

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5000234号
(P5000234)

(45) 発行日 平成24年8月15日(2012.8.15)

(24) 登録日 平成24年5月25日(2012.5.25)

(51) Int.Cl.

F I

G O 6 F 12/00 (2006.01)

G O 6 F 3/06 (2006.01)

G O 6 F 13/10 (2006.01)

G O 6 F 12/00 5 O 1 B

G O 6 F 12/00 5 1 4 E

G O 6 F 12/00 5 4 5 A

G O 6 F 3/06 3 O 1 K

G O 6 F 3/06 3 O 2 E

請求項の数 12 (全 52 頁) 最終頁に続く

(21) 出願番号 特願2006-227577 (P2006-227577)

(22) 出願日 平成18年8月24日(2006.8.24)

(62) 分割の表示 特願2004-51935 (P2004-51935)
の分割

原出願日 平成16年2月26日(2004.2.26)

(65) 公開番号 特開2006-331458 (P2006-331458A)

(43) 公開日 平成18年12月7日(2006.12.7)

審査請求日 平成18年8月24日(2006.8.24)

審判番号 不服2010-1008 (P2010-1008/J1)

審判請求日 平成22年1月18日(2010.1.18)

(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 110000062

特許業務法人第一国際特許事務所

(72) 発明者 江口 賢哲

神奈川県川崎市麻生区王禅寺1099番地

株式会社日立製作所システム開発研究所
内

(72) 発明者 山本 康友

神奈川県川崎市麻生区王禅寺1099番地

株式会社日立製作所システム開発研究所
内

最終頁に続く

(54) 【発明の名称】 制御装置

(57) 【特許請求の範囲】

【請求項1】

計算機と記憶装置とに接続される制御装置であって、
 前記計算機と接続される第一のポートと、
 前記記憶装置と接続される第二のポートと、
 前記第一のポート及び前記第二のポートと接続される制御部と、を有し、
 前記制御部は、
 前記記憶装置に含まれる第一の論理デバイスを該制御装置内に存在する論理ユニットとして前記計算機に提供し、

前記第一のポートで受信された、前記計算機から送信される前記論理ユニットに対する
 第一のアクセス要求を、前記論理ユニットと前記第一の論理デバイスとの対応関係に基づ
 いて、前記第一の論理デバイスに対する第二のアクセス要求に変換し、

前記第二のアクセス要求を、前記第二のポートを介して前記記憶装置へ送信し、

前記論理ユニットの入出力性能を測定するために前記第二のアクセス要求に応答する前
記第一の論理デバイスのI/O処理性能情報を計算する制御装置において、

前記制御部は、前記論理ユニットの性能を測定するために、ダミーデータによる性能情
 報解析のための入出力リクエストを生成し、前記第二のポートを介して前記第一の論理デ
 バイスに送信し、前記ダミーデータによる入出力リクエストが送受信される前記制御部と
前記記憶装置との間のI/O処理性能を測定し、

前記制御部は、目標入出力リクエスト処理性能値を決めて、データサイズの異なる複数

10

20

のダミーデータを生成し、前記第一の論理デバイスに送信し、

前記制御部は、前記ダミーデータを生成するにあたり、ダミーデータ長、目標入出力リクエスト処理性能値、入出力対象のアドレスの偏り、リード処理かライト処理かの処理命令種別を指定した計測指示に従って、前記ダミーデータを生成し、前記第一の論理デバイスに送信することを特徴とする制御装置。

【請求項 2】

前記入出力リクエストが送受信される前記制御部と前記記憶装置との間の I / O 処理性能の性能測定結果には、送出した前記ダミーデータのサイズごとに、目標入出力リクエスト処理性能値、アクセス範囲を示すローカリティ、リードライト比率、実測入出力リクエスト処理性能値、実測データ転送量、またはレスポンス時間のうちのいずれか 1 つを含む

10

【請求項 3】

前記制御部は、前記記憶装置内の前記第一の論理デバイスの入出力に関するハードウェア構成情報を取得するための要求を前記第二のポートを介して前記記憶装置に送信することを特徴とする請求項 1 記載の制御装置。

【請求項 4】

前記第二のアクセス要求に応答する I / O 処理性能情報には、前記記憶装置と前記第二のポートとの間における前記第二のアクセス要求に応答するレスポンス時間についての情報が含まれることを特徴とする請求項 1 記載の制御装置。

【請求項 5】

20

更に第二の記憶装置と接続される第三のポートを有し、

前記応答時間が所定の値を超えている場合、前記制御装置は、前記第二のポートを介して前記第一の論理デバイスに格納されているデータを前記記憶装置から受信して前記第二の記憶装置内の第二の論理デバイスに前記第三のポートを介して送信し、

前記第一の論理デバイスと前記論理ユニットとの対応関係を前記第二の論理デバイスと前記論理ユニットとの第二の対応関係に変更し、

前記第二の対応関係に基づいて、前記第一のポートで受信され、前記計算機から送信される前記論理ユニットに対する第一のアクセス要求を前記第二の論理デバイスに対する第三のアクセス要求に変換し、

前記第三のアクセス要求を、前記第三のポートを介して前記第二の記憶装置へ送信することを特徴とする請求項 4 記載の制御装置。

30

【請求項 6】

前記制御部は、前記論理ユニットの性能を測定するために、ダミー入出力リクエストを生成し、前記第三のポートを介して前記第二の論理デバイスに前記ダミーリクエストを送信し、前記ダミー入出力リクエストが送受信される前記制御部と前記記憶装置との間の I / O 処理性能を測定することを特徴とする請求項 5 記載の制御装置。

【請求項 7】

前記第二のアクセス要求に応答する I / O 処理性能情報には、前記記憶装置と前記第二のポートとの間における前記第二のアクセス要求のスループットについての情報が含まれることを特徴とする請求項 1 記載の制御装置。

40

【請求項 8】

更に第二の記憶装置と接続される第三のポートを有し、

前記スループットが所定の値以下の場合、該制御装置は、前記第二のポートを介して前記第一の論理デバイスに格納されているデータを前記記憶装置から受信して前記第二の記憶装置内の第二の論理デバイスに前記第三のポートを介して送信し、

前記第一の論理デバイスと前記論理ユニットとの対応関係を前記第二の論理デバイスと前記論理ユニットとの第二の対応関係に変更し、

前記第二の対応関係に基づいて、前記第一のポートで受信され、前記計算機から送信される前記論理ユニットに対する第一のアクセス要求を前記第二の論理デバイスに対する第三のアクセス要求に変換し、

50

前記第三のアクセス要求を、前記第三のポートを介して前記第二の記憶装置へ送信することを特徴とする請求項 7 記載の制御装置。

【請求項 9】

前記制御部は、前記論理ユニットの性能を測定するために、ダミー入出力リクエストを生成し、前記第三のポートを介して前記ダミー入出力リクエストを前記第二の論理デバイスに送信し、前記ダミー入出力リクエストが送受信される前記制御部と前記記憶装置との間の I / O 処理性能を測定することを特徴とする請求項 8 記載の制御装置。

【請求項 10】

前記第二のアクセス要求に応答する I / O 処理性能情報には、前記記憶装置と前記第二のポートとの間における前記第二のアクセス要求の入出力数についての情報が含まれることを特徴とする請求項 1 記載の制御装置。

10

【請求項 11】

更に第二の記憶装置と接続される第三のポートを有し、

前記入出力数が所定の値以下の場合、該制御装置は、前記第二のポートを介して前記第一の論理デバイスに格納されているデータを前記記憶装置から受信して前記第二の記憶装置内の第二の論理デバイスに前記第三のポートを介して送信し、

前記第一の論理デバイスと前記論理ユニットとの対応関係を前記第二の論理デバイスと前記論理ユニットとの第二の対応関係に変更し、

前記第二の対応関係に基づいて、前記第一のポートで受信された、前記計算機から送信される前記論理ユニットに対する第一のアクセス要求を前記第二の論理デバイスに対する第三のアクセス要求に変換し、

20

前記第三のアクセス要求を、前記第三のポートを介して前記第二の記憶装置へ送信することを特徴とする請求項 10 記載の制御装置。

【請求項 12】

前記制御部は、前記論理ユニットの性能を測定するために、ダミー入出力リクエストを生成し、前記第三のポートを介して前記ダミー入出力リクエストを前記第二の論理デバイスに送信し、前記ダミー入出力リクエストが送受信される前記制御部と前記記憶装置との間の I / O 処理性能を測定することを特徴とする請求項 11 記載の制御装置。

【発明の詳細な説明】

【技術分野】

30

【0001】

本発明は、ホストコンピュータに直接接続されないストレージサブシステムを有するストレージサブシステム全体の性能チューニング技術に関する。

【背景技術】

【0002】

近年のインターネットの普及、ブロードバンド化への対応などにより、計算機システムで扱われる情報量は年々増加し、その重要性も高まるばかりである。これに伴い、計算機システムにおいて、ホスト計算機が読み書きを行う情報を蓄積する記憶装置、特にホスト計算機の外部に接続する記憶装置サブシステムは、大容量化と高性能化と同時に、格納されているデータの保護といった信頼性への要求も高まっている。

40

【0003】

記憶装置サブシステムにおいて、これらの要求を同時に実現する手法の 1 つにディスクアレイシステムがある。

【0004】

ディスクアレイシステムでは、データを、アレイ状に配置された複数の物理記憶装置に分散し、さらに、冗長性を持たせて格納する。すなわち、大容量は、物理記憶装置を複数備えることによって、高い性能は、各物理記憶装置を並列に動作させることによって、そして、高い信頼性は、データの冗長化によって実現している。

【0005】

ディスクアレイシステムは、一般的に、その冗長性を持たせる構成によって、レベル 1

50

からレベル5の種別に分けられている（例えば、非特許文献1参照。）。冗長性を持たせず、単に、複数の物理記憶装置にデータを分割して格納する構成のディスクアレイシステムもあり、このようなものをレベル0と呼ぶこともある。以下、本明細書では、特定の上記レベルを実現する複数の物理記憶装置の組をパリティグループと呼ぶ。また、冗長性を持たせる構成をRAID構成と呼ぶ。

【0006】

ディスクアレイシステムを構成するためのコスト、ディスクアレイシステムの性能や特性等は、ディスクアレイシステムのレベルによって異なる。そこで、システムの使用目的に応じて、ディスクアレイシステムの構築の際にレベルの異なる複数のアレイ（ディスク装置の組）を混在させることも多い。

10

【0007】

このように、ディスクアレイシステムでは、複数の物理記憶装置を並列に動作させることによって性能を高めているため、処理内容に応じて、データを複数のパリティグループに効率よく分散すること、すなわち、性能チューニングが必要である。

【0008】

パリティグループを構成するそれぞれの物理記憶装置は、性能や容量などによりコストが異なる。このため、最適なコストパフォーマンスを実現するために、パリティグループ毎に異なる性能や容量の異なる物理記憶装置を組み合わせてパリティグループを構築することがある。このようなパリティグループによって異なる物理記憶装置が用いられるディスクアレイシステムの場合は、なおさら、性能チューニングは重要である。

20

【0009】

ディスクアレイシステムでの性能チューニングを実現する技術として、例えば、ディスクアレイシステムが、格納されているデータのホスト計算機からのアクセス頻度を監視し、監視結果に従って、アクセス頻度の高いデータをより高速な物理記憶装置に配置する技術がある。（例えば、特許文献1参照。）。

【0010】

また、計算機システムで行われる処理と処理に伴うI/Oが、ユーザによって作成されたスケジュールに則って行われ、日毎、月毎、年毎などの周期性を示す場合が多いという傾向を利用し、ディスクアレイシステムが、各物理記憶装置の使用状況を蓄積し、予め定められた処理スケジュールを考慮して、データの再配置を実行する技術などもある（例えば、特許文献2参照。）。

30

【0011】

ディスクアレイシステムでは、上述のようにデータを各物理記憶装置に分散し、また冗長性を持たせて配置する。接続するホスト計算機が、実際の物理記憶装置のデータの格納場所を意識しなくてもよいように、ホスト計算機が物理記憶装置にアクセスする際に用いる論理的なアドレス（以下、論理アドレスと呼ぶ）を、実際の物理記憶装置の物理的なアドレス（以下、物理アドレスと呼ぶ）と別個に持ち、論理アドレスと物理アドレスとの対応関係を示す情報を保持する構成となっている。

【0012】

このため、上記の技術では、データの再配置に伴い、ディスクアレイシステム内において、論理アドレスと物理アドレスとの対応付けを再配置前のものから再配置後のものに変更している。これにより、ホスト計算機からは、データの再配置後も同じ論理アドレスでアクセスが可能となる。このように、物理装置内でのデータ移動により、ホスト計算機側のその後のアクセスに影響がでないような移動を、ホスト透過的移動と呼ぶ。

40

【0013】

一方、増大する情報量に対応し、ホスト計算機からアクセス可能な記憶装置数を増やすため、アクセスインタフェースの不整合等でホスト計算機に直接入出力処理ができないディスクアレイシステムの記憶装置に、ホスト計算機からのアクセスを可能にする技術がある（例えば、特許文献3参照。）。

【0014】

50

特許文献3に開示されている技術においては、ホスト計算機が直接入出力処理できないディスクアレイシステムに対して、ホスト計算機が直接入出力可能なディスクアレイシステムが、ホスト計算機からのI/O要求などを送信する。

【0015】

【非特許文献1】D. A. Patterson, G. Gibson, and R. H. Kats, "A Case for Redundant Arrays of Inexpensive Disks (ディスクアレイ)" (in Proc. ACM SIGMOD, pp. 109 - 116, June 1988)

【0016】

【特許文献1】特開2000-293317号公報

【特許文献2】特開2001-67187号公報

【特許文献3】特開平10-283272号公報

【発明の開示】

【発明が解決しようとする課題】

【0017】

引用文献3に開示された技術を用い、ホスト計算機が直接入出力処理できないディスクアレイシステム(外部システム)にまで、ホスト計算機で使用するデータの格納領域を広げることが可能である。

【0018】

しかしながら、外部システムを付加した場合、ホスト計算機が直接入出力処理可能なディスクアレイシステムから、外部システムの使用状況、負荷状況などをモニタする機能やデータの再配置機能などがいないため、現状では、そのモニタ結果を用いて、外部システムまで含めて性能チューニングを行なうことができない。

【0019】

そこで、上記事情を鑑み、本発明は、ホスト計算機が入出力処理対象としない(すなわち直接接続しない)ストレージサブシステム(以下、外部ストレージサブシステムと呼ぶ。)を複数接続し、前記外部ストレージサブシステムに対するホスト計算機からのI/O処理要求を中継する機能を有するストレージサブシステムにおいて、外部ストレージサブシステムまで含めた性能チューニングを可能とすることを目的とする。

【課題を解決するための手段】

【0020】

上記目的を達成するために、本発明のストレージサブシステムは、自身に接続する外部ストレージサブシステムの稼動状況をモニタし、それに従って性能チューニングを行う。

【0021】

具体的には、本発明の制御装置は、計算機と記憶装置とに接続される制御装置であって、前記計算機と接続される第一のポートと、前記記憶装置と接続される第二のポートと、前記第一のポート及び前記第二のポートと接続される制御部と、を有し、前記制御部は、前記記憶装置に含まれる第一の論理デバイスを該制御装置内に存在する論理ユニットとして前記計算機に提供し、前記第一のポートで受信された、前記計算機から送信される前記論理ユニットに対する第一のアクセス要求を、前記論理ユニットと前記第一の論理デバイスとの対応関係に基づいて、前記第一の論理デバイスに対する第二のアクセス要求に変換し、前記第二のアクセス要求を、前記第二のポートを介して前記記憶装置へ送信し、前記論理ユニットの入出力性能を測定するために前記第二のアクセス要求に応答する前記第一の論理デバイスのI/O処理性能情報を計算する制御装置において、前記制御部は、前記論理ユニットの性能を測定するために、ダミーデータによる性能情報解析のための入出力リクエストを生成し、前記第二のポートを介して前記第一の論理デバイスに送信し、前記ダミーデータによる入出力リクエストが送受信される前記制御部と前記記憶装置との間のI/O処理性能を測定し、前記制御部は、目標入出力リクエスト処理性能値を決めて、データサイズの異なる複数のダミーデータを生成し、前記第一の論理デバイスに送信し、前記制御部は、前記ダミーデータを生成するにあたり、ダミーデータ長、目標入出力リクエ

10

20

30

40

50

スト処理性能値、入出力対象のアドレスの偏り、リード処理かライト処理かの処理命令種別を指定した計測指示に従って、前記ダミーデータを生成し、前記第一の論理デバイスに送信する。

【発明の効果】

【0022】

本発明によれば、ホスト計算機が入出力処理対象としない外部ストレージサブシステムを複数接続するストレージサブシステムにおいて、接続する外部ストレージサブシステムも含めたストレージサブシステム全体の性能チューニングを行うことが可能となる。

【発明を実施するための最良の形態】

【0023】

以下、本発明の実施の形態を、図面を参照しながら説明する。なお、これにより本発明が限定されるものではない

【0024】

<< 第一の実施形態 >>

『全体の構成』

図1は本発明の第一の実施形態における、計算機システムの構成の一例を説明するための図である。

【0025】

本図に示すように、計算機システムは少なくとも1つのホスト計算機1a、1b、・・・1n（ホスト計算機の台数は問わない。以降、ホスト計算機1で代表する。）と、ストレージサブシステム20a、20b、・・・20n（ストレージサブシステムの台数は問わない。以降、ストレージサブシステム20で代表する。）と、ストレージサブシステム20の保守管理等を行うために用いるサブシステム管理装置5と、ホスト計算機1およびストレージサブシステム20のI/O処理に用いられる第一のI/O経路61と、ホスト計算機1、ストレージサブシステム20およびストレージサブシステム管理装置5を接続するネットワーク7と、ホスト計算機1、I/Oネットワークおよびストレージサブシステム20によって構成されるストレージエリアネットワークの構成管理を行うSAN管理端末9と、ストレージサブシステム20に接続されるストレージサブシステム21a、21b、・・・21n（本ストレージサブシステムの台数は問わない。以後、ホスト計算機1が直接入出力処理を行うストレージサブシステム20と区別するため、外部ストレージサブシステムと呼び、外部ストレージサブシステム21で代表する。）と、ストレージサブシステム20と外部ストレージサブシステム21とを接続する第二のI/O経路62とを備える。

【0026】

ホスト計算機1は、パーソナルコンピュータ（PC）、ワークステーション（WS）、メインフレーム（MF）等の計算機である。ホスト計算機1では、その計算機の種類に応じたオペレーティングシステム（以下「OS」）とそのOSで動作可能な様々な業務、用途に対応したアプリケーションプログラム（AP）、たとえばデータベースマネジメントシステム（DBMS）等、が動作する。本実施形態では、簡単のため、ホスト計算機1を2つ記載しているが、ホスト計算機1は幾つあってもよい。

【0027】

ストレージサブシステム20、外部ストレージサブシステム21は、複数の物理記憶装置をアレイ状に配したディスクアレイ構成の記憶装置システムで、ホスト計算機1にデータの入出力対象領域として、論理記憶装置8を提供する。また、本実施形態におけるストレージサブシステム20aは外部ストレージサブシステム21に対して、ホスト計算機1と同様にI/O処理要求を出すことが出来る機能を有する。

【0028】

なお、外部ストレージサブシステム21は、ホスト計算機1が入出力処理に用いる第一のI/O経路61には接続されず、ストレージサブシステム20が入出力処理に用いる第二のI/O経路62に接続される。このため、外部ストレージサブシステム21は、ホス

10

20

30

40

50

ト計算機 1 から直接 I / O 処理要求を受けず、第一の I / O 経路 6 1 を介してホストからの I / O 要求処理を受けたストレージサブシステム 2 0 から、第二の I / O 経路 6 2 を介して I / O 処理を受ける。

