を実現するモジュールは、ハードウェア的なモジュールとコンピュータ及びプログラムとの混在により実現されてもよい。

[0184] そのプログラムは、例えば、磁気ディスクや半導体メモリなど、コンピュータが読み取り可能な非一時的記録媒体に記録され、コンピュータに提供される。そして、そのプログラムは、コンピュータの立ち上げ時などに、非一時的記録媒体からコンピュータに読み取られる。この読み取られたプログラムは、そのコンピュータの動作を制御することにより、そのコンピュータを前述した各実施形態における構成要素として機能させる。

[0185] また、以上説明した各実施形態では、複数の動作をフローチャートの形式で順番に記載してあるが、その記載の順番は複数の動作を実行する順番を限定するものではない。このため、各実施形態を実施するときには、その複数の動作の順番は内容的に支障のない範囲で変更することができる。

[0186] 更に、以上説明した各実施形態では、複数の動作は個々に相違するタイミングで実行されることに限定されない。例えば、ある動作の実行中に他の動作が発生してよい。また、ある動作と他の動作との実行タイミングが部分的に乃至全部において重複してもよい。

[0187] 更に、以上説明した各実施形態では、ある動作が他の動作の契機になるように記載しているが、その記載はある動作と他の動作との全ての関係を限定するものではない。このため、各実施形態を実施するときには、その複数の動作の関係は内容的に支障のない範囲で変更することができる。また各構成要素の各動作の具体的な記載は、各構成要素の各動作を限定するものではない。このため、各構成要素の具体的な各動作は、各実施形態を実施する上で機能的、性能的、その他の特性に対して支障をきたさない範囲内で変更されてよい。

[0188] 上記の実施形態の一部または全部は、以下の付記のようにも記載されうるが、以下には限られない。

- [0189] (付記1) アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出するアドレス解決部と、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行するアクセス実行部と、を含む情報処理装置。
- [0190] (付記2) 前記アクセス実行部は、前記第1のデータの追加、削除及び更新のいずれかであるデータ操作を実行する際に、前記記憶装置から前記第1のデータに対応する前記アクセス識別子に基づいて前記第2のデータを取得し、取得された前記第2のデータに対して前記データ操作を実行し、前記データ操作を実行された前記第2のデータを前記記憶装置に書き込む付記1記載の情報処理装置。
- [0191] (付記3) 前記アドレス解決部は、前記アクセス識別子に対応する数値を算出し、
算出された前記数値と、前記記憶装置上の利用可能なアドレス空間の開始アドレスと、前記領域のサイズ及び前記アドレス空間のサイズを示す情報と、に基づいて前記アドレスを算出することを特徴とする付記1または2記載の情報処理装置。
- [0192] (付記4) 前記アドレス解決部は、前記アクセス識別子のハッシュ値を前記数値として算出することを特徴とする付記3記載の情報処理装置。
- [0193] (付記5) 前記管理情報は、前記第2のデータに含まれる前記第1のデータに対応するアクセス識別子と、前記第1のデータの前記領域内における位置を示すポインタと、の対応を示す情報を含むことを特徴とする付記1乃至4のいずれか1つに記載の情報処理装置。
- [0194] (付記6) 前記第1のデータに対応するアクセス識別子に基づいて算出された前記アドレスに基づいて読み出された領域に含まれる前記管理情報は、前記第1のデータが前記記憶装置上の他の領域に記憶されることを示す情報

、及び前記第1のデータが他の記憶装置上の領域に記憶されることを示す情報を含むことを特徴とする付記1乃至5のいずれか1つに記載の情報処理装置。

[0195] (付記7) 前記アドレスは、前記記憶装置上の物理アドレスであることを特徴とする付記1乃至6のいずれか1つに記載の情報処理装置。

[0196] (付記8) アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する記憶装置と、前記記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出するアドレス解決部と、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行するアクセス実行部と、を含む情報処理装置と、を含む情報処理システム。

[0197] (付記9) 前記アクセス実行部は、前記第1のデータの追加、削除及び更新のいずれかであるデータ操作を実行する際に、前記記憶装置から前記第1のデータに対応する前記アクセス識別子に基づいて前記第2のデータを取得し、取得された前記第2のデータに対して前記データ操作を実行し、前記データ操作を実行された前記第2のデータを前記記憶装置に書き込む付記8記載の情報処理システム。

[0198] (付記10) 前記アドレス解決部は、前記アクセス識別子に対応する数値を算出し、

算出された前記数値と、前記記憶装置上の利用可能なアドレス空間の開始アドレスと、前記領域のサイズ及び前記アドレス空間のサイズを示す情報と、に基づいて前記アドレスを算出することを特徴とする付記8または9記載の情報処理システム。

[0199] (付記11) 前記アドレス解決部は、前記アクセス識別子のハッシュ値を前記数値として算出することを特徴とする付記10記載の情報処理システム。

- [0200] (付記 1 2) 前記管理情報は、前記第 2 のデータに含まれる前記第 1 のデータに対応するアクセス識別子と、前記第 1 のデータの前記領域内における位置を示すポインタと、の対応を示す情報を含むことを特徴とする付記 8 乃至 1 1 のいずれか 1 つに記載の情報処理システム。
- [0201] (付記 1 3) 前記第 1 のデータに対応するアクセス識別子に基づいて算出された前記アドレスに基づいて読み出された領域に含まれる前記管理情報は、前記第 1 のデータが前記記憶装置上の他の領域に記憶されることを示す情報、及び前記第 1 のデータが他の記憶装置上の領域に記憶されることを示す情報を含むことを特徴とする付記 8 乃至 1 2 のいずれか 1 つに記載の情報処理システム。
- [0202] (付記 1 4) 前記アドレスは、前記記憶装置上の物理アドレスであることを特徴とする付記 8 乃至 1 3 のいずれか 1 つに記載の情報処理システム。
- [0203] (付記 1 5) コンピュータが、アクセス対象のデータを特定するアクセス識別子に対応する第 1 のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第 2 のデータに含まれる管理情報に基づいて、前記第 2 のデータに含まれる前記第 1 のデータを取得し、前記第 2 のデータに対して前記第 1 のデータの操作を実行するデータアクセス方法。
- [0204] (付記 1 6) アクセス対象のデータを特定するアクセス識別子に対応する第 1 のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第 2 のデータに含まれる管理情報に基づいて、前記第 2 のデータに含まれる前記第 1 のデータを取得し、前記第 2 のデータに対して前記第 1 のデータの操作を実行する処理をコンピュータに実行させるプログラム。
- [0205] (付記 1 7) プロセッサと、プロセッサがアドレス解決手段及びアクセス実行手段として動作するための、プロセッサによって実行される命令を保持

する記憶部とを含み、

前記アドレス解決手段は、アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、前記アクセス実行手段は、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行する情報処理装置。

[0206] (付記18) アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行する処理をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な非一時的記録媒体。

[0207] 以上、各実施形態を参照して本発明を説明したが、本願発明は上記実施形態に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解しえるさまざまな変更をすることができる。

[0208] この出願は、2014年2月5日に提出された日本出願特願2014-020317を基礎とする優先権を主張し、その開示の全てをここに取り込む。

産業上の利用可能性

[0209] 本発明は、ネットワークで接続された記憶装置を利用する、データベースシステムやキーバリューストアシステム、及び複数計算機で共通の記憶装置を共有するシェアード型分散データストアシステムなどに適用できる。

符号の説明

[0210] 10 情報処理システム

- 1 0 0 計算機
- 1 1 0 データストア機能実現部
- 1 1 1 アクセス要求受信部
- 1 1 2 アドレス解決部
- 1 1 3 アクセス実行部
- 1 2 0 アプリケーション
- 2 0 0 外部記憶装置
- 2 1 0 アクセス処理部
- 2 2 0 データ記憶部
- 2 2 1 アドレス空間
- 2 2 3 ページ
- 2 2 5 データレコード
- 2 2 6 管理情報
- 3 0 0 ネットワーク
- 7 0 0 コンピュータ
- 7 0 1 CPU
- 7 0 2 記憶部
- 7 0 3 記憶装置
- 7 0 4 入力部
- 7 0 5 出力部
- 7 0 6 通信部
- 7 0 7 記録媒体