【 0 0 2 9 】

サブシステム管理装置 5 a、5 b は夫々、ストレージサブシステム 2 0、外部ストレージサブシステム 2 1 の障害、保守、構成、性能情報等を該ストレージサブシステムから取得、保持し、ストレージサブシステム 2 0 および外部ストレージサブシステム 2 1 の管理のためのユーザインタフェースを提供する。

【 0 0 3 0 】

ここで、本実施形態における管理とは、例えば、障害や性能の監視、構成定義、ストレージサブシステム上で動作するプログラムのインストール等などのことである。サブシステム管理装置 5 a やサブシステム管理装置 5 b は、例えば、ストレージサブシステム 2 0 や外部ストレージサブシステム 2 1 に論理記憶装置 8 を設定する場合、データをバックアップするための記憶領域を設定する場合、または、データを複製する際の記憶領域の対を設定する場合など、ユーザからの指示を受け付け、ストレージサブシステム 2 0 や、外部ストレージサブシステム 2 1 に対して設定指示及び設定情報を送信する。

【 0 0 3 1 】

第一の I / O 経路 6 1 は、ホスト計算機 1 がストレージサブシステム 2 0 に対して、種々のコマンド及びデータの I / O 処理を行うために用いられる。第二の I / O 経路 6 2 は、ストレージサブシステム 2 0 が外部ストレージサブシステム 2 1 に対して、種々のコマンド及びデータの I / O 処理を行うために使用される。

【 0 0 3 2 】

ホスト計算機 1 からストレージサブシステム 2 0 への I / O 処理要求に係るコマンドおよびデータは、第一の I / O 経路 6 1 を介して伝送され、ホスト計算機 1 から外部ストレージサブシステム 2 1 への I / O 処理要求に係るコマンドおよびデータは、第 1 の I / O 経路 6 1 を介してストレージサブシステム 2 0 へ伝送され、そこから第二の I / O 経路 6 2 を介して外部ストレージサブシステム 2 1 へ伝送される。

【 0 0 3 3 】

第一の I / O 経路 6 1 及び第二の I / O 経路 6 2 は、光ケーブルまたは銅線が用いられ、第一の I / O 経路 6 1 及び第二の I / O 経路 6 2 で使用される通信プロトコルには、イーサネット（登録商標）、F D D I、ファイバチャネル（F C）、S C S I、I n f i n i b a n d、T C P / I P、i S C S I、などが用いられる。

【 0 0 3 4 】

ネットワーク 7 は、ストレージサブシステム 2 0、2 1 の障害、保守、構成、性能等に関する管理情報をストレージサブシステム 2 0、2 1 からサブシステム管理装置 5 に送信したり、サブシステム管理装置 5 からの設定情報をストレージサブシステム 2 0、2 1 に送信したり、前記障害、保守、構成、性能等の管理情報をサブシステム管理装置 5 から S A N 管理端末 9 やホスト計算機 1 に送信したりする際に使用される。ネットワーク 7 に使用されるケーブル、通信プロトコルは第一の I / O 経路 6 1 や第二の I / O 経路 6 2 と同一でも異なってもよい。

【 0 0 3 5 】

なお、第二の I / O 経路 6 2 と第一の I / O 経路 6 1 は、ネットワークの処理論理上分けられていればよい。すなわち、第二の I / O 経路 6 2 と第一の I / O 経路 6 1 との伝送路は、物理的に独立していても良いし、また、第一の I / O 経路 6 1 と第二の I / O 経路 6 2 とは、共通の I / O ネットワークスイッチに接続され論理的に伝送路が分離されていても良い。例えば、両経路は、ファイバチャネルにおいて F C スwitch に接続され、F C スwitch におけるゾーニング技術を用いてスitch で論理的に異なるネットワークを構成してもよい。この場合、第二の I / O 経路 6 2 にのみ接続された外部ストレージサブシステム 2 1 のポート配下にパス定義された論理記憶装置 8 は、ホスト計算機 1 からは検出できず、直接 I / O 対象にはできないように構成される。

【 0 0 3 6 】

『ストレージサブシステム 2 0、外部ストレージサブシステム 2 1 の構成』

図 2 は、ストレージサブシステム 2 0 および外部ストレージサブシステム 2 1 のハードウェア構成とそれらの接続状態とを説明するための図である。ここでは、第一の I / O 経路 6 1 及び第二の I / O 経路 6 2 上に I / O ネットワークスイッチ 1 3 0 を用いた場合を例にあげて説明する。

【 0 0 3 7 】

図 1 に示すように、ストレージサブシステム 2 0 は、第一の I / O 経路 6 1 を介してホスト計算機 1 に接続され、外部ストレージサブシステム 2 1 は、第二の I / O 経路 6 2 を介してストレージサブシステム 2 0 に接続される。

10

【 0 0 3 8 】

ストレージサブシステム 2 0 は、少なくとも 1 つのストレージサブシステム制御装置 1 1 2 と、共有メモリ 1 0 7 と、ディスクキャッシュ 1 0 8 と、論理記憶装置 8 を構成する物理記憶装置 1 1 0 と、ストレージサブシステム制御装置 1 1 2 と共有メモリ 1 0 7 とディスクキャッシュ 1 0 8 と記憶装置 1 1 0 とをつなぐ内部ネットワーク 1 0 9 と、を備える。

【 0 0 3 9 】

ストレージサブシステム制御装置 1 1 2 は、少なくとも 1 つの第一の I / O 経路用のポート 1 0 4 a および少なくとも 1 つの第二の I / O 経路用のポート 1 0 4 b を有する I / O アダプタ 1 0 2 と、ネットワーク 7 用のポート 1 0 5 を有するネットワークアダプタ 1 0 3 と、制御プロセッサ 1 0 0 と、ローカルメモリ 1 0 1 と、ディスクアダプタ 1 0 6 と、を備え、I / O アダプタ 1 0 2 と、ネットワークアダプタ 1 0 3 と、制御プロセッサ 1 0 0 と、ローカルメモリ 1 0 1 と、ディスクアダプタ 1 0 6 とは、内部バスで接続する。

20

【 0 0 4 0 】

ポート 1 0 4 a は、I / O ネットワークスイッチ 1 3 0 を介して第一の I / O 経路 6 1 によりホスト計算機 1 に接続され、ホスト計算機 1 からの I / O 処理要求を受信するターゲットポートである。

【 0 0 4 1 】

ポート 1 0 4 b は、I / O ネットワークスイッチ 1 3 0 を介して第二の I / O 経路 6 2 により外部ストレージサブシステム 2 1 に接続され、外部ストレージサブシステム 2 1 に I / O 処理要求を送信するイニシエータポートである。

30

【 0 0 4 2 】

ポート 1 0 5 は、サブシステム管理端末 5 a とネットワーク 7 とを介して接続され、先に述べたように、サブシステム管理端末 5 a からの要求指示や情報を受領し、必要に応じて情報を送信するために使用される。例えばストレージサブシステム 2 0 の構成情報や障害情報、性能情報をサブシステム管理端末 5 a に対して送信する際等に用いられる。

【 0 0 4 3 】

ホスト計算機 1 から外部ストレージサブシステム 2 1 への I / O 処理要求は、第一の I / O 経路 6 1 を介してポート 1 0 4 a において受領され、ポート 1 0 4 b から第二の I / O 経路 6 2 を介して外部ストレージサブシステムに送信される。

40

【 0 0 4 4 】

ここで、ポート 1 0 4 a とポート 1 0 4 b とを物理的に別個のものとせず、一つのポートにイニシエータポートおよびターゲットポート双方の機能を持たせてもよい。

【 0 0 4 5 】

制御プロセッサ 1 0 0 は、ストレージサブシステム 2 0 を制御するためのプログラムを実行する。本実施形態では、制御プロセッサ 1 0 0 は複数備えるものとする。そして、処理するプログラムの制御対象に応じて、状態が設定される。状態は、後述するとおり、ホスト計算機 1 からの I / O を処理するものをターゲットプロセッサ、外部ストレージサブシステム 2 1 との I / O を処理するものをイニシエータプロセッサと設定する。

【 0 0 4 6 】

50

共有メモリ 107 とローカルメモリ 101 にはストレージサブシステム 20 を動作させるために必要なプログラムおよびデータが格納される。

【0047】

ディスクアダプタ 106 は、ストレージサブシステム制御装置 112 と内部ネットワーク 109 とを接続し、物理記憶装置 110 と入出力処理を行うためのインタフェースを提供するためのものである。

【0048】

外部ストレージサブシステム 21 は、基本的にストレージサブシステム 20 と同様の構成を備える。例えば、外部ストレージサブシステム 21 の全体を制御するストレージサブシステム制御装置 120、ポート 122 とストレージサブシステム制御装置 120 の内部バスとをつなぐ I/O アダプタ 123、ネットワーク 7 用のポート 121、論理記憶装置 8 を構成する物理記憶装置 124 などを用意する。

【0049】

また、ポート 122 は、I/O ネットワークスイッチ 130 を介して第二の I/O 経路 62 によりストレージサブシステム 21 の I/O アダプタ 102 にあるポート 104b と接続される。

【0050】

『ストレージサブシステム 20 の機能構成』

次に、ストレージサブシステム 20 の機能について説明する。これらは、上述したように、共有メモリ 107 およびローカルメモリ 101 に格納されているプログラムを制御プロセッサ 100 が実行することにより実現される。また、これらの機能を実現するために用いられるデータ等についても併せて説明する。

【0051】

図 3 は、ストレージサブシステム制御部 112 の機能を説明するための図である。

【0052】

ストレージサブシステム制御部 112 は、I/O ネットワーク処理部 200 と、ネットワーク処理部 201 と、コマンド処理部 202 と、論理記憶装置稼働情報処理部 206 と、物理記憶装置稼働情報処理部 207 と、外部ストレージ領域稼働情報取得処理部 208 と、キャッシュ Hit / Miss 判定処理部 210 と、キャッシュ量管理処理部 212 と、ポート制御処理部 213 と、プロセッサ稼働情報取得処理部 214 と、物理記憶装置 I/O 処理部 215 と、外部ストレージ I/O 処理部 216 と、構成定義処理部 217 と、構成変更計画立案処理部 218 と、構成変更計画実行処理部 219 と、外部ストレージ記憶装置属性情報取得処理部 221 と、マネージャ 223 とを用意する。これらは、プログラムとしてローカルメモリ 101 に格納される。

【0053】

これらの各処理部で処理されたデータ、または、処理に必要なデータは、それぞれ、論理 / 物理対応情報 203、論理記憶装置属性情報 204、物理記憶装置属性情報 205、外部ストレージ稼働情報 209、キャッシュ量カウンタ 211、外部ストレージ記憶装置属性情報 220、スケジュール情報 222、論理記憶装置稼働情報 224、物理記憶装置稼働情報 225、プロセッサ稼働情報 226、構成変更計画立案情報 227、ポート設定情報 228、ポート稼働情報 229 として、ストレージサブシステム制御装置 112 のローカルメモリ 101 または共有メモリ 107 に格納される。

【0054】

また、ローカルメモリ 101 上には、図示しないが、ストレージサブシステムの時刻情報を持つタイマプログラムがある。なお、ストレージサブシステムがストレージサブシステム制御部 112 を複数持つ場合がある。このような場合は、サブシステム管理装置 5 を介して予め 1 の代表のストレージサブシステム制御部 112 を設定する。そして、設定された代表のストレージサブシステム制御部 112 のタイマプログラムが保持する時刻情報を、共通の時刻情報として、共有メモリに格納する。代表のストレージサブシステム制御部 112 以外のストレージサブシステム制御部 112 は、共有メモリ 107 に格納されて

10

20

30

40

50

いる時刻情報を参照する。このような構成により、全ストレージサブシステム制御部 1 1 2 間で共通の時刻情報を得る。

【 0 0 5 5 】

以下、ストレージサブシステム制御部 1 1 2 の、各処理部、保持情報の詳細について説明する。

【 0 0 5 6 】

『 I / O ネットワーク処理部 2 0 0 』

I / O ネットワーク処理部 2 0 0 は、ポート 1 0 4 a、1 0 4 b と I / O アダプタ 1 0 2 とを制御する。ストレージ管理装置 5 を介して受け付けたストレージ管理者からの指示に従い、ポート 1 0 4 a、1 0 4 b を、イニシエータポート、ターゲットポート、もしくは、混在モードの 3 つの設定状態のいずれかの状態に設定する。

10

【 0 0 5 7 】

『 ネットワーク処理部 2 0 1 』

ネットワーク処理部 2 0 1 は、ネットワークポート 1 0 5 とネットワークアダプタ 1 0 3 とを制御する。

【 0 0 5 8 】

『 構成定義処理部 2 1 7 』

構成定義処理部 2 1 7 は、論理記憶装置 8 と物理記憶装置 1 1 0 または 1 2 4 との対応関係を定義し、論理物理対応情報 2 0 3 として共有メモリ 1 0 7 に格納する。

【 0 0 5 9 】

20

一般的に計算機システムでは、ホスト計算機 1 の起動直後に、例えば S C S I の場合、ホスト計算機 1 が I n q u i r y コマンドを送信しデバイスを検出することにより、接続するストレージサブシステムの論理記憶装置の検出を行う。

【 0 0 6 0 】

本実施形態では、同様に、ホスト計算機 1 の起動直後に、ホスト計算機 1 が、ストレージサブシステム 2 0 の、ターゲットポート 1 0 4 a 及び該ターゲットポート 1 0 4 a を用いて入出力処理可能な論理記憶装置 8 の検出を行う。そして、構成定義処理部 2 1 7 は、論理記憶装置 8 と物理記憶装置 1 1 0 との対応関係をユーザの指示により論理記憶装置 8 を定義する際に設定する。論理記憶装置 8 と物理記憶装置 1 1 0 との対応関係は、論理物理対応情報 2 0 3 として共有メモリ 1 0 7 に格納される。

30

【 0 0 6 1 】

また、システムの開始直後、または、管理者からの指示に従って、構成定義処理部 2 1 7 は、所定のコマンドを外部ストレージサブシステム 2 1 に送出し、外部ストレージサブシステム 2 1 の論理記憶装置 8 を、ストレージサブシステム 2 0 の論理記憶装置 8 として定義する。そして、定義情報を論理 / 物理対応情報 2 0 3 として共有メモリ 1 0 7 に格納する。

【 0 0 6 2 】

本実施形態では、構成定義処理部 2 1 7 は、イニシエータポート 1 0 4 b を介して、外部ストレージサブシステム 2 1 の、ターゲットポート 1 2 2 及び該ターゲットポート 1 2 2 を用いて入出力処理可能な論理記憶装置 8 を検出する。ストレージサブシステム管理端末 5 a において、ストレージサブシステム 2 0 の管理者から、検出した論理記憶装置 8 をストレージサブシステム 2 0 の論理記憶装置 8 として設定する旨の指示を受け付け、ストレージサブシステム管理端末 5 a は、受け付けた指示をストレージサブシステム 2 0 に送信する。該指示を受領したストレージサブシステム 2 0 の構成定義処理部 2 1 7 は、検出した論理記憶装置 8 を、ストレージサブシステム 2 0 の論理記憶装置 8 として定義する。

40

【 0 0 6 3 】

ここで、ストレージサブシステム 2 0 が、外部ストレージサブシステム 2 1 の入出力処理可能な論理記憶装置 8 を検出した際、構成定義処理部 2 1 7 が自動的にストレージサブシステム 2 0 の論理記憶装置 8 として定義するよう構成してもよい。

【 0 0 6 4 】

50

なお、論理記憶装置 8 の定義は、ホスト計算機 1 の起動直後に限られない。システム運用中に、構成定義処理部 2 1 7 が、ストレージサブシステム管理者からの指示を受け付け、論理記憶装置 8 の定義を行ってもよい。

【 0 0 6 5 】

『論理物理対応情報 2 0 3 』

次に、構成定義処理部 2 1 7 が、共有メモリ 1 0 7 に格納する論理物理対応情報 2 0 3 を説明する。この論理物理対応情報 2 0 3 は、上述のように構成定義処理部 2 1 7 によって生成および更新される。また、後述するように、性能チューニング処理において最適化が行われ、論理記憶装置 8 と、物理記憶装置 1 1 0 および 1 2 4 との対応関係に変更があった場合は、構成変更実行処理部 2 1 9 によっても、更新される。

10

【 0 0 6 6 】

論理物理対応情報 2 0 3 には、ホスト計算機 1 がストレージサブシステム 2 0 の記憶装置 1 1 0 にアクセスするために用いる論理アドレスと、ストレージサブシステム 2 0 および外部ストレージサブシステム 2 1 の記憶装置 1 1 0 および記憶装置 1 2 4 の物理アドレスとの対応関係を示す情報が格納される。

【 0 0 6 7 】

図 4 に、論理物理対応情報 2 0 3 の一例を示す。本図に示すように、論理物理対応情報 2 0 3 は、論理記憶装置 8 のアドレスを格納する論理アドレス格納部 6 0 0 と、実際にデータが格納される物理記憶装置 1 1 0 のアドレスが格納される物理アドレス格納部 6 0 1 とを備える。なお、論理物理対応情報 2 0 3 は、ポート 1 0 4 a 毎に作成される。

20

【 0 0 6 8 】

論理アドレス格納部 6 0 0 は、ストレージサブシステム 2 0 のポート 1 0 4 a にアクセスすることによって検出される論理記憶装置 8 (例えば、SCSIにおけるLU(Logical Unit)、以下ターゲット論理記憶装置と呼ぶ。)の論理記憶装置番号(例えばSCSIであればLU番号)が格納されるターゲット論理記憶装置番号格納部 6 0 2 と、そのターゲット論理記憶装置内のアドレスが格納されるターゲット論理記憶装置アドレス格納部 6 0 3 と、ストレージサブシステム 2 0 全体に内部的に付与される論理記憶装置番号(以下、LDEV(Logical Device)番号と呼ぶ。)が格納されるLDEV格納部 6 0 4 と、そのアドレス(以下、LDEVアドレスと呼ぶ。)が格納されるLDEVアドレス格納部 6 0 5 とを備える。

30

【 0 0 6 9 】

ターゲット論理記憶装置番号は、ホスト計算機 1 からみて、入出力対象となるターゲットポート 1 0 4 a ごとに一意に定まるものであり、ホスト計算機 1 は、このLU番号を用いて、データのリードライトを、ストレージサブシステム 2 0 に対して行う。

【 0 0 7 0 】

ターゲット論理記憶装置は、LDEVをターゲットポートに関連づけることで定義される。LDEVを複数連結して1つのターゲット論理記憶装置を定義してもよい。また、ホスト計算機 1 毎にターゲット論理記憶装置番号に割り当てられているLDEVは異なっても良く、同じであってもよい。

【 0 0 7 1 】

また、物理アドレス格納部 6 0 1 には、論理記憶装置番号格納部 6 0 2 に格納されているターゲット論理記憶装置番号に対応する物理アドレスが格納される。

40

【 0 0 7 2 】

物理アドレス格納部 6 0 1 は、パリティグループの番号が格納されるパリティグループ番号格納部 6 0 6 と、データが格納されるディスク装置の情報が格納されるデータ格納部 6 0 7 と、パリティに関する情報が格納されるパリティ格納部 6 0 8 と、外部ストレージサブシステムに関するデータが格納される外部ストレージ格納部 6 0 9 とを備える。

【 0 0 7 3 】

また、データ格納部 6 0 7 は、物理記憶装置(ディスク装置)の番号を格納するディスク装置格納部 6 1 0 とディスク装置内のアドレスを格納するディスク装置内アドレス格納

50

部 6 1 1 とを備え、パリティ格納部 6 0 8 は、ディスク装置格納部 6 1 2 とディスク装置内のアドレスを格納するディスク装置アドレス格納部 6 1 3 とを備える。外部ストレージ格納部 6 0 9 は、外部ストレージサブシステム内の物理記憶装置へアクセス用のポート番号と物理記憶装置の番号とを格納するポート番号ディスク装置番号格納部 6 1 4 と、ディスク装置内のアドレスを格納する論理装置内アドレス格納部 6 1 5 とを備える。