請求の範囲

- [請求項1] アクセス対象のデータを特定するアクセス識別子に対応する第1のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出するアドレス解決手段と、
- 算出された前記アドレスに基づいて読み出される、前記領域のデータである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、前記第2のデータに対して前記第1のデータの操作を実行するアクセス実行手段と、を含む情報処理装置。
- [請求項2] 前記アクセス実行手段は、前記第1のデータの追加、削除及び更新のいずれかであるデータ操作を実行する際に、
- 前記記憶装置から前記第1のデータに対応する前記アクセス識別子に基づいて前記第2のデータを取得し、取得された前記第2のデータに対して前記データ操作を実行し、前記データ操作を実行された前記第2のデータを前記記憶装置に書き込む
- 請求項1記載の情報処理装置。
- [請求項3] 前記アドレス解決手段は、
- 前記アクセス識別子に対応する数値を算出し、
- 算出された前記数値と、前記記憶装置上の利用可能なアドレス空間の開始アドレスと、前記領域のサイズ及び前記アドレス空間のサイズを示す情報と、に基づいて前記アドレスを算出することを特徴とする請求項1または2記載の情報処理装置。
- [請求項4] 前記アドレス解決手段は、前記アクセス識別子のハッシュ値を前記数値として算出する
- ことを特徴とする請求項3記載の情報処理装置。
- [請求項5] 前記管理情報は、前記第2のデータに含まれる前記第1のデータに対応するアクセス識別子と、前記第1のデータの前記領域内における位置を示すポインタと、の対応を示す情報を含む

ことを特徴とする請求項 1 乃至 4 のいずれか 1 項に記載の情報処理装置。

[請求項6] 前記第 1 のデータに対応するアクセス識別子に基づいて算出された前記アドレスに基づいて読み出された領域に含まれる前記管理情報は、前記第 1 のデータが前記記憶装置上の他の領域に記憶されることを示す情報、及び前記第 1 のデータが他の記憶装置上の領域に記憶されることを示す情報を含む

ことを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の情報処理装置。

[請求項7] 前記アドレスは、前記記憶装置上の物理アドレスであることを特徴とする請求項 1 乃至 6 のいずれか 1 項に記載の情報処理装置。

[請求項8] 請求項 1 乃至 7 のいずれか 1 項に記載の情報処理装置と、前記記憶装置と、を含む情報処理システム。

[請求項9] コンピュータが、アクセス対象のデータを特定するアクセス識別子に対応する第 1 のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、

算出された前記アドレスに基づいて読み出される、前記領域のデータである、第 2 のデータに含まれる管理情報に基づいて、前記第 2 のデータに含まれる前記第 1 のデータを取得し、

前記第 2 のデータに対して前記第 1 のデータの操作を実行するデータアクセス方法。

[請求項10] アクセス対象のデータを特定するアクセス識別子に対応する第 1 のデータを記憶する、記憶装置上の、領域を特定するアドレスを前記アクセス識別子に基づいて算出し、

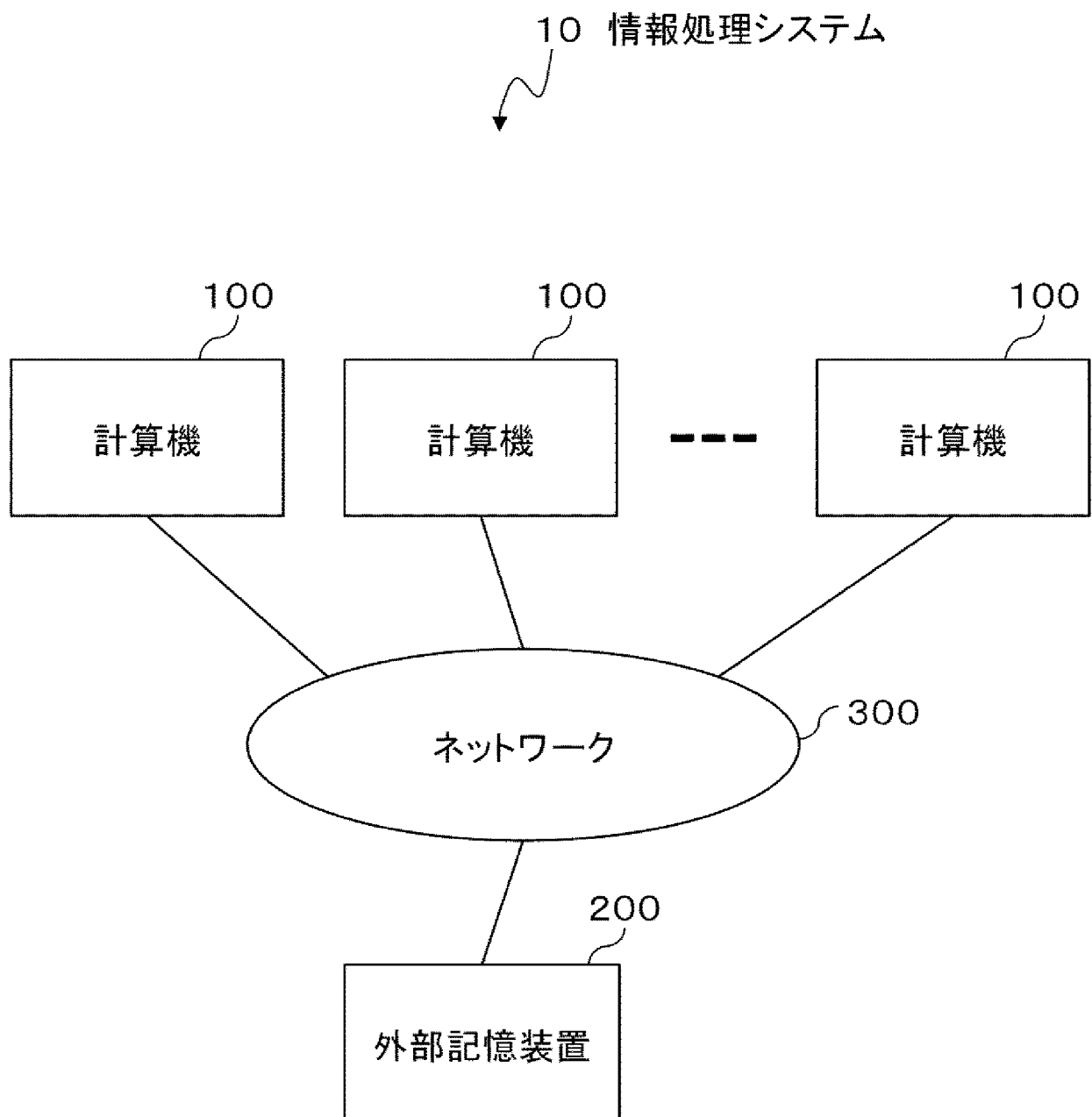
算出された前記アドレスに基づいて読み出される、前記領域のデー

タである、第2のデータに含まれる管理情報に基づいて、前記第2のデータに含まれる前記第1のデータを取得し、

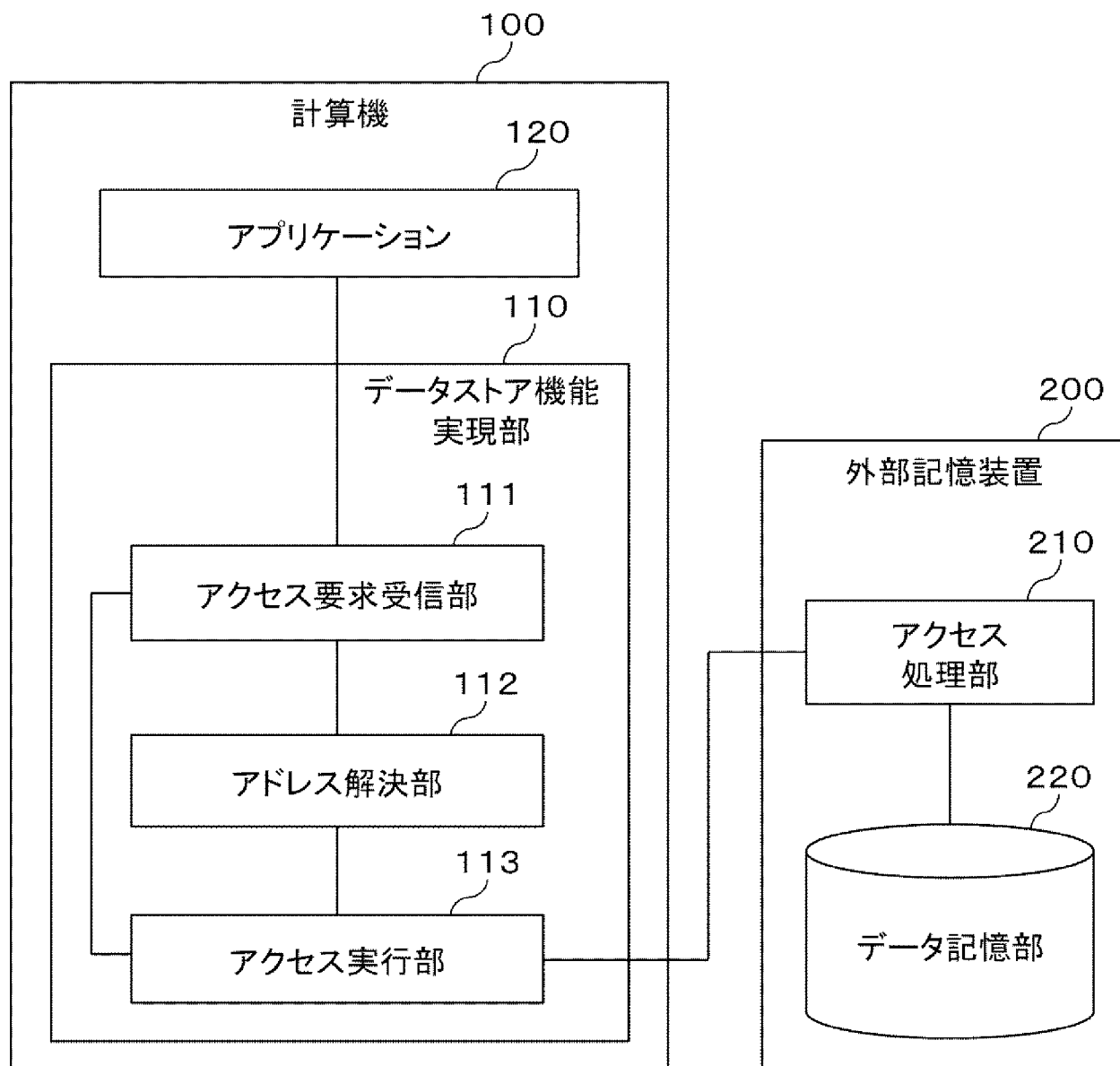
前記第2のデータに対して前記第1のデータの操作を実行する処理をコンピュータに実行させる