【 0 0 7 4 】

ディスク装置番号格納部 6 1 2 とディスク装置内アドレス格納部 6 1 3 には、パリティグループのレベルに応じた冗長データが格納されるディスク装置とそのアドレスとが格納される。

【 0 0 7 5 】

パリティグループ番号格納部 6 0 6 とディスク装置番号格納部 6 1 0 とアドレス格納部 6 1 1 とには、L D E V 番号格納部 6 0 4 に格納されている L D E V 番号とその L D E V アドレス格納部 6 0 5 に格納されている L D E V アドレスとで定まる論理記憶装置のデータ格納アドレスに対応する物理アドレスを一意に示す、パリティグループ番号、ディスク装置番号、アドレスが格納される。

【 0 0 7 6 】

本実施形態では、ターゲット論理記憶装置が、ストレージサブシステム 2 0 内の物理記憶装置で構成するパリティグループ内の記憶装置アドレスに対応する場合、パリティグループ番号格納部 6 0 6 とデータ格納部 6 0 7 とパリティ格納部 6 0 8 とに有効な値が格納される。そして、外部ストレージ格納部 6 0 9 には無効値（例えば、図 4 では、- 1 ）が

【 0 0 7 7 】

また、論理記憶装置番号格納部 6 0 2 に示すターゲット論理記憶装置番号に対応する物理アドレス 6 0 1 が、外部ストレージサブシステム 2 1 の論理記憶装置である場合（例えば、論理記憶装置番号 6 0 2 が F 3 のエントリの場合）は、外部ストレージ格納部 6 0 9 に有効な値が格納され、データ格納部 6 0 7 およびパリティ格納部 6 0 8 には、無効な値（例えば、図 4 では - 1 ）が格納される。

【 0 0 7 8 】

『 コマンド処理部 2 0 2 』

次に、ホスト計算機 1 から受信した I / O 処理要求に従って、処理を行うコマンド処理部 2 0 2 について説明する。I / O 処理要求は、リード要求とライト要求とストレージサブシステムの状態情報（例えば、構成情報、障害情報等）の取得要求（例えば、S C S I における I n q u i r y コマンド）とがある。コマンド処理部 2 0 2 は、ホスト計算機 1 から受信した I / O 処理要求から処理対象の論理アドレスを抽出し、論理物理対応情報 2 0 3 を参照して、対応する物理アドレスを読み出し、データの入出力処理を行ったり対応するターゲット論理記憶装置の状態をホストコンピュータ計算機 1 に通知したりする。なお、データの入出力処理（データのリードライト処理）の詳細は後述する。

【 0 0 7 9 】

『 論理記憶装置属性情報 2 0 4 』

論理記憶装置属性情報 2 0 4 には、予めサブシステム管理装置 5 などを介して入力された論理記憶装置 8 の属性情報、例えば、サイズ、エミュレーションタイプ、リザーブ情報、パス定義情報、I / O 対象ホスト計算機 1 の情報（ホストの I / O ポートの識別子、たとえばファイバチャネル（F C ）における W o r l d W i d e N a m e （W W N ）等が保持される。

【 0 0 8 0 】

図 5 に、論理記憶装置属性情報 2 0 4 に保持されているデータの一例を示す。本図に示すように、論理記憶装置属性情報 2 0 4 は、ホスト計算機 1 が入出力処理対象としてアクセス可能なターゲットポート識別子（番号）と該ターゲットポートから入力可能なターゲット論理記憶装置の識別子（番号）および該ターゲット論理記憶装置のアドレスに対応したマッピング先のパリティグループ（物理記憶装置）を示す情報である。

【 0 0 8 1 】

論理記憶装置属性情報 2 0 4 は、ターゲットポート 1 0 4 a の識別子 (I D) を格納する I D 格納部 1 1 0 1 とターゲット論理記憶装置 8 の識別子とを格納するターゲット論理記憶装置番号格納部 1 1 0 2 とを備える。

【 0 0 8 2 】

また、論理記憶装置属性情報 2 0 4 は、図 4 に示す論理物理対応情報 2 0 3 に従って、ターゲット論理記憶装置番号格納部 1 1 0 2 に格納されているターゲット論理記憶装置を構成する L D E V の識別子 (L D E V 番号) を格納する L D E V 番号格納部 1 1 0 3 と、該 L D E V が存在するパリティグループ (P G) の識別子を格納する P G 番号格納部 1 1 0 4 とを備える。

10

【 0 0 8 3 】

さらに、論理記憶装置属性情報 2 0 4 は、それぞれの論理記憶装置 8 が、ストレージサブシステム 2 0 のようにホスト計算機 1 から第一の I / O 経路 6 1 を介して直接アクセス可能なストレージサブシステムであるか、ストレージサブシステム 2 0 および第二の I / O 経路 6 2 を介してアクセスしなければならない外部ストレージサブシステムであるかを示すストレージ種類格納部 1 1 0 5 を備える。

【 0 0 8 4 】

ストレージ種類格納部 1 1 0 5 には、外部ストレージサブシステム 2 1 であることを示す情報が格納される場合は、当該外部ストレージサブシステム 2 1 へのアクセスに必要なアドレス、例えば、ファイバチャネル (F C) における W o r l d W i d e N a m e (W W N) と論理記憶装置番号 (L U N) 等も格納される。

20

【 0 0 8 5 】

その他、論理記憶装置属性情報 2 0 4 は、エミュレーションタイプおよび容量に関する情報が格納されるエミュレーションタイプおよび容量格納部 1 1 0 6 とパス定義情報格納部 1 1 0 7 と論理記憶装置の状態情報を格納する状態情報格納部 1 1 0 8 とを備える。

【 0 0 8 6 】

ここで、エミュレーションタイプとは種々のメインフレーム用それぞれの論理記憶装置かオープン系ホスト計算機アクセス対照用論理記憶装置か双方からアクセス可能な論理記憶装置か等の論理記憶装置のエミュレーション情報であり、容量は論理記憶装置の記憶容量のことである。

30

【 0 0 8 7 】

論理記憶装置の状態とは、ホスト計算機 1 から I / O 処理要求を受けるオンライン状態、ある論理記憶装置のある時点のコピー (スナップショットデータ) の格納先や遠隔バックアップ用やディザスタリカバリ用の遠隔コピー先等のための予約状態であるリザーブ状態、論理記憶装置障害による閉塞状態等のことである。

【 0 0 8 8 】

また、パス定義とは、ターゲットポート番号と論理記憶装置番号との組を指定することで、ホスト計算機 1 から入出力処理対象となる論理記憶装置 8 にアクセスできるように、論理記憶装置 8 に、ストレージサブシステムの I / O ポートに関連付けてその配下の論理装置の番号を定義する処理である。

40

【 0 0 8 9 】

『 パリティグループ (物理記憶装置) 属性情報 2 0 5 』

次に、ストレージサブシステム 2 0 および外部ストレージサブシステム 2 1 内の各パリティグループを構成するハードディスク等の物理記憶装置の属性情報 2 0 5 について説明する。

【 0 0 9 0 】

物理記憶装置の属性情報 2 0 5 は、ストレージサブシステム 2 0 の管理者によって、予め、サブシステム管理装置 5 などを介して設定され、共有メモリ 1 0 7 もしくはローカルメモリ 1 0 1 に格納される。物理記憶装置属性情報 2 0 5 として格納される属性情報は、例えば、物理記憶装置 1 1 0 の使用している型や反応速度や回転数、サステイン性能、回

50

転待ち時間、コマンドオーバーヘッド、物理記憶装置の R A I D 構成情報等である。なお、R A I D 構成情報は、パリティグループの構成低後を行う際に設定する情報なので、その情報を用いても良い。

【 0 0 9 1 】

図 6 に物理記憶装置属性情報 2 0 5 に格納されているデータの一例を示す。本図に示すように、物理記憶装置属性情報 2 0 5 は、その物理記憶装置が属するパリティグループの番号を格納するパリティグループ番号格納部 8 0 1 と、各パリティグループに属するディスクの性能指標情報を格納するディスク性能指標格納部 8 0 2 と、各パリティグループの R A I D 構成を格納する R A I D レベル格納部 8 0 3 と、R A I D 構成の詳細を格納する詳細構成格納部 8 0 4 と、各パリティグループの運用上の属性情報を格納するシーケンシャルボリューム格納部 8 0 5 と、リードライトに関する特性を示す属性情報 8 0 6 を備える。

10

【 0 0 9 2 】

ディスク性能指標格納部 8 0 2 には、各パリティグループで使用されているディスクのコマンドオーバーヘッドとシーク時間と平均待ち時間とメディア転送時間等の情報が保持される。また、シーケンシャルボリューム格納部 8 0 5 に格納されている情報は、後述の性能チューニングを行う際に利用される。例えば、R A I D 構成やディスクストライプサイズ上、シーケンシャルアクセスに適していると考えられるパリティグループをシーケンシャルボリュームとして定義し、サブシステム内のある論理デバイス (L D E V) へのアクセス比率でシーケンシャルが高いものを当該パリティグループに移動する。また、ランダムアクセス比率の高い L D E V のデータを、当該シーケンシャルボリュームに定義したパリティグループに移動させない、もしくは、追い出すようにする。

20

【 0 0 9 3 】

なお、後述する性能チューニングは、構成変更計画立案処理部 2 1 8 によって、性能上必要がある場合に行われる。

【 0 0 9 4 】

R A I D レベル格納部 8 0 3 と詳細構成格納部 8 0 4 とに格納される情報により、例えば、本図の第 1 行目に格納されているパリティグループ番号 0 0 0 0 のパリティグループは、その R A I D 構成が R A I D 5 であり、3 D 1 P (4 台のディスクで R A I D 構成 (R A I D 5) を組み、3 つのディスクにデータを格納し、残りの 1 つのディスクにパリティデータを格納する構成) であることが分かる。

30

【 0 0 9 5 】

属性情報 8 0 6 には、リード処理のみを受け付け、ライト更新処理を拒否するボリュームであるか否かを示す属性情報が格納される。

【 0 0 9 6 】

『論理記憶装置稼動情報取得処理部 2 0 6、論理記憶装置稼動情報 2 2 4』

論理記憶装置稼動情報取得処理部 2 0 6 は、ストレージサブシステム 2 0 内の論理記憶装置 8 に対する I / O 処理に関して入出力に係る稼動情報を取得し、論理記憶装置稼動情報 2 2 4 として共有メモリ 1 0 7 またはローカルメモリ 1 0 1 に保持する。

【 0 0 9 7 】

40

論理記憶装置稼動情報取得処理部 2 0 6 は、論理記憶装置 8 を構成する L D E V ごともしくは論理記憶装置 8 ごとに、ある一定時間を単位 (例えば、1 秒、1 分、1 0 分等) として、例えば、ストレージサブシステム 2 0 の単位時間当たりの該論理記憶装置に対する全 I / O 処理に係る時間をカウントし、単位時間当たりの平均 I / O 処理時間を算出するとともに、該単位時間内の最大 I / O 処理時間を随時求める。また、該単位時間内の該論理記憶装置に対する全 I / O 数や単位時間当たりの平均 I / O カウント数、単位時間内での最大 I / O カウント数、単位時間当たりの平均データ転送量、単位時間当たりの最大データ転送量等を算出する。

【 0 0 9 8 】

また、論理記憶装置稼動情報取得処理部 2 0 6 は、論理記憶装置 8 を構成する L D E V

50

ごともしくは論理記憶装置 8 ごとに、キャッシュ Hit / Miss 判定処理がカウントしたキャッシュ量カウンタ値を元に、単位時間当たりのキャッシュヒット率を算出する。さらに、リードアクセス時間、ライトアクセス時間、シーケンシャルアクセス時間、総占有時間等をモニタし、シーケンシャル比率、リードライト比率、平均ディスク占有率、最大ディスク占有率などを、算出する。

【 0 0 9 9 】

本実施形態では、以上の取得した情報と算出した情報とは、論理記憶装置稼動情報 2 2 4 として共有メモリ 1 0 7 またはローカルメモリ 1 0 1 に保持される。

【 0 1 0 0 】

図 7 および図 8 に論理記憶装置稼動情報取得処理部 2 0 6 が計測し、取得、算出した論理記憶装置 8 の、それぞれ、L U 番号および L D E V 番号ごとの論理記憶装置稼動情報 2 2 4 a、2 2 4 b の一例を示す。

【 0 1 0 1 】

図 7 はポート 1 0 4 a に関連付けられてパス定義された L U の稼動情報を格納する論理記憶装置稼動情報 2 2 4 a に格納されているデータの一例を説明するための図である。論理記憶装置稼動情報 2 2 4 a は、ここで、列 9 0 1 には、取得時間がサンプリング時間間隔ごとに記録されている。列 9 0 2 には、ポート 1 0 4 a ごとに論理記憶装置 8 の番号がサンプリング時間ごと記録され、列 9 0 4 には当該時間の当該論理記憶装置の一秒あたりの平均 I / O カウント数 (I O P S)、列 9 0 5 には当該時間の一秒あたりの最大 I / O カウント数 (I O P S)、列 9 0 6 には当該時間の当該論理記憶装置の一秒あたりの平均データ転送量 (M B P S)、列 9 0 7 には当該時間の一秒あたりの最大データ転送量 (M B P S)、列 9 0 8 には当該時間の当該論理記憶装置の一秒あたりの平均 I O 処理時間 (μ s)、列 9 0 9 には当該時間の一秒あたりの最大 I O 処理時間 (μ s) がそれぞれ記録される。

【 0 1 0 2 】

図 8 は、L D E V の稼動情報を格納する論理記憶装置稼動情報テーブル 2 2 4 b に格納されているデータの一例を説明するための図である。本図に示すように、論理記憶装置稼動情報テーブル 2 2 4 b は、シーケンシャルリード 1 0 0 4 と、シーケンシャルライド (データ) 1 0 0 5 と、シーケンシャルライト (パリティ) 1 0 0 6 と、ランダムリード 1 0 0 7 と、ランダムライト (データ) 1 0 0 9 と、総占有時間 1 0 1 0 と、リードライト比率 1 0 1 1 と、シーケンシャル比率 1 0 1 2 とを備える。

【 0 1 0 3 】

ここで、シーケンシャルリード 1 0 0 4 と、シーケンシャルライド (データ) 1 0 0 5 と、シーケンシャルライト (パリティ) 1 0 0 6 と、ランダムリード 1 0 0 7 と、ランダムライト (データ) 1 0 0 9 と、総占有時間 1 0 1 0 とには、それぞれの処理の、所定の時刻ごとの L D E V ごとの、ディスクアクセスの占有時間が格納される。また、リードライト比率 1 0 1 1 には、所定の時刻ごとの L D E V ごとの、ライトアクセス時間に対するリードアクセス時間の割合が格納される。そして、シーケンシャル比率 1 0 1 2 には、所定の時刻ごとの L D E V ごとの、総占有時間に占めるシーケンシャルアクセスの時間の割合が格納される。

【 0 1 0 4 】

『物理記憶装置稼動情報取得処理部 2 0 7、物理記憶装置稼動情報 2 2 5』

物理記憶装置稼動情報取得処理部 2 0 7 は、物理記憶装置 1 1 0 に対する I / O 処理による稼動情報 (物理稼動情報) を取得し、物理記憶装置稼動情報 2 2 5 として保持する。

【 0 1 0 5 】

物理記憶装置稼動情報処理部 2 0 7 は、物理記憶装置 1 1 0 ごとに、ある一定時間を単位として、平均 I / O 処理時間、最大 I / O 処理時間、平均 I / O カウント数、最大 I / O カウント数、平均データ転送量、最大データ転送量、キャッシュヒット率、シーケンシャル比率、リードライト比率、平均ディスク占有率、最大ディスク占有率等の情報を取得し、物理記憶装置稼動情報 2 2 5 として共有メモリ 1 0 7 またはローカルメモリ 1 0 1 に

保持する。このように、物理記憶装置稼動情報 2 2 5 には、ホスト計算機 1 の記憶装置への I/O 処理に必要な時間を計測した結果が記録される。

【 0 1 0 6 】

図 9 は、物理記憶装置稼動情報 2 2 5 の一例を示す図である。

【 0 1 0 7 】

本図に示すように、物理記憶装置稼動情報 2 2 5 は、シーケンシャルリード 1 3 0 4 と、シーケンシャルライト（データ） 1 3 0 5 と、シーケンシャルライト（パリティ） 1 3 0 6 と、ランダムリード 1 3 0 7 と、ランダムライト（データ） 1 3 0 8 と、ランダムライト（パリティ） 1 3 0 9 と、総占有時間 1 3 1 0 と、リードライト比率 1 3 1 1 と、シーケンシャル比率 1 3 1 2 とを備える。

10

【 0 1 0 8 】

ここで、シーケンシャルリード 1 3 0 4 と、シーケンシャルライト（データ） 1 3 0 5 と、シーケンシャルライト（パリティ） 1 3 0 6 と、ランダムリード 1 3 0 7 と、ランダムライト（データ） 1 3 0 8 と、ランダムライト（パリティ） 1 3 0 9 と、総占有時間 1 3 1 0 とには、所定の時刻ごとのパリティグループごとの、それぞれの処理がディスクアクセスを占有する時間が格納される。リードライト比率 1 3 1 1 には、所定の時刻ごとのパリティグループごとの、ライトアクセス時間に対するリードアクセス時間の割合が格納される。そして、シーケンシャル比率 1 3 1 2 には、所定の時刻ごとのパリティグループごとの、総占有時間に占めるシーケンシャルアクセスの時間の割合が格納される。

20

【 0 1 0 9 】

なお、論理記憶装置稼動情報取得処理部 2 0 6 および論理記憶装置稼動情報 2 2 4 と、物理記憶装置稼動情報取得処理部 2 0 7 および物理記憶装置稼動情報 2 2 5 とは、どちらか一方があれば、論理/物理対応情報 2 0 3 に格納されているデータから作成可能であるため、どちらか一方が保持されていれば良い。

【 0 1 1 0 】

『外部ストレージ稼動情報取得処理部 2 0 8、ストレージ稼動情報 2 0 9』

外部ストレージ稼動情報取得処理部 2 0 8 は、外部ストレージの稼動情報を取得して、外部ストレージ稼動情報 2 0 9 として共有メモリ 1 0 7 またはローカルメモリ 1 0 1 に保持する。

30

【 0 1 1 1 】

外部ストレージ稼動情報取得処理部 2 0 8 は、外部ストレージサブシステム 2 1 が提供する論理記憶装置 8 へのストレージサブシステム 2 0 からの I/O 処理要求に対するレスポンスやスループットを稼動情報として計測して取得し、外部ストレージ稼動情報 2 0 9 として保持する。

【 0 1 1 2 】

図 1 0 は外部ストレージ稼動情報 2 0 9 を説明するための図である。

【 0 1 1 3 】

本図に示すように、外部ストレージ稼動情報 2 0 9 は、時間 1 8 0 1 と、イニシエータポート番号 1 9 0 1 と、汎用ストレージサブシステム WWN 1 9 0 2 と、汎用ストレージ LUN 1 9 0 3 と、IOPS 1 9 0 4 と、MBPS 1 9 0 5 と、レスポンス 1 9 0 6 と、リンク状態 1 9 0 7 とを備える。

40

【 0 1 1 4 】

『キャッシュ Hit / Miss 判定処理部 2 1 0、キャッシュ量カウンタ 2 1 1』

キャッシュ Hit / Miss 判定処理部 2 1 0 は、コマンド処理部 2 0 2 に呼び出され、コマンド処理部 2 0 2 において処理するコマンドの対象アドレスのデータがディスクキャッシュ 1 0 8 上にあるか無いかを判定する。図 1 1 は、キャッシュ量カウンタ 2 1 1 を説明するための図である。