プログラムを記録したコンピュータ読み取り可能な非一時的記録媒体。

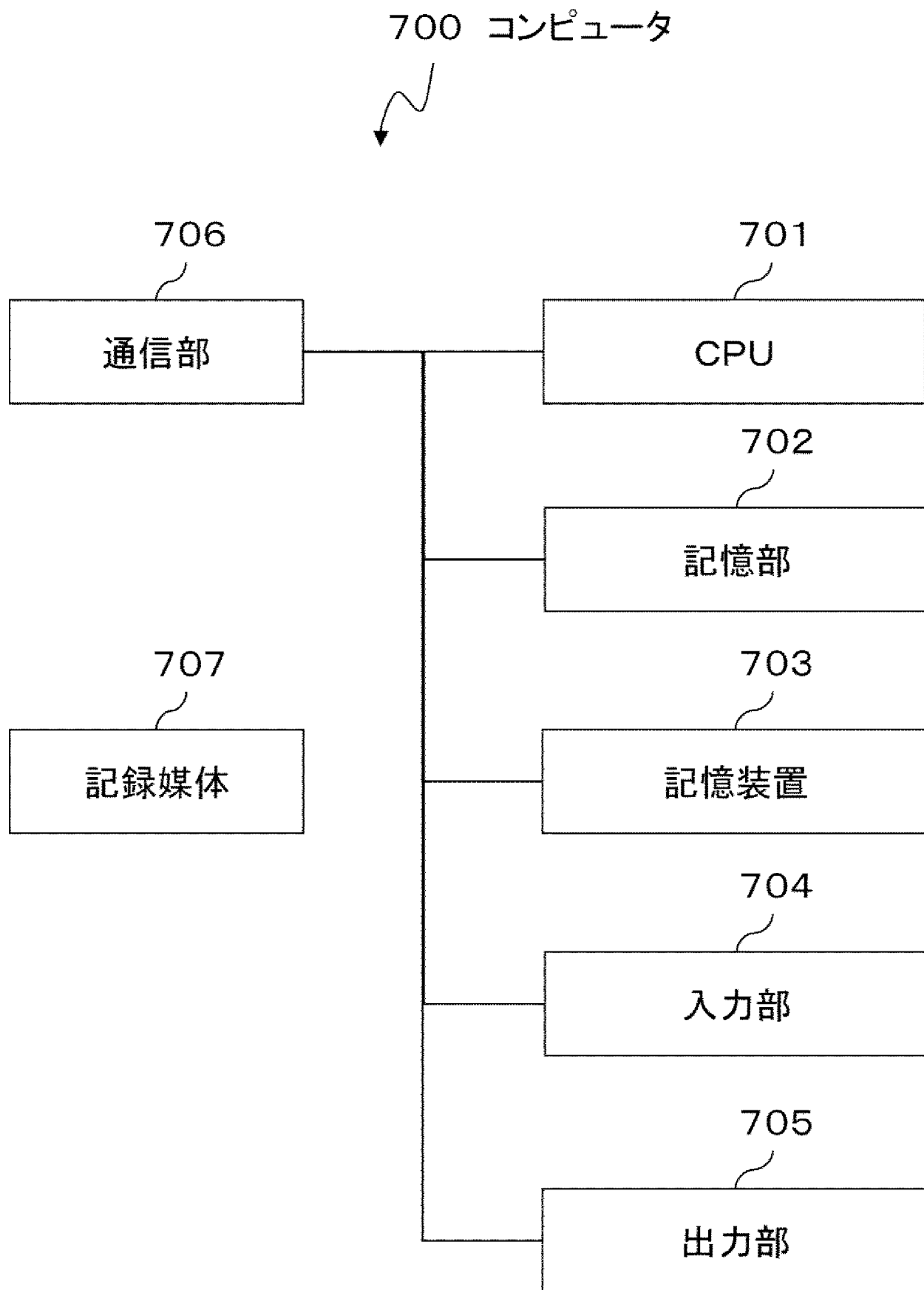
[図1]