【 0 1 1 5 】

キャッシュ Hit / Miss 判定処理部 2 1 0 は、処理するコマンドの対象アドレスのデータが、ディスクキャッシュ 1 0 8 上にあれば、コマンド処理部 2 0 2 に、対象アドレ

50

スのデータがディスクキャッシュ 108 上に存在することと、その存在領域のアドレスとを通知する。そのとき、ディスクキャッシュ 108 上の所定の領域情報と該当する論理記憶装置 8 のキャッシュ量カウンタ 211 の図示しないヒット情報カウンタを一つ増加させる。

【0116】

なお、キャッシュ量カウンタ 211 は、上記のヒット情報の他に、論理記憶装置 8 毎の、ディスクキャッシュ 108 上にあって物理記憶装置 110 に書き込まれていないダーティ状態のデータの量（これをダーティ量と以下する）と、ディスクキャッシュ 108 上と物理記憶装置 110 との双方にあるクリーン状態のデータ量（これをクリーン量と以下する）と、キャッシュ全体のダーティ量、クリーン量、空き容量であるフリー量、および、

10

【0117】

『キャッシュ量管理部 212』

キャッシュ量管理部 212 は、ディスクキャッシュ 108 に格納されているデータ量を管理する。

【0118】

キャッシュ量管理部 212 は、キャッシュのクリーン量またはダーティ量の変動に応じてコマンド処理部 202 の起動間隔を制御する。具体的には、キャッシュ量カウンタ 211 のキャッシュ全体のダーティ量とクリーン量とを示すカウンタを常に参照し、両者の合計がある一定の割合以上になったら、コマンド処理部 202 の起動間隔を通常より長くし、ある一定の割合以下になれば、通常の起動間隔に戻すといった制御を行なう。

20

【0119】

ホスト計算機 1 からのライト処理が頻繁に行なわれると、ディスクキャッシュ 108 上にダーティ状態のデータがたまる。例えば主電源が落ちて、自己の内臓電池でサブシステムを動作させなければならなくなった場合、ディスクキャッシュ 108 上のダーティ状態のデータを所定の物理記憶装置に書き込む必要があるが、ダーティ量が大いいと、書き込みに時間を要して、物理記憶装置上にデータが反映されずに失う危険性が増える。

【0120】

また、RAID 構成をとる場合、パリティを生成するための領域がキャッシュ上に必要なため、ホスト計算機 1 から送信されたライトデータは、キャッシュ容量全体のある一定量以下に抑える必要がある。

30

【0121】

これらの事情から、キャッシュ量管理部 212 は、キャッシュ量を監視し、その量を常に一定以下にするようシステムの動作を制御する。

【0122】

『ポート流量管理処理部 213、ポート設定情報 228、ポート稼動情報 229』

ポート流量管理処理部 213 は、ストレージサブシステム 20 のポート 104 a、104 b ごとのデータ流量を管理する。

【0123】

ポート流量管理処理部 213 は、I/O ポート 104 a、104 b において I/O 処理要求を受領するごとに、ホスト計算機 1 の WWN 毎に、データ転送量、I/O 数を計測し、それらをポート稼動情報 229 として共有メモリ 107 またはローカルメモリ 101 に格納する。

40

【0124】

図 12 は、ポート稼動情報 229 の一例を説明するための図である。本図に示すように、ポート稼動情報 229 は、時間 1201 と、ポート番号 1201 a と、平均レスポンス時間および最大のレスポンス時間とをそれぞれ保持する、平均 IOPS 1204 および最大 IOPS 1205 と、平均のスループット時間と最大のスループット時間とをそれぞれ保持する平均 MBPS 1206 および最大 MBPS 1207 と、平均 I/O 処理時間と最大 I/O 処理時間とをそれぞれ保持する平均 I/O 処理時間 1208 および最大 I/O 処

50

理時間 1 2 0 9 と、シーケンシャル比率 1 2 1 0 とリードライト比率 1 2 1 1 とを備える。

【 0 1 2 5 】

また、ポート流量管理処理部 2 1 3 は、ポート稼動情報 2 2 9 として保持されている情報を用い、ポートでの I / O 処理制御を行う。

【 0 1 2 6 】

なお、ホスト計算機 1 の WWN 毎の、I / O 数 (I O P S)、データ転送量 (M B P S) などの上限値の設定情報は、ストレージサブシステム管理者からストレージサブシステム管理装置 5 を介して予め入力され、ネットワーク 7 を介してストレージサブシステム 2 0 に送信され、ポート設定情報 2 2 8 として共有メモリ 1 0 7 またはローカルメモリ 1 0 1 に格納される。

10

【 0 1 2 7 】

ポート設定情報 2 2 8 は、ポート 1 0 4 a、1 0 4 b ごとの、ホスト計算機 1 との I / O 処理に係る稼動情報、例えばスループット (単位時間当たりの I O 数 (I O P S) やデータ転送量 (M B P S)) やレスポンス (I / O 処理応答時間) の最悪値、平均値、最良値等を基に、上限値、下限値、目標値が格納される。

【 0 1 2 8 】

これらは、例えば、テスト I / O をストレージサブシステム 2 0 から外部ストレージサブシステム 2 1 に出し、その際の応答時間などを見て、ストレージサブシステム管理装置 5 を介して設定される。なお、元々ホスト計算機 1 と I / O 処理を行っていた外部ストレージサブシステム 2 1 であれば、そのときに得られた情報を利用する。

20

【 0 1 2 9 】

図 1 3 は、ポート設定情報 2 2 8 の一例を示す図である。

【 0 1 3 0 】

本図に示すように、ポート設定情報 2 2 8 は、ポート番号 1 4 0 0 と、状態 1 4 0 1 と、接続 WWN 1 4 0 2 と、最大 I O P S 1 4 0 3 と、最大 M B P S 1 4 0 4 と、最小 I O P S 1 4 0 5 と、最小 M B P S 1 4 0 6 と、目標 I O P S 1 4 0 7 と、目標 M B P S 1 4 0 8 と、帯域 1 4 0 9 と、プロトコル 1 4 1 0 とを備える。

【 0 1 3 1 】

『 プロセッサ稼動情報取得処理部 2 1 4、プロセッサ稼動情報 2 2 6 』

30

プロセッサ稼動情報取得処理部 2 1 4 は、制御プロセッサ 1 0 0 が各処理を実行する時間を計測し、プロセッサ稼動情報 2 2 6 として記録する。

【 0 1 3 2 】

図 1 4 にプロセッサ稼動情報 2 2 6 の一例を示す。本図に示すように、プロセッサ稼動情報 2 2 6 は、プロセッサ番号 1 4 5 1 と、状態 1 4 5 2 と、稼働率 1 4 5 3 とを備える。ここで、状態 1 4 5 2 には、当該プロセッサが、ホスト計算機 1 からの I / O を処理するターゲットプロセッサであるか、外部ストレージサブシステム 2 1 との I / O 処理を制御するイニシエータプロセッサであるかを示す情報が格納される。各プロセッサの状態は、予め管理者などによって定められる。

【 0 1 3 3 】

40

『 物理記憶装置 I / O 処理部 2 1 5 』

物理記憶装置 I / O 処理部 2 1 5 は、ホスト計算機 1 の I / O 処理要求に従い、物理記憶装置 1 1 0 からディスクキャッシュ 1 0 8 へ、または、ディスクキャッシュ 1 0 8 から物理記憶装置 1 1 0 に、データを書き込む。

【 0 1 3 4 】

ホスト計算機 1 からライト処理要求があった場合、コマンド処理部 2 0 2 においてその処理要求の相手先を解析する。その結果、その処理要求の相手先がストレージサブシステム 2 0 内の物理記憶装置 1 1 0 であった場合、物理記憶装置 I / O 処理部 2 1 5 は、コマンド処理部 2 0 2 がディスクキャッシュ 1 0 8 にライト処理を行った後、コマンド処理部 2 0 2 からの指示に従い、ダーティデータを、該当する物理記憶装置 1 1 0 に書き込む。

50

以下、物理記憶装置 I / O 処理部 2 1 5 がディスクキャッシュ 1 0 8 から物理記憶装置 1 1 0 にデータを書き込む処理をライトアフタ処理と呼ぶ。

【 0 1 3 5 】

また、ホスト計算機 1 からリード処理要求があり、その相手先が、ディスクキャッシュ 1 0 8 上になく、かつ、ストレージサブシステム 2 0 内の物理記憶装置 1 1 0 であった場合、物理記憶装置 I / O 処理部 2 1 5 は、コマンド処理部 2 0 2 からの指示に従い、当該物理記憶装置 1 1 0 の所定のアドレスに読み出しに行く。

【 0 1 3 6 】

『外部ストレージ I / O 処理部 2 1 6 』

外部ストレージ I / O 処理部 2 1 6 は、ホスト計算機 1 の I / O 処理要求に従い、外部
ストレージサブシステムの物理記憶装置 1 2 4 からディスクキャッシュ 1 0 8 へ、または
、ディスクキャッシュ 1 0 8 から物理記憶装置 1 2 4 に、データを書き込む。

10

【 0 1 3 7 】

コマンド処理部 2 0 2 が解析したライト処理要求の相手先が外部ストレージサブシステム 2 1 内の物理記憶装置 1 2 4 であった場合、コマンド処理部 2 0 2 の指示に従い、外部
ストレージ I / O 処理部 2 1 6 は、ライトアフタ時に、ダーティデータを、ディスクキャ
ッシュ 1 0 8 から該当する物理記憶装置 1 2 4 に書き込む。

【 0 1 3 8 】

また、コマンド処理部 2 0 2 が解析したリード処理要求の相手先が、ディスクキャッ
シュ 1 0 8 上になく、かつ、外部ストレージサブシステム 2 1 内の物理記憶装置 1 2 4 であ
った場合、外部ストレージ I / O 処理部 2 1 6 は、コマンド処理部 2 0 2 の指示に従い、
当該物理記憶装置 1 2 4 の所定のアドレスに読み出しに行く。

20

【 0 1 3 9 】

『構成変更計画立案処理部 2 1 8、構成変更計画立案情報 2 2 7 』

構成変更計画立案処理部 2 1 8 は、物理記憶装置稼動情報 2 2 5、ポート稼動情報 2 2
9、論理記憶装置稼動情報 2 2 4、プロセッサ稼動情報 2 2 6、キャッシュ量カウンタ 2
1 1などを参照して、性能上の要求に応じて構成変更計画を立てる。

【 0 1 4 0 】

構成変更計画立案処理部 2 1 8 は、マネージャ 2 2 3 により、所定の時間間隔で起動さ
れ、ストレージサブシステム 2 0 または外部ストレージサブシステム 2 1 のホスト I / O
処理性能が当初想定より所定以上低下している場合に、データの再配置、イニシエータポ
ート 1 0 4 b 数の増減等を計画し、構成変更計画立案情報 2 2 7 として共有メモリ 1 0 7
またはローカルメモリ 1 0 1 に保持する。構成変更計画が必要な理由は、ストレージサブ
システム 2 0 のホスト I / O 処理性能が、システムの運用経過に従い時間とともに変化する
ためである。

30

【 0 1 4 1 】

計画立案した情報は、後述する構成変更計画実行処理部 2 1 9 によって実行されるか、
または、ストレージ管理者などからキャンセルの指示を受け付けるまで保持される。

【 0 1 4 2 】

なお、I / O 処理性能として、例えば、スループットやレスポンス時間などで判断する
。

40

【 0 1 4 3 】

『構成変更計画実行処理部 2 1 9 』

構成変更計画実行処理部 2 1 9 は、マネージャ 2 2 3 の指示により、構成変更計画立案
情報 2 2 7 に従って構成変更を行う。本処理の実行により、ストレージサブシステム 2 0
および外部ストレージサブシステム 2 1 の性能が改善される。

【 0 1 4 4 】

『スケジュール情報 2 2 2 』

スケジュール情報 2 2 2 は、性能チューニングに係る各種処理を行なうスケジュールを
示す情報であり、共有メモリ 1 0 7 またはローカルメモリ 1 0 1 に格納される。スケジ
ュ

50

ール情報 2 2 2 は、例えば、構成変更計画を立案するタイミング、立案した計画を実行するタイミングなどである。

【 0 1 4 5 】

『 マネージャ 2 2 3 』

マネージャ 2 2 3 はスケジュール情報 2 2 2 に従い、構成変更計画立案処理部 2 1 8 および構成変更計画実行処理部 2 1 9 を実行する。

【 0 1 4 6 】

『 外部ストレージ記憶装置属性情報取得処理部 2 2 1、外部ストレージ記憶装置属性情報 2 2 0 』

外部ストレージ記憶装置属性情報取得処理部 2 2 1 は、サブシステム管理装置 5 から外部ストレージサブシステム 2 1 の記憶装置属性情報を取得し、外部ストレージ記憶装置属性情報 2 2 0 として共有メモリ 1 0 7 またはローカルメモリ 1 0 1 に格納する。なお、本処理部は、ネットワーク 7 を介し外部ストレージサブシステム 2 1 の管理装置 5 から外部ストレージサブシステム 2 1 の R A I D 構成や使用している H D D の種類等が取得できる I / F が用意されている場合に処理が行なわれる。

【 0 1 4 7 】

『 サブシステム管理装置 5 』

図 1 5 に、本実施形態のサブシステム管理装置 5 のハードウェア構成の一例を示す。

【 0 1 4 8 】

本図に示すように、サブシステム管理装置 5 は、それぞれ、伝送路である内部バス 3 0 8 で接続された、制御プロセッサ 3 0 1 と、電氣的に不揮発な記憶部であるメモリ 3 0 2 と、ネットワーク I / F 3 0 4 と、入力部 3 0 6 と、表示部 3 0 5 とを備える。

【 0 1 4 9 】

ストレージサブシステム制御プロセッサ 3 0 1 は、サブシステム管理装置 5 の機能を実現するためのプログラムを実行する。

【 0 1 5 0 】

メモリ 3 0 2 には、制御プロセッサ 3 0 1 が実行するプログラムおよびそのプログラムが使用する情報等が格納される。

【 0 1 5 1 】

ネットワーク I / F 3 0 4 は、ネットワーク 7 とのインタフェースである。ストレージサブシステム管理装置 5 は、ネットワーク 7 を介してストレージサブシステム 2 0、2 1 のシステム構成に関する情報を取得し、また、管理者から受け付けた構成定義（例えば、R A I D の構成定義や論理記憶装置の定義とそのバス定義処理、スナップショットペア定義など）、をストレージサブシステム 2 0、2 1 に送信する。

【 0 1 5 2 】

入力部 3 0 6 および表示部 3 0 5 は、ストレージサブシステム 2 0、2 1 の管理者とのインタフェースである。入力部 3 0 6 は、ストレージサブシステム 2 0、2 1 の保守 / 管理やりストア処理の指示や構成変更計画立案のための情報、例えば参照するリソースの稼働情報の期間や閾値、構成変更計画立案を実行する時刻などの入力を受け付け、また、表示部 3 0 5 に必要な情報を表示する。

【 0 1 5 3 】

『 ホスト計算機 1 』

図 1 6 に、本実施形態のホスト計算機 1 のハードウェア構成の一例を示す。

【 0 1 5 4 】

本図に示すように、ホスト計算機 1 は、所定のプログラムを実行する C P U 4 0 1 と、C P U 4 0 1 が実行する O S、アプリケーションプログラム（以下、A P と呼ぶ。）および、A P が使用するデータなどを格納メモリ 4 0 2 と、O S、A P、および、A P が使用するデータが格納されるローカルディスク装置 4 0 3 と、第一の I / O 経路 6 1 とホスト計算機 1 とを接続するホストバスアダプタ 4 0 4 と、ネットワーク 7 とホスト計算機 1 とを接続するためのネットワークアダプタ 4 0 5 と、フレキシブルディスク等の可搬記憶

10

20

30

40

50

メディアからのデータの読み出し等を制御するリムーバブル記憶ドライブ装置 407 と、これらの構成要素間を接続し、OS、AP、データ、制御データなどの転送に用いられる内部バスであるローカル I/O ネットワーク 409 とを備える。

【0155】

可搬記憶媒体としては、CD-ROM、CD-R、CD-RW、DVD や MO 等の光ディスク、光磁気ディスクや、ハードディスクやフレキシブルディスク等の磁気ディスク等が使用可能である。

【0156】

なお、以下に説明する各処理部は、可搬記憶媒体に格納されたプログラムをリムーバブル記憶ドライブ装置 407 を介して読み出す、または、外部のネットワークなどを経由して

10

ホスト計算機 1 にインストールされる。

【0157】

また、ホスト計算機 1 は、CPU 401、ローカルディスク 403、メモリ 402、ホストバスアダプタ 404、ネットワークアダプタ 408 を複数備えてもよい。

【0158】

『SAN 管理端末 9』

図 17 に、本実施形態の SAN 管理端末 9 のハードウェア構成の一例を示す。

【0159】

本図に示すように、SAN 管理端末 9 は、制御プロセッサ 501 と、電氣的に不揮発な記憶部であるメモリ 502 と、ネットワーク I/F 504 と、入力部 506 と、表示部 505 と、各々の構成要素を接続してデータや制御命令等を伝送する内部バスである伝送路 508 とを備える。

20

【0160】

メモリ 502 には、制御プロセッサ 501 が実行するプログラムおよびそのプログラムが使用する情報等が格納され、制御プロセッサ 501 は、SAN 管理端末 9 そのプログラムを実行する。

【0161】

ネットワーク I/F 304 は、ネットワーク 7 とのインタフェースである。

【0162】

『外部ストレージサブシステム 21 への I/O 処理』

30

次に、外部ストレージサブシステム 21 への I/O 処理の流れを説明する。図 18、図 19 は、外部ストレージサブシステム 21 への I/O 処理の流れを説明するための図である。

【0163】

ホスト計算機 1 からの I/O 処理要求は、ストレージサブシステム 20 が受領し、外部ストレージサブシステム 21 へ転送する。この場合、外部ストレージサブシステム 21 への I/O 処理要求であっても、ディスクキャッシュ 108 は使用される。

【0164】

まず、リード要求があった場合の処理について図 18 を用いて説明する。

【0165】

40

コマンド処理部 202 は、ホスト計算機 1 からリード処理要求を受領すると、受領した要求を解析し（ステップ 1601）、対象となるリードデータのアドレスおよびホスト計算機 1 の入出力対象であるターゲット論理アドレスである該アドレスを、対応する LDEV 番号と LDEV アドレスとに変換する（ステップ 1602）。

【0166】

次に、キャッシュ Hit / Miss 判定処理部 210 は、前記 LDEV アドレスのデータがディスクキャッシュ 108 に存在するか否かを判定する。（ステップ 1603）。

【0167】

ここで、ディスクキャッシュ 108 に存在すると判定された場合（キャッシュヒット時）、コマンド処理部 202 は、そのデータをディスクキャッシュ 108 から読み出し、該

50

データをホスト計算機 1 に送信した後、完了報告をホスト計算機 1 に送信する（ステップ 1604、1605）。

【0168】

ステップ 1603 において、存在しないと判定された場合（キャッシュミス時）は、コマンド処理部 202 は、論理/物理対応情報 203 にアクセスし、ステップ 1602 において割り出した L D E V アドレスが、ストレージサブシステム 20 か外部ストレージサブシステム 21 に存在するか、判断する（ステップ 1612）。

【0169】

ここで、ストレージサブシステム 20 に存在する場合は、コマンド処理部 202 は、物理記憶装置 I / O 処理部 215 に I / O 処理対象アドレス、データ長等とともに、データのリード要求を送信し（ステップ 1618）、物理記憶装置 I / O 処理部 215 は、I / O 処理を行う（ステップ 1619）。

10

【0170】

ステップ 1612 において、外部ストレージサブシステム 21 に存在すると判断された場合は、コマンド処理部 202 は、外部ストレージ I / O 処理部 216 に、受領した I / O 処理要求を送信する（ステップ 1615）。

【0171】

I / O 処理を受領した外部ストレージ I / O 処理部 216 は、I / O 処理要求内のアドレスに従って、外部ストレージサブシステム 21 にアクセスし、指定されたデータを読み出し（ステップ 1616）、ディスクキャッシュ 108 に格納し（ステップ 1616）、格納したことを、コマンド処理部 202 に通知する（ステップ 1617）。

20

【0172】

通知を受け取ったコマンド処理部 202 は、ディスクキャッシュ 108 から通知を受けたデータを読み出し、該データをホスト計算機 1 に送信し（ステップ 1621）、その後、完了報告をホスト計算機 1 に送信する（ステップ 1622）。

【0173】

次にライト処理要求があった場合の処理について図 19 を用いて説明する。

【0174】

ホスト計算機 1 からライト処理要求を受信したコマンド処理部 202 は、リード時と同様、受領した要求を解析し（ステップ 1701）、ライト処理準備ができていることをホスト計算機 1 に通知し、その後送信されてくるライトデータのアドレス、ホスト計算機 1 の入出力対象であるターゲット論理アドレスである該アドレス、を対応する L D E V 番号と L D E V のアドレスとに変換する（ステップ 1702）。

30

【0175】

そして、キャッシュ Hit / Miss 判定処理部 210 は、そのアドレスを基にディスクキャッシュ 108 上のデータの Hit / Miss 判定を行う（ステップ 1703）。

【0176】

ステップ 1703 において、キャッシュヒット時は、コマンド処理部 202 は、ヒットしたキャッシュの領域に受領したライトデータを上書きする。そして、キャッシュミス時は、ディスクキャッシュ 108 内に新たな領域を確保し、受領したライトデータをそこに格納し（ステップ 1704）、完了報告をホスト計算機 1 に通知する（ステップ 1705）。

40

【0177】

なお、ライトデータをキャッシュに格納した場合、キャッシュ上のアドレス情報や格納されるべき物理記憶装置のアドレス情報などを処理要求情報として、ダーティキューに登録する。

【0178】

次に、一旦ディスクキャッシュ 108 に格納したデータを物理記憶装置 110、124 に実際に書き込むライトアフト時の処理について説明する。

【0179】

50

物理記憶装置 I / O 処理部 2 1 5 または外部ストレージ I / O 処理 2 1 6 は、前記ダーティキューを参照し、処理すべきキューが存在した場合（ステップ 1 7 1 1）、キャッシュ上に存在する転送データ（ダーティデータ）を確定して（ステップ 1 7 1 2）ダーティデータをリード（ステップ 1 7 1 3）し、それぞれの物理記憶装置 1 1 0、1 2 4 にライト処理を行う（ステップ 1 7 1 4）。それぞれの物理記憶装置 1 1 0、1 2 4 からライト完了の報告を受領すると（ステップ 1 7 1 5）、前記ダーティキューをクリーンキューにつながり変える。

【 0 1 8 0 】

『性能劣化について』

上記機能、情報を用いて、性能チューニングを行なう処理を説明する前に、本実施形態のストレージサブシステムにおいて、性能チューニングを要する性能の劣化について説明する。

10

【 0 1 8 1 】

ホスト計算機 1 に直接接続されているストレージサブシステム 2 0 の場合は、アクセス衝突などにより、性能が劣化する。

【 0 1 8 2 】

すなわち、ストレージサブシステム 2 0 がホスト計算機 1 に提供する論理記憶装置 8 は、複数の物理記憶装置 1 1 0 からなるパリティグループによって提供される。このため、ホスト計算機 1 からみて、異なる論理記憶装置 8 へのアクセスであっても、それが、同一パリティグループを構成する物理記憶装置 1 1 0 へのアクセスとなることがあり、アクセス衝突が発生し、処理が遅延する。

20

【 0 1 8 3 】

また、外部ストレージサブシステム 2 1 を含めたシステム全体の性能劣化は、I / O 付加の上昇などによっても発生する。

【 0 1 8 4 】

本実施形態のように、外部ストレージサブシステム 2 1 への I / O 処理を、ストレージサブシステム 2 0 を介して行なう状況で、システム全体の性能を劣化させる要因となる、I / O 負荷の上昇について、説明する。

【 0 1 8 5 】

外部ストレージサブシステム 2 1 への I / O 負荷は、外部ストレージサブシステム 2 1 への I / O 処理要求の回数と I / O 処理応答時間との積で得られる。また、負荷が大きくなればなるほど、I / O 処理応答時間はさらに長くなる。このような負荷上昇に伴う I / O 処理応答の時間の遅延は様々な要因が考えられる。

30

【 0 1 8 6 】

例えば、ポート 1 0 4 b の競合がある。外部ストレージサブシステム 2 1 への I / O 処理はストレージサブシステム 2 0 を介して行なわれる。このストレージサブシステム 2 0 は、外部ストレージサブシステム 2 1 が複数ある場合であっても、それぞれの外部ストレージサブシステム 2 1 に対して別個のイニシエータポート 1 0 4 b を備えず、ポート 1 0 4 b を共用することがある。このような場合、ストレージサブシステム 2 0 側のポート 1 0 4 b という共有リソースでの処理が競合することにより I / O 処理の遅延が発生し、それが I / O 処理応答時間の遅延となって表面化する。

40

【 0 1 8 7 】

また、制御プロセッサ 1 0 0 の競合も考えられる。外部ストレージサブシステム 2 1 に対する I / O 処理要求を処理する外部ストレージ I / O 処理部 2 1 6 が処理を行なうにあたり、ストレージサブシステム 2 0 内部の I / O 処理等と競合し、制御プロセッサ 1 0 0 のリソースが十分に割り振れないことがある。このような要因で、I / O 処理応答の時間的な遅延が発生する。

【 0 1 8 8 】

このように、外部ストレージサブシステム 2 1 を接続した構成の場合、現象として現れている性能劣化が、単に同じパリティグループを構成する物理記憶装置へのアクセス衝突

50

によるものだけでなく、ポート 104b の競合、制御プロセッサ 100 の競合によるものである場合も考えられる。従って、性能チューニングにあたっては、これらの要因を考慮する必要がある。

『性能チューニング処理の流れ』

次に、上記の各機能を用い、外部ストレージサブシステム 21 を含めた性能チューニング処理の流れを以下に説明する。本実施形態においては、まず、所定の時間間隔ごとに自動で性能チューニングを行う、すなわち、性能スケジュール情報 222 に従って、構成変更を行うための種々の処理を自動的に進めることを指示する性能改善機能が ON であるか否かを判断し、当該機能が ON の場合、処理を進める。ここでの処理は、構成変更計画の立案、ストレージサブシステム 20 の性能チューニング、外部ストレージサブシステム 21 の性能チューニングの順に行われる。そして、外部ストレージサブシステム 21 の性能チューニングを行う場合は、前もって、管理端末を介して、管理者の承認を得る。

10

【0189】

図 20 に、ストレージサブシステム 20 の性能チューニング時の処理フローを示す。

【0190】

マネージャ 223 は、一定時間間隔ごとにスケジュール情報 222 を参照し、ストレージサブシステム 20 の、性能改善機能が ON、すなわち、性能チューニングを行う状態になっているかを確認する（ステップ 2001）。なお、参照時間間隔や性能改善機能の指示は、ストレージサブシステム管理装置 5 を介して、管理者より入力される。

【0191】

20

ストレージサブシステム 20 の性能改善機能が ON である場合、マネージャ 223 は、性能スケジュール情報 222 を参照し、構成変更計画立案情報 227 の更新を行う時刻であるかを確認する（ステップ 2002）。

【0192】

構成変更計画立案情報 227 の更新時刻であれば、マネージャ 223 は、構成変更計画立案処理部 218 に構成変更計画立案情報を作成させる（ステップ 2031）。本ステップ 2031 の詳細な処理フローは図 21 に従って後述する。

【0193】

ステップ 2002 において、構成変更計画立案情報 227 の更新時刻でなければ、マネージャ 223 は、スケジュール情報 222 を参照し、構成変更計画の実行時刻であるかを確認する（ステップ 2003）。

30

【0194】

ステップ 2003 において、実行時刻でない場合は、処理を終了する。実行時刻である場合は、マネージャ 223 は、構成変更計画立案情報 227 を参照し、構成変更計画が格納されているか否かを確認する（ステップ 2004）。

【0195】

ステップ 2004 において、構成変更計画が格納されていない場合は終了する。格納されている場合は、マネージャ 223 は、その構成変更計画情報 227 を参照し、構成変更計画に、ストレージサブシステム 20 に関するものがあるかを確認する（ステップ 2005）。

40

【0196】

ステップ 2005 において、ストレージサブシステム 20 に関する変更計画が格納されていた場合は、マネージャ 223 は、構成変更計画実行処理部 219 に、その変更計画に従い構成変更を実行させ（ステップ 2011）、ステップ 2006 に進む。

【0197】

ステップ 2005 において、ストレージサブシステム 20 に関する変更計画がなかった場合は、マネージャ 223 は、構成変更計画情報 227 に、外部ストレージサブシステム 21 に関するものが格納されているか否かを確認する（ステップ 2006）。

【0198】

ステップ 2006 において、外部ストレージサブシステムに関する構成変更計画が格納

50

されていない場合は、処理を終了する。格納されている場合は、マネージャ 2 2 3 は、ストレージサブシステム管理装置 5 をの表示部に外部ストレージサブシステム 2 1 に関する構成変更計画があることと、構成変更計画情報 2 2 7 から抽出したその変更計画を表示し、管理者に提示する（ステップ 2 0 0 7）。

【 0 1 9 9 】

サブシステム管理者による前記推奨構成変更の実行指示を受け付けると、マネージャ 2 2 3 は、構成変更計画実行処理部 2 1 9 は、構成変更を実行し（ステップ 2 0 2 1）終了し、指示がなければ、そのまま終了する。

【 0 2 0 0 】

なお、サブシステム管理装置 5 を介して、構成変更計画を提示された利用者は、その構成変更計画が必要でないものと判断すれば、変更計画のキャンセルをサブシステム管理装置 5 指示することもできる。

10

【 0 2 0 1 】

次に、上記のステップ 2 0 3 1 において、構成変更計画立案情報 2 2 7 を作成する処理を、以下に説明する。本実施形態では、パリティグループの負荷をモニタし、高負荷のものがあった場合、必要に応じて構成変更計画を立案する。高負荷のパリティグループがストレージサブシステム 2 0 のものであった場合は、特許文献 2 に開示されている方法を用いて立案する。

【 0 2 0 2 】

ここで、高負荷のパリティグループが、外部ストレージサブシステム 2 1 のものであった場合、上述のように、性能劣化の要因は、単に同じパリティグループを構成する物理記憶装置へのアクセス衝突によるものだけでなく、イニシエータポート 1 0 4 b の競合、制御プロセッサ 1 0 0 の競合によるものである場合も考えられる。このため、本実施形態では、まず、イニシエータポート 1 0 4 b の競合、制御プロセッサ 1 0 0 の競合によるものかどうか、確認する。そして、これらが要因でない場合、アクセス衝突によるものと判断し、データの再配置を検討する。本実施形態では、データの再配置として、外部ストレージサブシステム 2 1 からストレージサブシステム 2 0 へデータを移行する案を立案する。また、本実施形態では、イニシエータポートの競合 1 0 4 b や制御プロセッサ 1 0 0 の競合などの場合の対処も、それぞれ、構成変更計画案の 1 つとしてステップ 2 0 3 1 で立案し、実行前に管理者の許可を得る。

20

30

【 0 2 0 3 】

ここでは、アクセス衝突以外の要因であるか否かの判断として、スループットなどの稼働情報、キャッシュの格納状態、イニシエータポート、プロセッサの負荷状況等の、ストレージサブシステム 2 0 において収集している外部ストレージサブシステム 2 1 の状態を示す各種情報をチェックする。

【 0 2 0 4 】

図 2 1 は、構成変更計画立案情報 2 2 7 作成時の構成変更計画立案処理部 2 1 8 の処理フローである。

【 0 2 0 5 】

構成変更計画立案処理部 2 1 8 は、スケジュール情報 2 2 2 に従い、物理記憶装置稼働情報 2 2 5 を参照し、I/O 処理性能を示す、例えば、総占有時間 1 3 1 0 が、予め設定されている閾値よりも大きいものを、高負荷のパリティグループをとして抽出する（ステップ 1 5 0 1）。

40

【 0 2 0 6 】

構成変更計画立案処理部 2 1 8 は、ステップ 1 5 0 1 において抽出したパリティグループがストレージサブシステム 2 0 のものか、外部ストレージサブシステム 2 1 のものかを、物理記憶装置稼働情報 2 2 5 の P . G . 番号 1 3 0 2 によって判別する（ステップ 1 5 0 2）。

【 0 2 0 7 】

ステップ 1 5 0 2 において、ストレージサブシステム 2 0 内のパリティグループと判断

50

された場合、構成変更計画立案処理部 218 は、特許文献 2 に開示されている方法などを用い、構成変更計画を立案する（ステップ 1521）。

【0208】

ステップ 1502 において、外部ストレージサブシステム 21 内のデバイスと判断された場合、構成変更計画立案処理部 218 は、当該外部ストレージサブシステム 21 に対して I/O 処理を行うイニシエータポート 104b のレスポンス時間およびスループットを、稼動情報 229 およびポート設定情報 228 を参照して確認する（ステップ 1503）。

【0209】

ここでは、レスポンス時間を確認するために、稼動情報 229 の平均 IOPS 1204 とポート設定情報 228 の目標 IOPS 1407 とを比較し、スループット性能を確認するために、稼動情報 229 の平均 MBPS 1206 とポート設定情報 228 の目標 MBPS 1408 とを比較する。

【0210】

構成変更計画立案処理部 218 は、平均 MBPS 1206 および平均 IOPS 1204 によって示される性能がともに目標としてポート設定情報 228 に設定されている値により示される性能を上回る場合、特に問題なしと判断し、処理を終える。

【0211】

ここで、平均 MBPS 1206 または平均 IOPS 1204 のどちらか一方の値によって示される性能のみが、目標としてポート設定情報 228 に設定されている値により示される性能を下回っている場合（ステップ 1504）、まず、P.G. 稼動情報 225 のシーケンシャル比率 1312 を確認し、その時に送信されているデータの種別がシーケンシャルデータかランダムデータかを判別する。シーケンシャル比率 1312 の値が所定の閾値より大きい、すなわち、シーケンシャルデータが送信されている場合は（ステップ 1511）、レスポンス時間が大きい、または、スループット性能が悪化していても問題がないものと判断し、処理を終える。

【0212】

一方、稼動情報 225 として与えられるレスポンス時間およびスループット性能が、ともに、目標としてポート設定情報 228 に与えられている性能を下回っている場合（ステップ 1504）、または、どちらか一方が下回り、かつ、シーケンシャル比率が所定の閾値より小さい場合（ステップ 1511）は、ストレージサブシステム 20 側にボトルネックがある場合が考えられる。すなわち、ストレージサブシステム 20 と外部ストレージサブシステム 21 との間の物理的な接続に問題がある可能性がある。

【0213】

その場合、構成変更計画立案処理部 218 は、当該デバイスに関するキャッシュのダーティ量が増加しているか否かを確認する（ステップ 1505）。ここでは、キャッシュ管理情報 211 を参照し、まず、全体のキャッシュ量にダーティ量が占める割合を、クリーンカウンタ 1804、ダーティカウンタ 1805、フリーカウンタ 1806 とを用いて算出する。

【0214】

ここで、ダーティ量が占める割合が所定の値以上の場合、その原因となっている論理記憶装置を抽出する。すなわち、キャッシュ管理情報 211 に格納されている個々の論理記憶装置毎のダーティカウンタ 1805 に着目し、大きい値を示す論理記憶装置番号を抽出する。抽出にあたっては、値の大きいものから上位いくつかを抽出してもよいし、予め定められた閾値以上のものを抽出してもよい。

【0215】

抽出したものの中に、外部ストレージサブシステム 21 の論理記憶装置があった場合、物理的な接続に関して何らかの問題が発生し、データが送出できない事態が考えられるため、接続状態を確認する（ステップ 1512）。接続状態に問題があった場合は、構成変更計画立案処理部 218 は、その旨を示すアラートを表示し（ステップ 1522）、処理

10

20

30

40

50

を終了する。

【0216】

ステップ1522において、ダーティ量が所定の割合以下であった場合、ストレージサブシステム20の論理記憶装置が原因でダーティ量が増えている場合、または、ステップ1523で接続状態に問題がなかった場合は、構成変更計画立案処理部218は、イニシエータポート104bの負荷の状況を確認する(ステップ1506)。

【0217】

ここで、構成変更計画立案処理部218は、ステップ1501で高負荷パリティグループと判断されたデバイスに対するイニシエータポート104bについて、処理能力の限界であるか否か判別する。すなわち、ポート稼働情報229の平均データ転送量1206および平均レスポンス時間1204とポート設定情報228の目標データ転送量1408および目標レスポンス時間1407とをそれぞれ比較し、平均データ転送量1206が目標データ転送量1408を下回っている場合、または、平均レスポンス時間1204が目標レスポンス時間1407を超えている場合は、高負荷と判断する。

【0218】

ステップ1506において、イニシエータポート104bが高負荷と判断された場合、構成変更計画立案処理部218は、代替可能な余裕のあるイニシエータポート104bがあるか否か判断する(ステップ1513)。本実施形態では、ポート設定情報228の目標データ転送量1408とポート稼働情報229の平均データ転送量1206とを比較し、上記高負荷と判断されたイニシエータポート104bの負荷分を加えても目標データ転送量1408を超えないイニシエータポート104bがあれば、当該イニシエータポート104bを代替可能なイニシエータポート104bと判断し、抽出する。

【0219】

そして、代替可能なイニシエータポート104bを使用して外部ストレージサブシステム21とのI/O処理を行うよう構成を変更する案を構成変更計画として作成して構成変更計画立案情報227に格納し(ステップ1523)、処理を終える。

【0220】

また、ステップ1513で、代替可能なイニシエータポート104bがない場合は、複数のイニシエータポート104bに、前記高負荷と判断されたイニシエータポート104bの負荷分を分散可能か確認する(ステップ1514)。

【0221】

複数のイニシエータポート104bであって、分散して負荷を負担することが可能なイニシエータポート104bがある場合、それらのイニシエータポート104bを使用して外部ストレージサブシステム21とのI/O処理を行うよう構成を変更する案を構成変更計画として作成して構成変更計画立案情報227に格納し(ステップ1524)、処理を終える。

【0222】

ステップ1514において、交替可能なイニシエータポート104bがない場合、外部ストレージサブシステム21からストレージサブシステム20へのデータの移動を考慮するため、ステップ1508の処理に進む。

【0223】

ステップ1505において、イニシエータポート104bの負荷が限界でないと判断した場合、構成変更計画立案処理部218は、プロセッサ稼働情報226を参照し、制御プロセッサ100の稼働状態を確認する(ステップ1507)。

【0224】

ステップ1507において、プロセッサ稼働率が予め定められた閾値と比較して高いと判断した場合は、構成変更計画立案処理部218は、上記のイニシエータポート104bの場合の処理と同様の手順で、プロセッサ稼働情報226の、状態1452がイニシエータである他のプロセッサの中から、代替可能な制御プロセッサ100があるか否か(ステップ1515)、または、複数の制御プロセッサ100によって処理負荷を分担可能か(

10

20

30

40

50

ステップ1516)を判断する。

【0225】

代替可能な制御プロセッサ100がある場合、構成変更計画立案処理部218は、当該制御プロセッサ100を使用して外部ストレージサブシステム21とのI/O処理を行うよう構成を変更する案を、構成変更計画として作成して構成変更計画立案情報227に格納し(ステップ1525)、処理を終える。

【0226】

また、分散可能な複数の制御プロセッサ100がある場合、構成変更計画立案処理部218は、それらの複数の制御プロセッサ100を使用して外部ストレージサブシステム21とのI/O処理を行うよう構成を変更する案を、構成変更計画として作成して構成変更計画立案情報227に格納し(ステップ1526)、処理を終える。