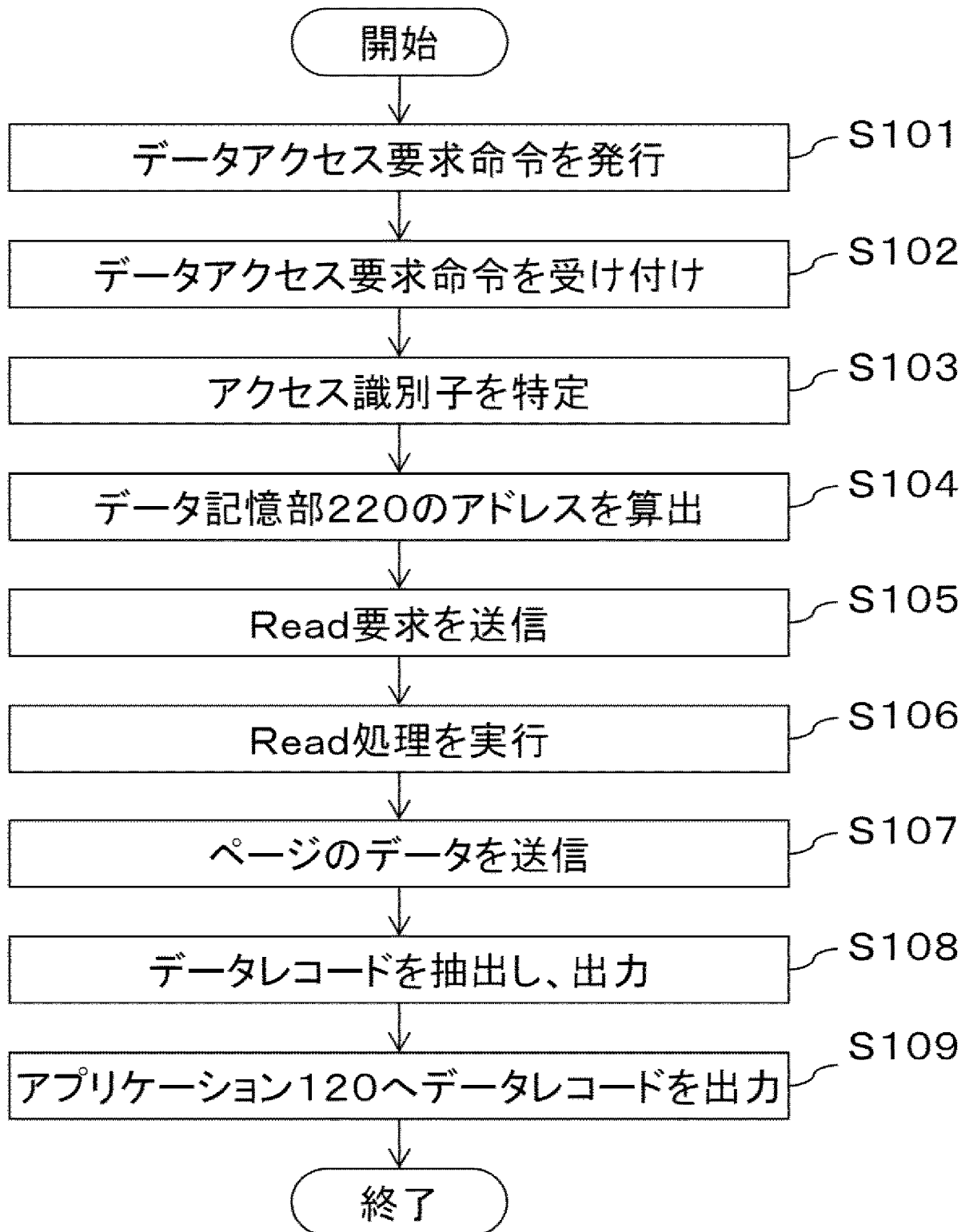
[図2]



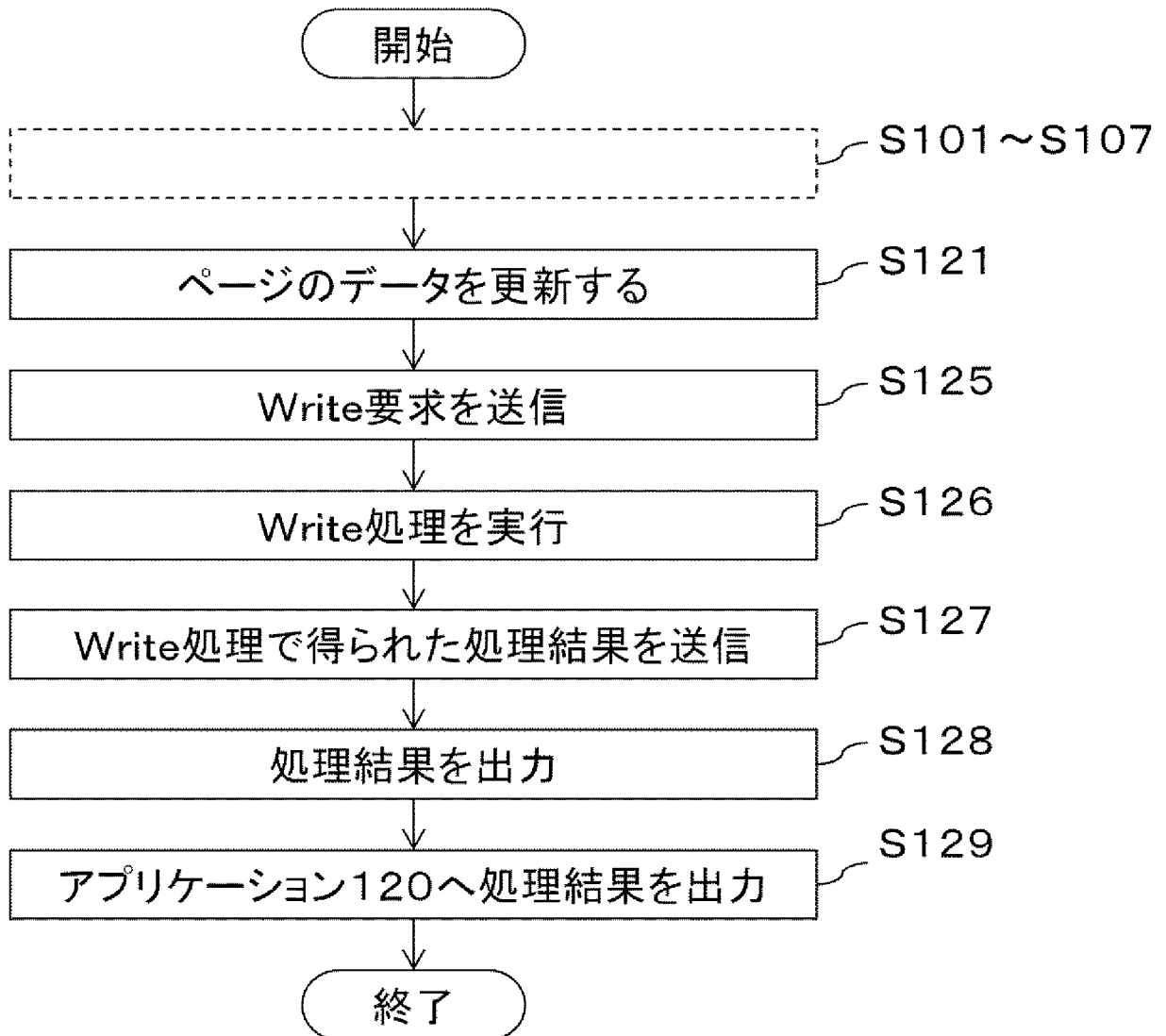
[図3]