10

【0227】

ステップ1516において、交替可能な制御プロセッサ100がない場合、外部ストレージサブシステム21からストレージサブシステム20へのデータの移動を考えるため、ステップ1508の処理に進む。

【0228】

また、ステップ1507において、制御プロセッサ100の稼働率が予め定めた閾値よりを超えていなかった場合も、同様に、ステップ1508に進む。

【0229】

ステップ1508において、構成変更計画立案処理部218は、外部ストレージサブシステム21からストレージサブシステム20へのデータの移動を考える。

20

【0230】

構成変更計画立案処理部218は、ストレージサブシステム20内に外部ストレージサブシステム21のデータを移行可能なだけの空き容量があるかを判断するため、物理記憶装置稼働情報225および図示しない空き容量管理情報を参照する(ステップ1509)。ここで、空き容量管理情報は、各パリティグループごとの容量および利用率を管理するデータベースである。

【0231】

ステップ1509において、物理記憶装置稼働情報225と空き容量管理情報とに従って、ストレージサブシステム20内に、外部ストレージサブシステム21パリティグループであって、ステップ1501で高負荷と判断されたパリティグループの容量分の余裕のあるパリティグループがある場合(ステップ1510)、構成変更計画立案処理部218は、ステップ1501で高負荷と判断された外部ストレージサブシステム21のパリティグループを、ステップ1510で余裕があると判断されたストレージサブシステム20のパリティグループに移動する構成変更計画を作成して構成変更計画立案情報227に登録し(ステップ1511)、処理を終了する。

30

【0232】

ステップ1510において、移動可能なパリティグループを見つけることができない場合は、外部ストレージサブシステム21へのI/O処理に問題があることをAlert情報としてサブシステム管理装置5へ通知することでユーザに提示し(ステップ1527)、処理を終了する。

40

【0233】

以上の処理により、構成変更計画は立案され、構成変更計画立案情報227に格納される。

【0234】

なお、外部ストレージサブシステム21の論理記憶装置8内にあるデータをディザスタリカバリ用に遠隔地にあるストレージサブシステム20にコピーする場合、ストレージサブシステム20にはあるが外部ストレージサブシステム21には無い機能を用いる場合、または、ストレージサブシステム20から外部ストレージサブシステム21へのI/Oネットワークの帯域が狭くかつ外部ストレージサブシステム21の論理記憶装置へのI/O

50

処理が頻繁である場合などに、外部ストレージサブシステム 21 からストレージサブシステム 20 へのデータの移行は、性能改善のために特に有効である。

【0235】

また、ストレージサブシステム 20 のディスクキャッシュ 108 にデータを常駐させると更に性能向上が望める。

【0236】

なお、上記のステップ 1523、1524 におけるイニシエータポート 104b の変更、および、ステップ 12525、1526 における制御プロセッサ 100 の変更は、構成変更計画として提示せず、上記のステップにおいて、処理可能なイニシエータポート 104b および制御プロセッサ 100 が確定した時点で、実行するよう構成してもよい。

10

【0237】

本実施形態において、上記の性能チューニングにおいては、外部ストレージサブシステム 21 の属性情報である外部ストレージ記憶装置属性情報 220 を予め保持していることを前提としている。

【0238】

外部ストレージサブシステム 21 の物理記憶装置 124 の性能がストレージサブシステム 20 内部の物理記憶装置 110 と同様には評価できない場合がある。例えば、第二の I/O 経路 62 上を介して行われるストレージサブシステム 20 からの I/O 処理は、外部ストレージサブシステム 21 の当該論理記憶装置 8 への I/O 以外に、他の外部ストレージサブシステムの論理記憶装置 8 に対する I/O もある。注目する外部ストレージサブシステム 21 の論理記憶装置 8 に対する I/O 性能は、それらの各処理の干渉によるネットワークへの負荷や、スイッチ 130 への負荷の影響を受ける。しかし、これらの負荷がどの程度かが把握できない。また、外部ストレージサブシステム 21 にもディスクキャッシュ 108 を有する場合もある。外部ストレージサブシステム 21 内でキャッシュヒットしたか否かはストレージサブシステム 20 からは把握できない。

20

【0239】

このように、不確定な要素があり、ディスクの性能が評価できない場合は、性能改善のための構成変更には、外部ストレージサブシステム 21 を接続する機能をもつストレージサブシステム 20 のストレージ管理者による判断があるほうが望ましい。

【0240】

30

『ホスト計算機 1 に透過的にデバイスの移動を行う方法』

次に、ホストに透過的にデバイスの移動を行う方法を、図を用いて説明する。

【0241】

図 22 は、第 1 の論理記憶装置から第 2 の論理記憶装置にデータをコピーする場合の処理を説明するための図である。

【0242】

ここでは、構成変更計画立案処理部 218 において立案された計画が、第 1 の内部論理記憶装置 (LDEV1) 2104 内のデータを第 2 の内部論理記憶装置 (LDEV2) 2105 にコピーするものである場合を例にあげ、説明する。

【0243】

40

コマンド処理部 202 は、論理 / 物理対応情報 203 を参照して、ホスト計算機 1 の I/O 処理対象である論理アドレスに対応する物理アドレスに対して、キャッシュ Hit / Miss 判定処理部 210 に、ディスクキャッシュ 108 のメモリ領域を割り当てさせる。

【0244】

物理記憶装置 I/O 処理部 215 は、ライト処理要求の場合は、割り当てられたメモリ領域のデータを物理記憶装置 110 に書き込み、リード処理要求の場合は、割り当てられたメモリ領域上に格納されたデータを物理記憶装置 110 から読み込む。コマンド処理部 202 は、割り当てられたメモリ領域上に格納されたデータを、リードデータとしてホスト計算機 1 に転送する。

50

【 0 2 4 5 】

ここで、ホスト計算機 1 は、ストレージサブシステム 2 0 の、I / O ポートアドレス x、論理記憶装置番号 y で識別される論理記憶装置 2 1 0 1 に対して I / O 処理を行うものとする。

【 0 2 4 6 】

また、ストレージサブシステム 2 0 の第 1 の内部論理記憶装置 (L D E V 1) 2 1 0 4 と第 2 の内部論理記憶装置 (L D E V 2) 2 1 0 5 とのデータは、それぞれ、第 1 のパリティグループ (P G 1) 2 1 0 8 と第 2 のパリティグループ (P G 2) 2 1 0 9 とを構成する物理記憶装置 1 1 0 上に格納されているものとする。そして、第 1 のパリティグループ 2 1 0 8 は、第 1 の内部論理記憶装置 2 1 0 4 の他に第 3 の内部論理記憶装置 (L D E V 3) 2 1 0 6 を備える。また、第 2 のパリティグループ 2 1 0 9 は、第 2 の内部論理記憶装置 2 1 0 5 の他に第 4 の内部論理記憶装置 2 1 0 7 を備える。

10

【 0 2 4 7 】

ただし、第 2 内部論理記憶装置 (L D E V 2) 2 1 0 5 はリザーブ状態とし、ホスト計算機 1 からのデータの入出力対象にはならないようにガードがかけてあるものとする。また、第 2 の内部論理記憶装置 (L D E V 2) 2 1 0 5 は第 1 の内部論理記憶装置 (L D E V 1) 2 1 0 4 と同じエミュレーションタイプとする。

【 0 2 4 8 】

構成変更実行処理部 2 1 9 は、第 1 の内部論理記憶装置 (L D E V 1) 2 1 0 4 内のデータを第 2 の内部論理記憶装置 (L D E V 2) 2 1 0 5 にコピーする。

20

【 0 2 4 9 】

このとき、コピー処理中も第 1 の内部論理記憶装置 (L D E V 1) 2 1 0 4 へのホスト計算機 1 からのリード / ライト処理要求は受け付ける。例えば、ライト処理要求を受け付けた場合、コマンド処理部 2 0 2 は、第 1 の内部論理記憶装置 2 1 0 4 のライト更新が指示されたデータが、既に第 2 の内部論理記憶装置 2 1 0 5 へコピー済みであるか否か判断する。コピー先にコピーされていないければ、コピー元にそのままライトし、コピー済みであれば、そのつどコピー先にコピーを行う。

【 0 2 5 0 】

ある時刻 T 1 (ステップ 2 1 1 1) において、コピー元の第 1 の内部論理記憶装置 (L D E V 1) 2 1 0 4 からコピー先の第 2 の内部論理記憶装置 (L D E V 2) 2 1 0 5 に対するコピー処理が、コマンド処理部 2 0 2 がキャッシュメモリ 1 0 8 にライト領域を確保して書き込むデータ (しかかり中のライトデータ)、および、ディスクキャッシュ 1 0 8 上にあってまだ物理記憶装置 1 1 0 にライトされていないデータを除いて、すべて終了したら、コマンド処理部 2 0 2 は、その瞬間から第 1 の内部論理記憶装置 2 1 0 4 に対するコマンド処理およびライトデータ受領を、I / O アダプタ 1 0 2 上のバッファメモリにキューイングして、しかかり中のライトデータおよびキャッシュメモリ 1 0 8 上のデータをコピー先にコピーする。また、構成変更実行処理部 2 1 9 は、論理 / 物理対応情報 2 0 3 のコピー元の第 1 の内部論理記憶装置 2 1 0 4 とコピー先の第 2 の内部論理記憶装置 2 1 0 5 との物理記憶装置のアドレス対応情報を入れ替える。

30

【 0 2 5 1 】

時刻 T 以降のある時刻 T 2 (ステップ 2 1 1 2) では、コマンド処理部 2 0 2 は、キューイングしていたコマンドおよびライトデータの処理を、対応情報が入れ替えられた論理 / 物理対応情報 2 0 3 を用いて行ない、その応答をホスト計算機 1 に返す。

40

【 0 2 5 2 】

これによって、ストレージサブシステム 2 0 に対して I / O 処理を行うホスト計算機 1 は、常に I / O ポートアドレスが x、論理記憶装置番号が y で識別される論理記憶装置 2 1 0 1 に対して I / O 処理を実行することとなる。従って、ホスト計算機 1 は、データの物理的位置が変更されたにもかかわらず、それを意識せずに、I / O 処理を続けることができる。

【 0 2 5 3 】

50

本実施形態では、上述のように、接続されている外部ストレージサブシステム 21 も含めて性能チューニングが行なわれる。このため、外部ストレージサブシステム 21 に対する I/O 処理負荷が上昇し、必要な性能が得られなくなった場合に、外部ストレージサブシステム 21 から内部ストレージサブシステム 20 へデータを移動するといった構成変更計画が立案されることがある。

【0254】

この場合も、移動元の外部ストレージサブシステム 21 の論理記憶装置 8 を上記の説明における第 1 の内部論理記憶装置、移動先のストレージサブシステム 20 の論理記憶装置 8 を、上記の第 2 の内部論理記憶装置とし、同様に、コピー処理と論理 / 物理対応情報 203 の書き換えによって、ホスト透過的に移動を実現することができる。

10

【0255】

以上説明したホスト透過的に移動する際の、構成変更実行処理部 219 による一連の処理の流れを図 23 に示す。

【0256】

LDEV1 から LDEV2 にデータの移動を行う場合、コピーポインタ (CP) を LDEV1 の先頭アドレスに設定する (ステップ 2201)。

【0257】

次に、LDEV1 内のデータをアドレス CP から N (Byte) コピーする (ステップ 2202)。そして、コピーポインタ CP をコピーデータ量分 (N (Byte)) 進める (ステップ 2203)。

20

【0258】

CP の値と LDEV1 の容量を比較して (ステップ 2204)、小さければ CP に更に次に LDEV1 から LDEV2 にコピーするデータ量 N (Byte) 分を足した CP + N と LDEV1 の容量を比較して (ステップ 2205)、小さければ、次にコピーするデータ量 N (Byte) の値を (LDEV1 の容量) - CP に設定する (ステップ 2206)。

【0259】

ステップ 2202 からステップ 2206 までの流れをステップ 2204 での評価が NO になるまで行い、LDEV1 から LDEV2 へのコピー量が LDEV1 の容量分実施されたら、次に LDEV1 へのライトデータでディスクキャッシュ 108 上において物理記憶措置 110 に反映されていないものがあるかを確認する (ステップ 2207)。

30

【0260】

あれば LDEV1 のある PG1 上のディスクに優先的にデータをライトする (ステップ 2210)。

【0261】

ステップ 2210 のライト処理終了後、ステップ 2207 での評価を再度行い、LDEV1 へのライトデータでディスクキャッシュ 108 上において物理記憶措置 110 に反映されていないものがなくなるまで、ステップ 2207 および 2210 の処理を繰り返す。

【0262】

次に、コマンド処理部 202 が LDEV1 に関してディスクキャッシュ 108 を確保してディスクキャッシュ 108 に書き込む処理を行っているかを確認し (しかかり中のライト処理) (ステップ 2208)、しかかり中のライト処理がある場合、ディスクキャッシュ 108 へのそのライト処理を行わせ、優先的に LDEV1 のある PG1 上のディスクに優先的にデータをライトする (ステップ 2211)。

40

【0263】

ライト処理終了後、ステップ 2208 の評価を再度行い、しかかり中のライト処理が無いかを確認し、しかかり中のライト処理がない場合、キューイングした I/O 処理のうち、LDEV1 に対するものを、コマンド処理部 202 による処理を待ち状態にする (ステップ 2209)。

【0264】

50

その後 L D E V と P . G . のアドレスマップを入れ替える（ステップ 2 2 1 2）。すなわち、論理物理対応情報 2 0 3 の、対応情報を入れ替える。

【 0 2 6 5 】

また、ステップ 2 2 0 2 から 2 2 0 6 までのコピー処理とは独立に、当該コピー処理中またはステップ 2 2 0 9 の前までに、ライトデータが新たに処理された場合の処理について、図を用いて、説明する。

【 0 2 6 6 】

図 2 4 は、ライトデータが新たに処理された場合の処理フローである。

【 0 2 6 7 】

上記のタイミングでライトデータが新たに処理された場合、L D E V 1 のライトデータかどうかを評価する（ステップ 2 2 2 1）。10

【 0 2 6 8 】

L D E V 1 のライトデータであれば、ライト処理アドレスを抽出し（ステップ 2 2 2 2）、ライトアドレスとコピーポインタを評価する（ステップ 2 2 2 3）。

【 0 2 6 9 】

ステップ 2 2 2 3 において、ライトアドレスがコピーポインタよりも先のアドレスであれば、当該アドレスにデータをライトし（ステップ 2 2 2 4）処理を終了する。

【 0 2 7 0 】

一方、ステップ 2 2 2 3 において、ライトアドレスがコピーポインタよりも後のアドレスであれば、L D E V 1 のアドレス W A にデータをライトするとともに、移動先の L D E V 2 の該当するアドレスにもデータをライト（ステップ 2 2 2 5）し、処理を終了する。20

【 0 2 7 1 】

また、ステップ 2 2 2 1 において、L D E V 1 へのライト処理以外の場合は、当該要求処理を実行して終了する。

【 0 2 7 2 】

以上説明したように、本実施形態によれば、ホスト計算機 1 から直接アクセスできない外部ストレージサブシステム 2 1 を接続したシステムであっても、その外部ストレージサブシステム 2 1 を含めて性能チューニングを行なうことができる。

【 0 2 7 3 】

『外部ストレージサブシステム 2 1 と第 2 の外部ストレージサブシステム 2 2 との間のデータの移行』30

次に、第二の I / O 経路 9 を介してストレージサブシステム 2 0 に接続される、外部ストレージサブシステム 2 1 と第 2 の外部ストレージサブシステム 2 2 との間でのデータ移行について説明する。

【 0 2 7 4 】

ここで、外部ストレージサブシステム 2 2 は、第二の I / O 経路 6 2 を介してストレージサブシステム 2 0 に接続される。また、他の外部ストレージサブシステム 2 1 などと同様にサブシステム管理装置 5 を接続し、そのサブシステム管理装置 5 を介して、ネットワーク 7 に接続されている。

【 0 2 7 5 】

以下、図 2 5 を参照し、外部ストレージサブシステム間での移動の手順を説明する。40

【 0 2 7 6 】

構成定義処理 2 1 7 は、第二の I / O 経路 6 2 を介して、第二の I / O 経路 6 2 上にある外部ストレージサブシステム 2 2 を定義する。この定義方法は、I / O 処理に用いるプロトコルによって方法が異なるが、ここでは発明に直接関係しないので詳細は述べない。そして、外部ストレージサブシステム 2 2 の論理記憶装置を、I / O 処理対象論理記憶装置としてストレージサブシステム 2 0 の論理 / 物理記憶装置対応情報 2 0 3 に登録する。

【 0 2 7 7 】

ここで、外部ストレージサブシステム 2 2 の I / O ポートに対し、すでに、ホスト計算機 1 の識別子によるアクセス制御の設定が行なわれていた場合は、そのアクセス制限を外50

す。具体的には、外部ストレージサブシステム 22 に接続されるサブシステム管理装置 5 を介して、利用者からの指示により、そのアクセス制限を外すか、または、ストレージサブシステム 20 の I/O 処理要求ネットワーク処理部 200 により、外部ストレージサブシステム 22 に第二の I/O 経路を用いて I/O 処理を行うポートの識別子を、アクセス許可対照に設定する。

【0278】

アクセスの制限を外した後、ストレージサブシステム 20 のストレージサブシステム制御装置 112 により、外部ストレージサブシステム 22 の論理記憶装置をストレージサブシステム 20 の物理記憶装置として定義し、ストレージサブシステム 20 の論理記憶装置を定義する。そして、構成定義処理部 217 により、論理/物理対応情報 203 を更新/作成する(ステップ 2801)。

10

【0279】

次に、ダミーデータを用いて外部ストレージサブシステム 22 の I/O 処理の性能を計測する(ステップ 2802)。

【0280】

外部ストレージサブシステム 22 の I/O 処理性能が予め定められた期待値を満たさない場合は(ステップ 2803)、ストレージサブシステム制御装置 112 は、異なる外部ストレージサブシステムを採用するか(ステップ 2804)、または、ストレージサブシステム 20 内の物理記憶装置へデータの移行を行う(ステップ 2805)。

【0281】

20

外部ストレージサブシステム 22 に対する I/O 処理性能が、予め定められた期待値を満たす場合、または、I/O 処理性能に対する期待値が必要ない場合は、外部ストレージサブシステム 22 に対してのデータの移行を行う(ステップ 2806)。

【0282】

データの移行は、ストレージサブシステム 20 上で定義された外部ストレージサブシステム 21 の論理記憶装置を物理記憶装置としてマッピングするストレージサブシステム 20 の第 1 の論理記憶装置と、外部ストレージサブシステム 22 の論理記憶装置を物理記憶装置としてマッピングするストレージサブシステム 20 の第 2 の論理記憶装置との間のデータの移行方法と同様の方法で行なう。これは既に図 23 に示しているので、ここでは、再度説明しない。

30

【0283】

『外部ストレージサブシステム 21 内でのデータの移動』

次に、外部ストレージサブシステム 21 内でのデータの移動に関して説明する。外部ストレージサブシステム 21 内でのデータ移動時のストレージサブシステム制御装置 112 の処理フローを図 26 に示す。

【0284】

ストレージサブシステム 20 が I/O 処理対象とする外部ストレージサブシステム 21 内の第 1 の論理記憶装置と第 2 の論理記憶装置との稼動情報は、既に述べたように、ストレージサブシステム 20 からみた I/O 処理性能およびその履歴としてストレージサブシステム 20 で保持している。

40

【0285】

外部ストレージサブシステム 21 の第 1 の論理記憶装置に比べ第 2 の論理記憶装置のレスポンス性能がある時期を境に悪化した場合、何らかの性能問題があると考えられる。このような場合に、ストレージサブシステム制御装置 112 は、ストレージサブシステム 20 において外部ストレージサブシステム 21 の論理記憶装置に対するアクセスパタン(シーケンシャル比率とリードライト比率)を論理記憶装置稼動情報 224 を参照することにより取得する(ステップ 2701)。