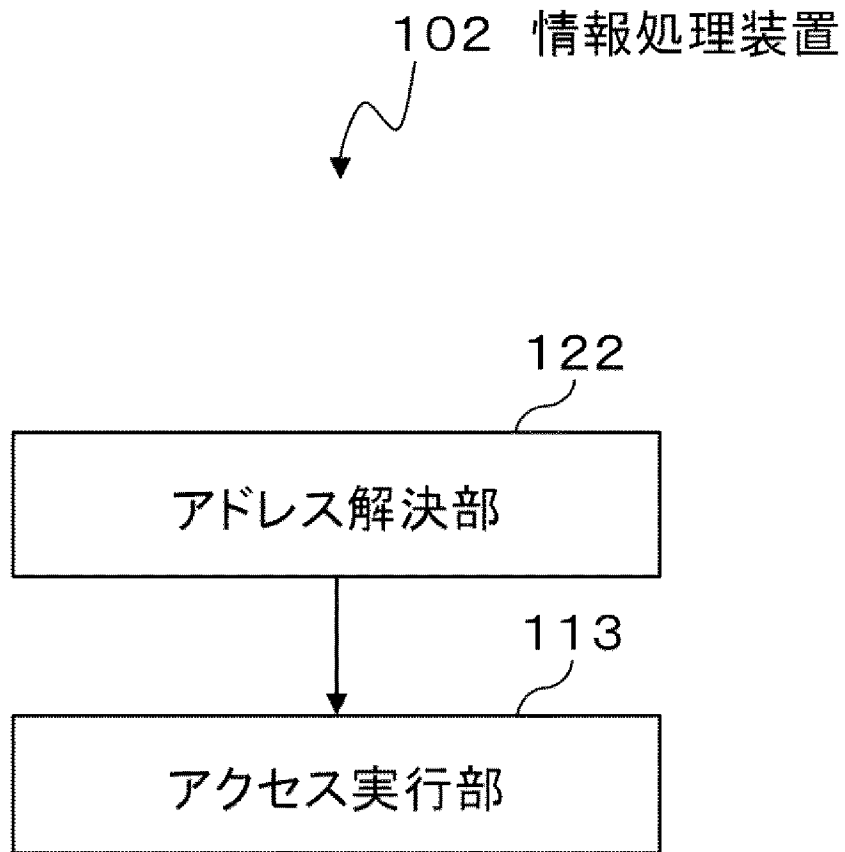
[図4]



[図7]



[図8]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2015/000504

A. CLASSIFICATION OF SUBJECT MATTER
G06F12/00(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F12/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 2012/121316 A1 (NEC Corp.), 13 September 2012 (13.09.2012), paragraphs [0245] to [0259] & US 2013/0346365 A1	1, 5, 6, 8-10 2-4, 7
Y	JP 2007-4234 A (Hitachi, Ltd.), 11 January 2007 (11.01.2007), paragraphs [0057] to [0140] & US 2006/0288153 A1 & EP 1739565 A1	2, 7
Y	JP 2010-128670 A (Hitachi, Ltd.), 10 June 2010 (10.06.2010), paragraph [0043] (Family: none)	3, 4

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 03 April 2015 (03.04.15)	Date of mailing of the international search report 21 April 2015 (21.04.15)
---	--

Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.
--	---

A. 発明の属する分野の分類（国際特許分類（IPC）） Int.Cl. G06F12/00(2006.01)i		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） Int.Cl. G06F12/00		
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2015年 日本国実用新案登録公報 1996-2015年 日本国登録実用新案公報 1994-2015年		
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X Y	WO 2012/121316 A1（日本電気株式会社）2012.09.13, 第245-259段落 & US 2013/0346365 A1	1, 5, 6, 8-10 2-4, 7
Y	JP 2007-4234 A（株式会社日立製作所）2007.01.11, 第57-140段落 & US 2006/0288153 A1 & EP 1739565 A1	2, 7
Y	JP 2010-128670 A（株式会社日立製作所）2010.06.10, 第43段落（ファミリーなし）	3, 4
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー 「A」特に関連のある文献ではなく、一般的技術水準を示すもの 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） 「O」口頭による開示、使用、展示等に言及する文献 「P」国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」同一パテントファミリー文献		
国際調査を完了した日 03.04.2015	国際調査報告の発送日 21.04.2015	
国際調査機関の名称及びあて先 日本国特許庁（ISA/J P） 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員） 加内 慎也 電話番号 03-3581-1101 内線 3565	5U 9745