次に、ストレージサブシステム制御装置 112 は、論理記憶装置への I/O 処理アドレスと第 2 の論理記憶装置への I/O 処理アドレスの偏りを I/O 対象アドレスの重なり具合から把握する(ステップ 2702)。

50

【 0 2 8 6 】

ここで、第 1 の論理記憶装置に比べ第 2 の論理記憶装置へのホスト計算機 1 の I / O 処理のアクセスローカリティが低い場合は、外部ストレージサブシステム 2 1 内でのディスクキャッシュ 1 0 8 に I / O 対象データが無い場合が多い可能性が高いため、レスポンス性能が悪いと判断する。

【 0 2 8 7 】

さほど I / O アクセス範囲のローカリティが変わらない場合は、ストレージサブシステム制御部 1 1 2 は、I / O 処理のデータ長を見る。データ長が長ければシーケンシャルアクセスなので、キャッシュ上にデータを残さないように外部ストレージサブシステム 2 1 側のキャッシュ制御が実行されていると判断し、スループットの値を参照し、予め定めた期待値が出ていれば問題ないと判断する（ステップ 2 7 0 3 ）。

10

【 0 2 8 8 】

データ長が短い場合、リードライト比率をみる。そして、第 2 の論理記憶装置への I / O 処理におけるリード比率が第 1 の論理記憶装置への I / O 処理のリード比率に比べて高い場合、ストレージサブシステム制御部 1 1 2 は、ディスクキャッシュからではなく物理記憶装置にデータを読みに行っているためにレスポンス性能が悪いと判断し、スループットの値を参照し、予め定めた期待値が出ていれば良いと判断する（ステップ 2 7 0 4 ）。

【 0 2 8 9 】

第 2 の論理記憶装置へのリード比率が第 1 の論理記憶装置へのリード比率以下の場合、ストレージサブシステム制御部 1 1 2 は、第 2 の論理記憶装置の配置されている物理記憶装置でのアクセスに衝突が起こっている、または、第 2 の論理記憶装置に何らかの障害が内在しているため外部ストレージサブシステム 2 1 内部で物理記憶装置へのアクセスに対してリトライがかかっている、等の理由によりアクセス性能が出ていないと判断し、データを移動するための処理を行う（ステップ 2 7 0 7 ）。

20

【 0 2 9 0 】

予め保持する空き領域情報を参照し、ストレージサブシステム制御部 1 1 2 は、外部ストレージサブシステム 2 1 内の空き領域を抽出する。例えば、外部ストレージサブシステム 2 1 が旧機種で、外部ストレージサブシステム 2 1 内のいくつかの論理記憶装置内のデータをストレージサブシステム 2 0 の論理記憶装置へ移行していた場合は、ストレージサブシステム 2 0 の論理記憶装置へ移行して空きとなっている論理記憶装置を、移行先候補として抽出する（ステップ 2 7 0 8 ）。

30

【 0 2 9 1 】

次に、ストレージサブシステム制御部 1 1 2 は、上述の手法を用いて、移行先の論理記憶装置に対してダミーデータを用いて I / O 処理を行い、I / O 性能を計測する（ステップ 2 7 0 9 ）。

【 0 2 9 2 】

ステップ 2 7 0 9 の計測の結果、ストレージサブシステム制御部 1 1 2 は、I / O 性能が所定の基準を満たしている論理記憶装置（第 3 の論理記憶装置と呼ぶ）を選択し（ステップ 2 7 1 0 ）、第 2 の論理記憶装置のデータを第 3 の論理記憶装置に移行する（ステップ 2 7 1 1 ）。

40

【 0 2 9 3 】

移行の方法は様々で、外部ストレージサブシステム 2 1 にストレージサブシステム 2 0 と同様のホストに透過的に論理記憶装置の移動を行う機能があればそれを使用する。また、外部ストレージサブシステム 2 1 にストレージサブシステム 2 0 と同様のホストに透過的に論理記憶装置の移動を行う機能が無い場合でも、ミラーを作成する機能があれば、外部ストレージサブシステム 2 1 で第 2 の論理記憶装置と第 3 の論理記憶装置とでミラーを作成し、第 1 の論理記憶装置で保持する論理 / 物理対応情報 2 0 3 の P . G . 番号 6 0 5 と汎用ストレージの論理記憶装置を識別する（P o r t アドレス、論理記憶装置番号）情報 6 1 4 との対応を変更する。

【 0 2 9 4 】

50

外部ストレージサブシステム 21 に対して制御が何もできない場合、ストレージサブシステム制御部 112 は、第 2 の論理記憶装置と第 3 の論理記憶装置とに対してデータのリードライト処理を行い、第 2 の論理記憶装置から作成されていた P・G 上にマッピングされていた L D E V を第 3 の論理記憶装置から作成されていた P・G 上にマッピングされるように論理 / 物理対応情報 203 を変更する (ステップ 2711)。

【0295】

以上、説明してきたように、本実施形態では、外部ストレージサブシステム 21 のホスト計算機 1 の I / O を制御するストレージサブシステム 20 において、外部ストレージサブシステム 21 への I / O を制御するために外部ストレージサブシステム 21 の L U をストレージサブシステム 20 の L U にマッピングしてホスト計算機 1 に提供し、その構成情報を保持する。また、外部ストレージサブシステム 21 への I / O 処理に要した時間を計測し、ネットワークの性能や外部ストレージサブシステム 21 の I / O 処理性能を属性情報として保持する。これらの情報を元に、性能チューニングを実行する。

【0296】

このように、本実施形態では、外部ストレージサブシステム 21 も含めたストレージサブシステムの、負荷状況に従って、外部ストレージサブシステム 21 を接続するストレージサブシステム 20 において、外部ストレージサブシステム 21 まで含めた性能チューニングを実現している。

【0297】

本実施形態によれば、外部ストレージサブシステムを複数接続し、その外部ストレージサブシステムに対するホスト計算機からの I / O 処理要求を受領して、その外部ストレージサブシステムに I / O 処理要求を中継する機能を有するストレージサブシステムにおいて、そのストレージサブシステム自身のリソースの性能管理だけでなく、接続する外部ストレージサブシステムも含めたストレージサブシステムの性能情報を管理し、かつ、管理した性能情報を用いて、ホストに透過的に性能チューニングが可能となる。

【0298】

『階層管理記憶機能の導入』

次に、外部ストレージサブシステム 21 を接続する際に、ストレージサブシステム 20 において、階層記憶管理機能を用いて、データの配置を決める方法を述べる。本処理は、図示しない、再配置計画立案処理部によって行われる。

【0299】

再配置計画立案処理部は、外部ストレージサブシステム 21 とホスト計算機 1 との間の運用状況をホスト計算機 1、S A N 管理端末 9、または、外部ストレージサブシステム 21 に接続されているストレージサブシステム管理装置 5 から取得する。

【0300】

運用状況として取得する情報は、ホスト計算機 1 と外部ストレージサブシステム 21 との間の I / O ネットワークの帯域と平均 I / O 処理数と最大 I / O 処理数と平均データ転送量と最大データ転送量とシーケンシャル比率とリードライト比率の履歴である。なるべく長期に渡って取得することが望ましい。

【0301】

再配置計画立案処理部は、取得した情報の I / O 処理量から外部ストレージサブシステム 21 の L U のデータをストレージサブシステム 20 内の記憶装置に移行するほうが良いか、外部ストレージサブシステム 21 内に置き、ストレージサブシステム 20 に接続しストレージサブシステム 20 の L U として提供するか、外部ストレージサブシステム 21 内に置き、従来通りホスト計算機 1 と直接 I / O 処理を行うかを決め、ユーザに推奨案を提示する。

【0302】

ストレージサブシステム 20 と外部ストレージサブシステム 21 との間の第二の I / O 経路 62 の構築はホスト計算機 1 と外部ストレージサブシステム 21 との間の I / O ネットワークの帯域と同等以上に構築する方が望ましいが、それが何らかの理由、例えば、ス

10

20

30

40

50

イチとストレージサブシステム 20 との間の帯域が狭いなどの理由でできない場合でも、最大 I/O 処理数と最大データ転送量の履歴とから I/O ネットワークが処理できる場合は、外部ストレージサブシステム 21 内に置いておいてストレージサブシステム 20 に接続し、ストレージサブシステム 20 の LU として提供することを推奨する。

【0303】

サブシステム管理装置 5 等を介して、ユーザから推奨案に従う旨の指示を受け付けると、再配置計画立案処理部は、示された推奨案に従ってデータのコピーを行う。

【0304】

以上により、ストレージサブシステム 20 に外部ストレージサブシステム 21 を接続する際に、階層記憶管理機能を使ったデータ配置が完了する。

10

【0305】

なお、本実施形態では、ストレージサブシステム 20 が外部ストレージサブシステム 21 の I/O 処理状況を監視し、レスポンス等により、その性能の変化を解析していたが、外部ストレージサブシステム 21 の性能の推移の監視方法はこれに限られない。

【0306】

例えば、ストレージサブシステム 20 のストレージサブシステム制御装置 112 が、第二の I/O 経路 62 を介して外部ストレージサブシステム 21 の I/O 稼動情報と構成情報とを取得できる場合、外部ストレージサブシステム 21 に対して、前記 I/O 稼動情報と構成情報との取得要求コマンドを送信し、前記 I/O 稼動情報と構成情報を取得する。

【0307】

20

この場合、外部ストレージサブシステム 21 の論理記憶装置 8 と物理記憶装置 124 との対応情報、論理記憶装置稼動情報、ポート稼動情報などが、ストレージサブシステム単位の論理/物理対応情報 203、論理記憶装置稼動情報 224、物理記憶装置稼動情報 225 などとしてストレージサブシステム制御装置 112 に保持される。

【0308】

構成変更計画立案処理部 218 は、上記情報に従い、外部ストレージサブシステム 21 の性能劣化を判別し、性能チューニングのための構成変更計画を立案する。

【0309】

<< 第二の実施形態 >>

次に、第二の I/O 経路 62 を介して、ストレージサブシステム 20 から外部ストレージサブシステム 21 へ性能情報解析のための I/O 要求を発行し、外部ストレージサブシステム 21 の性能情報を取得する技術を、第二の実施形態として説明する。

30

【0310】

本実施形態のシステム構成は、基本的に第一の実施形態と同様である。本実施形態のストレージサブシステム 20 は、第一の実施形態の構成に加え、性能情報解析のための I/O 要求としてダミーデータを生成し、送出するダミーデータ生成送出機能を備える。以下において、第一の実施形態と異なる構成についてのみ、主として説明する。

【0311】

例えば、特定のデータサイズのデータを連続してある特定のアドレスに対して送信することで、外部ストレージサブシステム 21 の I/O 処理性能が計測できる。

40

【0312】

ダミーデータ生成送出機能による外部ストレージサブシステム 21 の I/O 処理性能の予測の手順について説明する。

【0313】

ダミーデータ生成送出機能で生成されるダミーデータは、特定のデータサイズの 1 つ以上のデータであり、生成されると、第二の I/O 経路 62 のプロトコルに応じた形式で送出される。

【0314】

図 27 に、ダミーデータ生成送出機能によるストレージサブシステム 20 から外部ストレージサブシステム 21 への I/O 処理性能の測定時の処理のイメージを示す。また、図

50

28に、ダミーデータ生成送出機能が送出したダミーデータによる性能測定結果3000の一例を示す。

【0315】

サブシステム管理装置5もしくはSAN管理端末9は、管理者から受け付けた入力に従い、ダミーデータ生成送出機能に対して、外部ストレージサブシステム21のI/Oポートのアドレス、論理記憶装置番号、ダミーデータ長、ダミーデータの目標IOPS、前記論理記憶装置のI/O対象アドレスの偏り、リード処理かライト処理かを指定して計測指示を出す。ダミーデータ生成送出機能は、上記計測指示に従って、ダミーデータを生成し、指示された外部ストレージサブシステムのI/Oポートに当該データを送出する。

【0316】

性能測定結果3000は、図28に示すように、送出したダミーデータのサイズごとに、目標IOPSと、ローカリティと、リードライト比率と、実測IOPSと、実測MBPSと、レスポンスと、を備える。

【0317】

目標IOPSは1秒間に出すI/O処理コマンド数の目標値であり、外部ストレージサブシステム21もしくは第二のI/O経路62が当該目標値を十分に出す処理性能または帯域があれば、ほぼ目標値通りにストレージサブシステム20はコマンドを発行できる。この目標値を上げていくことで、ストレージサブシステム20と第2のストレージサブシステム間のI/O処理性能が分かる。

【0318】

また、ダミーデータ長を変えることで、ランダムアクセスやシーケンシャルアクセスに対するストレージサブシステム20と外部ストレージサブシステム21との間のI/O処理性能が分かる。

【0319】

I/O対象アドレスの偏りを指定することで、たとえば偏りが大きければ外部ストレージサブシステム21がディスクキャッシュ108にヒットする確率が高くなるため、キャッシュヒット性能とキャッシュミス性能との違いが、ストレージサブシステム20と外部ストレージサブシステム21との間のI/O処理性能として分かる。

【0320】

さらに、ある小さな一定のアドレス範囲にI/O処理を行っても、偏りを全くなくしても、処理性能に違いが無ければ、ディスクキャッシュ108が無い、または、その容量がきわめて小さい、などの障害が潜んでいる可能性があると考えられる。

【0321】

ストレージサブシステム制御装置112は、計測後、計測指示を送信してきた外部ストレージサブシステム21に対し、終了報告を送信する。終了報告を受信すると、ストレージサブシステム管理装置5もしくはSAN管理装置9は、表示部に終了報告を表示させ、管理者からの指示を待つ。計測結果を読み出す指示を受け付けると、これらの装置は、ストレージサブシステム制御装置112から計測結果情報をネットワーク7経由で読み出し、表示部に表示する。管理者は、この表示を見ることにより、第二のI/O経路62を介した外部ストレージサブシステム21へのI/O処理性能を知ることができる。

【0322】

なお、本ダミーデータ生成送出機能は、稼動中にストレージサブシステム20から外部ストレージサブシステム21のI/O処理性能を得る目的以外に、外部ストレージサブシステム21を接続した情報処理システムの設計時にも用いることができる。設計時には、外部ストレージサブシステム21のI/O処理性能を見積もる必要がある。

【0323】

新規にストレージサブシステム20を導入するにあたり、導入コストを抑えるため、従来用いていたストレージサブシステムから全てのデータを移行せず、一部を従来のストレージサブシステムに残す場合がある。

【0324】

10

20

30

40

50

このような場合に、残された従来のストレージサブシステムを、上記の実施形態でいう、外部ストレージサブシステム 2 1 として情報処理システムを設計する。しかし、この場合、外部ストレージサブシステム 2 1 として、ホスト計算機 1 から、ストレージサブシステム 2 0 および第二の I / O 経路 6 2 を介してアクセスすることになるため、従来からのホスト計算機 1 との I / O 処理性能は、そのまま採用できない。

【 0 3 2 5 】

そこで、本実施形態では、外部ストレージサブシステム 2 1 の I / O 処理性能の見積もりを、ストレージサブシステム 2 0 から外部ストレージサブシステム 2 1 へダミーデータを送出することによって行なう。

【 0 3 2 6 】

例えば、特定のデータサイズのデータを連続してある特定のアドレスに対して送信することで、外部ストレージサブシステム 2 1 の I / O 処理性能が計測でき、これにより、ストレージサブシステム 2 0 を介したホスト計算機 1 との I / O 処理性能を見積もることができる。

【 0 3 2 7 】

ストレージサブシステム 2 0 の上記ダミーデータ生成および送出機能による、外部ストレージサブシステム 2 1 の性能を見積もる際の具体的な手順は、以下のとおりである。

【 0 3 2 8 】

ストレージサブシステム 2 0 を新規に導入する前に、外部ストレージサブシステム 2 1 となるストレージサブシステムとホスト計算機 1 との I / O 処理性能は予め取得されているものとする。まず、ストレージサブシステム 2 0 のストレージサブシステム制御装置 1 1 2 は、第二の I / O 経路 6 2 を介してアクセス可能な外部ストレージサブシステム 2 1 のポートからアクセス可能な論理記憶装置を探し出す。見つけた論理記憶装置に対し、ダミーデータ生成および送出機能は、以下の手順でダミーデータを送出することにより、第二の I / O 経路 6 2 を介した外部ストレージサブシステム 2 1 の I / O 処理性能の測定を行う。

【 0 3 2 9 】

外部ストレージサブシステム 2 1 の I / O 処理性能が、所望の性能を満たしていた場合、ストレージサブシステム制御装置 1 1 2 は、第二の I / O 経路 6 2 上にある外部ストレージサブシステム 2 1 の論理記憶装置をストレージサブシステム 2 0 の論理記憶装置として論理 / 物理対応情報 2 0 3 に定義する。

【 0 3 3 0 】

< < 第三の実施形態 > >

『データのアクセス頻度に従った性能チューニング』

次に、データへのアクセス頻度に従い、外部ストレージサブシステム 2 1 を含めたシステムの性能チューニングを実現する実施形態を第三の実施形態として説明する。以下、本実施形態について、第一の実施形態と異なる構成を主として説明する。

【 0 3 3 1 】

図 2 9 に、ストレージサブシステム 2 0 が N A S (N e t w o r k A t t a c h e d S t o r a g e) として機能する I / F を備える場合のストレージサブシステム 2 0 および外部ストレージサブシステム 2 1 の機能構成を示す。

【 0 3 3 2 】

N A S はネットワークに直結してファイル共有サービスなどをネットワーク・クライアントに提供するストレージ機器であり、ネットワークを介してファイルを共用できる独立したファイルサーバとして機能することを特徴としている。

【 0 3 3 3 】

N A S を導入することにより、ファイルサーバと記憶装置とを一台の装置で管理できるため、管理対象が減り、ファイルサーバと記憶装置との 2 台の装置をそれぞれ別個に管理する場合と比べ、管理コストを低く抑えられるというメリットがある。

【 0 3 3 4 】

10

20

30

40

50

本実施形態は、基本的には、第一の実施形態と同様の機能構成を有する。以下に説明されていない構成は、第一の実施形態と基本的に同様である。

【0335】

さらに、ストレージサブシステム制御装置112は、NASを実現するネットワークファイルシステム制御部2401を備え、ネットワーク7を経由して、ホスト計算機1やホスト計算機1がサービスを行う情報処理システムのクライアント計算機と接続されている。クライアント計算機は、ネットワークファイルシステム制御部2401が提供するNASを介して、ストレージサブシステム20にアクセスできる。

【0336】

また、ネットワークファイルシステム制御部2401は、ポートやアダプタなどを制御するネットワークファイルI/O処理部2402と、ファイル処理を行うファイルシステム処理部2403と、ファイルの再配置を計画し、実行する再配置処理部2404と、ファイルの管理情報を格納するファイル管理情報2410とを備える。本処理部は、第一の実施形態の構成変更計画立案処理部218と構成変更計画実行処理部219とに代わるものである。

【0337】

ストレージサブシステム制御装置112は、NASを実現するネットワークファイルシステム制御部2401を持つことにより、後述するように、ファイル単位に、ファイルを作成した日時、最新のアクセス時間、アクセス頻度を管理することが可能となる。

【0338】

まず、ネットワークファイルシステム制御部2401による、ファイルとそれを格納する記憶装置のアドレスとの管理を、図を用いて説明する。

【0339】

図30は、ネットワークファイルシステム制御部2401によるファイル管理のイメージ図である。

【0340】

ネットワークファイルシステム制御部2401は、ストレージサブシステム制御装置112から論理記憶装置もしくは内部論理記憶装置の提供を受け、これらをボリュームとして管理する。

【0341】

図30に示すように、本実施形態におけるファイルシステムは、論理記憶装置2501をいくつかの区画2502に区切り、ファイルの管理を容易にし、障害の局所化を行うことができるよう構成される。

【0342】

ネットワークファイルシステム制御部2401は、区画2502に、ブートブロック2503、スーパーブロック2504、シリンダブロック2505、iノードリスト2506、データ領域2507を作成する。

【0343】

スーパーブロック2504には、区画2502の管理情報が格納され、スーパーブロック2504内にあるファイルはiノードで管理される。

【0344】

iノードは、iノードリスト2512として保持され、管理される。また、各iノードはiノード番号2511によって指定される。各iノードの内容はディレクトリ情報やファイル情報である。

【0345】

iノードがディレクトリ情報である場合、本図に示すように、そのディレクトリにあるディレクトリおよびファイルがエントリされ、例えばルートディレクトリ2513にはdirAというディレクトリがエントリされておりdirAのiノード番号は4であることがわかる。

【0346】

10

20

30

40

50

階層的なアクセス方法により、例えば、ディレクトリ `dir A / sub dir B / 2 5 1 4` にファイル `File C` があり、その `i` ノード番号が 8 であることがわかる。そして、`i` ノード番号 8 で指定される `File C` の `i` ノードは、そのファイルの所有者 2 5 1 5、所有者のグループ名 2 5 1 6、ファイルの種類 2 5 1 7（例えば、テキストファイル、バイナリファイル等）、最終アクセス時刻 2 5 1 8、最終更新時刻 2 5 1 9、`i` ノード採取更新時刻 2 5 2 0、ファイルのサイズ 2 5 2 1、ディスクのアドレス情報 2 5 2 2、等を保持している。

【 0 3 4 7 】

ディスクのアドレス情報 2 5 2 2 として、ファイル 2 5 2 6 が格納されているディレクトリ 2 5 2 4 とその中の位置 2 5 2 5 とを保持する。また、ファイル 2 5 2 6 が配置されているディスクとブロックと次の読み出しデータのブロックのアドレスとを保持する。次の読み出しデータのブロックのアドレスは、そのファイルが複数のデータブロック 2 5 2 7 に分散されて配置されていても読み出せるようにするために保持されているものである。

【 0 3 4 8 】

ネットワークファイルシステム制御部 2 4 0 1 は、ファイル管理情報 2 4 1 0 に、ファイル管理情報 2 4 1 0 を備える。図 3 1 に、ファイル管理情報 2 4 1 0 の一例を示す。

【 0 3 4 9 】

本図に示すように、ファイル管理情報 2 4 1 0 は、各ファイルごとに、それぞれのファイル名 2 4 1 1、インデックス 2 4 1 2、ファイルサイズ 2 4 1 3、ファイルの種類 2 4 1 4、作成時間 2 4 1 5、最終アクセス時間 2 4 1 6、最終更新時間 2 4 1 7、一定時間内のアクセス頻度 2 4 1 8、設定が可能であればファイルの重要度 2 4 1 9 などの情報を保持する。

【 0 3 5 0 】

ファイル管理情報 2 4 1 0 を用いることにより、再配置処理部 2 4 0 4 は、作成日時からの時間経過に応じて、ファイル移動の要否を判断し、必要に応じて性能チューニングを行う。

【 0 3 5 1 】

次に、再配置処理部 2 4 0 4 による、アクセス頻度に応じて構成を変更する、本実施形態の一連の処理を図に従って説明する。図 3 2 は、再配置処理部 2 4 0 4 による処理フローである。

【 0 3 5 2 】

再配置処理部 2 4 0 4 は、ファイル管理情報 2 4 1 0 を参照し（ステップ 2 6 0 1）、ファイルの最新参照日時 2 4 1 6 の項目でソートし（ステップ 2 6 0 2）、最新参照日時が予め定めた時間以上経っているものがあるか否かを判断する（ステップ 2 6 0 3）。

【 0 3 5 3 】

ステップ 2 6 0 3 において、再配置処理部 2 4 0 4 は、予め定めた時間以上経っているものがあれば、該当するファイルを抽出し（ステップ 2 6 1 0）、移動対象ファイルとして管理する。

【 0 3 5 4 】

ステップ 2 6 0 3 において、予め定めた時間以上経っているものがなければ、再配置処理部 2 4 0 4 は、ファイル管理情報 2 4 1 0 を参照頻度でソートし（ステップ 2 6 0 4）、ファイル参照頻度が 0 のものがあるかを判断する（ステップ 2 6 0 5）。

【 0 3 5 5 】

ステップ 2 6 0 5 において、ファイル参照頻度 0 のものがあれば、再配置処理部 2 4 0 4 は、該当するファイルを抽出して（ステップ 2 6 1 1）、移動対象ファイルとして管理する。

【 0 3 5 6 】

ステップ 2 6 0 5 において、ファイル参照頻度 0 のものがなければ、再配置処理部 2 4 0 4 は、ファイルの参照頻度が予め定めた値未満のものがあるかを判断する（ステップ 2

10

20

30

40

50

606)。

【0357】

ステップ2606において、所定値未満のものがあれば、再配置処理部2404は、該当するファイルを抽出し(ステップ2607)、抽出したファイルをファイル管理情報2410を作成日時でソートし(ステップ2608)、抽出したファイルに作成日時から予め定めた時間以上たっているファイルがあるかどうか判断する(ステップ2609)。

【0358】

ステップ2609において、所定時間以上経たファイルがあれば、再配置処理部2404は、該当するファイルを抽出し(ステップ2612)、移動対象ファイルとして管理する。

10

【0359】

その後、再配置処理部2404は、外部ストレージサブシステム21の論理記憶装置に、抽出した移動対象全ファイルを移動する(ステップ2613)。移動した後、再配置処理部2404は、iノードを書き換え(ステップ2614)、処理を終了する。

【0360】

なお、ステップ2606、および、ステップ2809において、条件を満たすファイルがない場合は、処理を終了する。

【0361】

以上、アクセス頻度に応じて、構成を変更する手順を説明した。

【0362】

20

本実施形態によれば、例えば、作成日時から1週間、または、1ヶ月アクセスが無いファイルは外部ストレージサブシステム21の論理記憶装置上にデータ領域を移動し、頻繁にアクセスするファイルはストレージサブシステム20の物理記憶装置110にデータを格納するといった性能チューニングが可能となる。

【0363】

また、参照頻度2418を直接参照し、この値が予め定めた値より低いファイルを外部ストレージサブシステム21の論理記憶装置に移動し、参照頻度が上がった場合は、ストレージサブシステム20の物理記憶装置110にファイルの格納箇所を移動するといった性能チューニングも可能である。

【0364】

30

本実施形態は、ウェブのデータ等のように作成された直後は参照頻度が高く、10日もすると殆ど参照されないといったデータを格納するストレージサブシステムに対し、有効である。

【0365】

なお、第一の実施形態では、マネージャ223の指示に従い、構成変更計画立案処理部218が立案した変更計画を、利用者に提示し、指示を受け付けてから構成変更計画実行処理部219が実行していた。本実施形態においても、上記の手順において、ステップ2613を実行する前に、利用者に提示し、許可を得るよう構成してもよい。

【0366】

このように、本実施形態によれば、外部ストレージサブシステムを接続したストレージサブシステムにおいて、性能チューニングを行なうにあたり、各ファイルのアクセス頻度に応じて最適にファイルを配置できる。このため、さらに、ストレージサブシステム全体としての性能を高めることができる。

40

【0367】

第一の実施形態および第二の実施形態では負荷を基準に、第三の実施形態では、アクセス頻度を基準に、性能チューニングの要否を判断し、実行している。しかしながら、性能チューニングの判断基準はこれらに限られない。例えば、ストレージサブシステム自体の障害兆候を予め検知して、データの移行を行なうといった性能チューニングも可能である。

【0368】

50

例えば、規制によって、企業では、帳簿やメール、患者のカルテや新薬開発におけるデータ等を予め定められた期間保持しなければならない場合がある。しかも、このようなデータは、要求に応じて所定の時間内に提示できる必要がある。従って、このようなデータを扱うシステムはそのような要求を満たす必要がある。

【0369】

ここで、上記システムが、第一の実施形態と同様に、第一のI/O経路61を介しホスト計算機1とのI/O処理を行い、第二のI/O経路62を介して外部ストレージサブシステム21とI/O処理を行うストレージサブシステムであって、外部ストレージサブシステム21が旧機種でありストレージサブシステム20が新機種である場合を考える。

【0370】

通常業務のデータはストレージサブシステム20に格納し、前記の規制で保存を義務付けられている帳簿、メール、患者のカルテ、新薬開発におけるデータ等は、外部ストレージサブシステム21に格納すれば、外部ストレージサブシステム21である旧機種内のデータをすべて新機種に移行して廃棄するより資産の有効活用になる可能性がある。しかし、旧機種は新機種に比べ使用時間が長い分装置寿命が早く訪れてしまう可能性が高い。

【0371】

このような場合、ストレージサブシステム20から、外部ストレージサブシステム21に対し、予め定められた一定の間隔でI/O処理要求を出し、外部ストレージサブシステム21の処理性能を計測することにより、外部ストレージサブシステム21に内在する障害兆候を検出して、他のストレージサブシステムへデータを移行する。

【0372】

ここで、移行する他のストレージサブシステムはストレージサブシステム20であってもよいし、ストレージサブシステム20でも外部ストレージサブシステム21でもどちらでもない、第3のストレージサブシステムであってもよい。

【0373】

なお、移行先のストレージサブシステムを、例えば、外部ストレージサブシステム21と同様に、ストレージサブシステム20に比べて旧機種であるもの、または、導入コストが比較的安価なものとすると、さほどアクセス頻度の高くないデータを格納する記憶装置のコストを抑えることができる。

【図面の簡単な説明】

【0374】

【図1】図1は、第一の実施形態の計算機システム全体の構成図である。

【図2】図2は、第一の実施形態のストレージサブシステムと外部ストレージサブシステムとの機能構成および接続状態を説明するための図である。

【図3】図3は、第一の実施形態のストレージサブシステム制御部112の機能を説明するための図である。

【図4】図4は、第一の実施形態の論理物理対応情報の一例を示す図である。

【図5】図5は、第一の実施形態の論理記憶装置稼動情報の一例を示す図である。

【図6】図6は、第一の実施形態の物理記憶装置属性情報の一例を示す図である。

【図7】図7は、第一の実施形態の論理記憶装置稼動情報の一例を示す図である。

【図8】図8は、第一の実施形態の論理記憶装置属性情報の一例を示す図である。

【図9】図9は、第一の実施形態の物理記憶装置稼動情報の一例を示す図である。

【図10】図10は、第一の実施形態の外部ストレージ稼動情報を説明するための図である。

【図11】図11は、第一の実施形態のキャッシュ量カウンタを説明するための図である。

【図12】図12は、第一の実施形態のポート稼動情報の一例を説明するための図である。

【図13】図13は、第一の実施形態のポート設定情報の一例を示す図である。

【図14】図14は、第一の実施形態のプロセッサ稼動情報の一例を示す図である。

10

20

30

40

50

【図 15】図 15 は、第一の実施形態のサブシステム管理装置のハードウェア構成の一例を示す図である。

【図 16】図 16 は、第一の実施形態のホスト計算機のハードウェア構成の一例を示す図である。

【図 17】図 17 は、第一の実施形態の S A N 管理端末のハードウェア構成の一例を示す図である。

【図 18】図 18 は、第一の実施形態の外部ストレージサブシステムへのリード要求があった場合の処理を説明するための図である。

【図 19】図 19 は、第一の実施形態の外部ストレージサブシステムへのライト要求があった場合の処理を説明するための図である。

【図 20】図 20 は、第一の実施形態のストレージサブシステムの性能チューニング時の処理フローである。

【図 21】図 21 は、第一の実施形態の構成変更計画立案処理部の処理フローである

【図 22】図 22 は、第一の実施形態の第 1 の論理記憶装置から第 2 の論理記憶装置にデータをコピーする場合の処理を説明するための図である。

【図 23】図 23 は、第一の実施形態の構成変更実行処理部の処理フローである。

【図 24】図 24 は、第一の実施形態の構成変更実行処理部の処理フローである。

【図 25】図 25 は、第一の実施形態の外部ストレージサブシステム間での移動の手順を説明するための図である。

【図 26】図 26 は、第一の実施形態の外部ストレージサブシステム内でのデータ移動時の処理フローである。

【図 27】図 27 は、第二の実施形態のストレージサブシステムから外部ストレージサブシステムへの I / O 処理性能の測定処理を説明するための図である。

【図 28】図 28 は、第二の実施形態の送出したダミーデータによる性能測定結果の一例を説明するための図である。

【図 29】図 29 は、第三の実施形態のストレージサブシステムおよび外部ストレージサブシステムの機能構成図である。

【図 30】図 30 は、第三の実施形態のネットワークファイルシステム制御部によるファイル管理のイメージ図である。

【図 31】図 31 は、第三の実施形態のファイル管理情報の一例を示す図である。

【図 32】図 32 は、第三の実施形態の再配置処理部の処理フローである。

【符号の説明】

【 0 3 7 5 】

1・・・ホスト計算機、1 a・・・ホスト計算機、1 n・・・ホスト計算機、2 0 n・・・ストレージサブシステム、2 0 a・・・ストレージサブシステム、2 1・・・外部ストレージサブシステム、5 n・・・サブシステム管理装置、5 a・・・サブシステム管理装置、5 b・・・サブシステム管理装置、6 1・・・第一の I / O 経路、6 2・・・第二の I / O 経路、7・・・ネットワーク、8・・・論理記憶装置、1 0 0・・・制御用プロセッサ、1 0 1・・・ローカルメモリ、1 0 3・・・ネットワークアダプタ、1 0 4 a・・・ポート、1 0 4 b・・・ポート、1 0 5・・・ポート、1 0 6・・・ディスクアダプタ、1 0 7・・・共有メモリ、1 0 8・・・ディスクキャッシュ、1 0 9・・・内部ネットワーク、1 1 0・・・記憶装置、1 1 2・・・ストレージサブシステム制御装置、1 2 0・・・ストレージサブシステム制御装置、1 2 1・・・ポート、1 2 2・・・ポート、1 2 3・・・I / O アダプタ、1 2 4・・・物理記憶装置、1 3 0・・・I / O ネットワークスイッチ

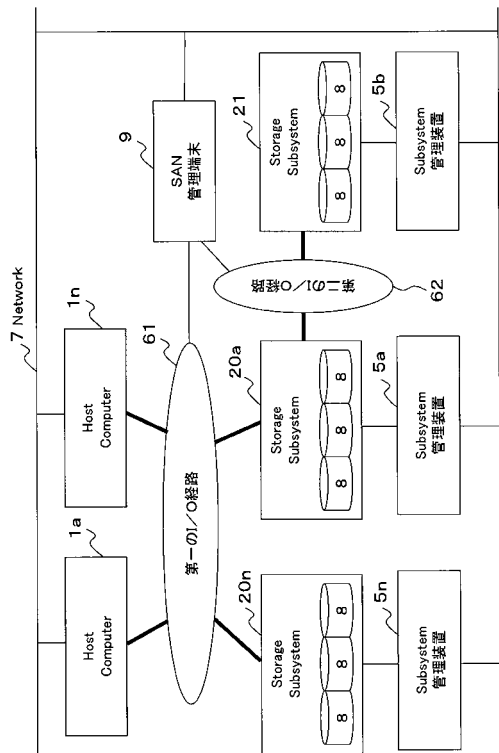
10

20

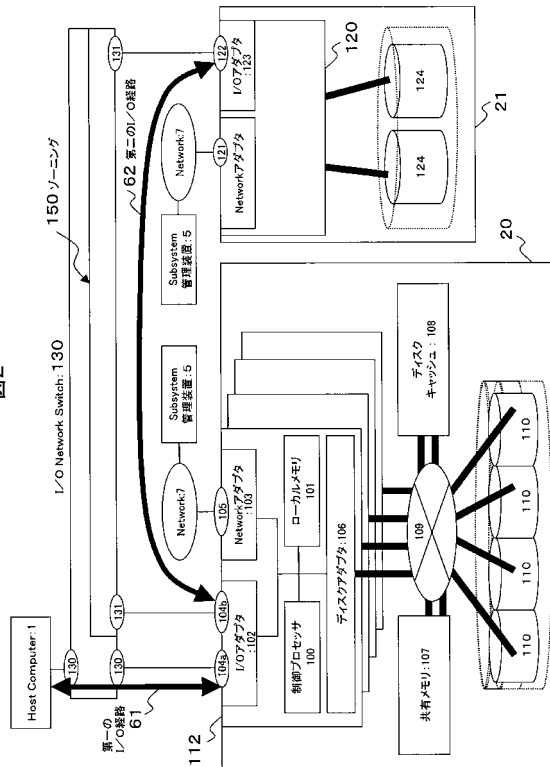
30

40

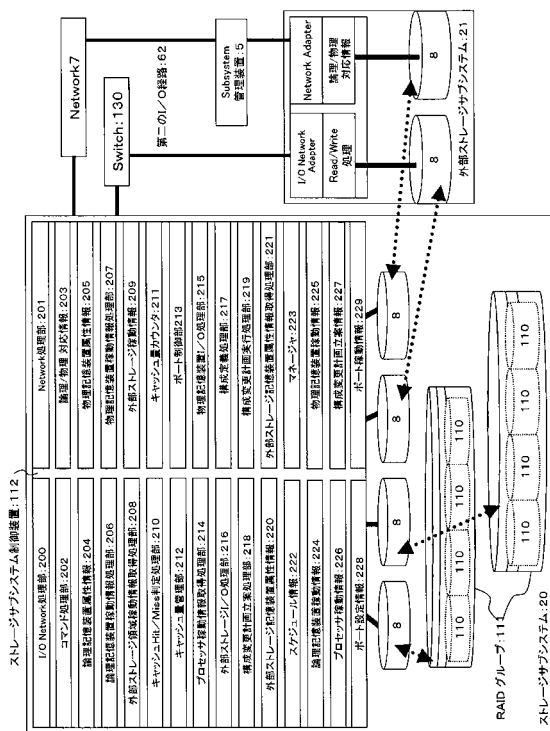
【 図 1 】



【圖 2】



【 図 3 】



【圖 4】

[illegible]

【 図 9 】

[illegible]

【 図 1 0 】

時間	イニシエータ ポート番号	汎用ストレージ サブシステムWWN	汎用ストレージ LUN	IOPS	MBPS	レスポンス	リンク状態
1801	1901	1902	1903	1904	1905	1906	1907
YYY. MM.D D.hh. mm. 00	10						
	11						
	...						
YYY. MM.D D.hh. mm. 00	10						
	11						
	...						
...	...	...	...	...	...	...	...

【 図 1 1 】


11

211

時間 1801		クリーンカウンタ 1804	ダーティカウンタ 1805	フリーカウンタ 1806
	論理記憶 装置番号			
YYYY. MM. DD. hh. mm. 00	0000			****
	0001			****
	****	****	****	****
	全体			
YYYY. MM. DD. hh. mm'. 00	0000			****
	0001			****
	****	****	****	****
	全体			
****	****	****	****	****

【 図 1 2 】

[illegible]

【図 13】

228

図13

ポート番号 1400	状態 1401	接続WWN 1402	最大 IOPS 1403	最大 MBPS 1404	最小 IOPS 1405	最小 MBPS 1406	目標 IOPS 1407	目標 MBPS 1408	帯域 1409	プロトコル 1410
00	ターゲット	ホストWWN	100	50	10	10	100	50	2G	FC
01	ターゲット	ホストWWN	100	50	10	10	50	25	2G	FC
02	ターゲット	ホストWWN	100	50	10	10	100	50	2G	iSCSI
...	...	...	...	...	...	...	...	...	...	...
10	イニシエータ	利用ストレージ WWN	100	50	10	10	25	...	2G	FC
11	イニシエータ	利用ストレージ WWN	100	50	10	10	50	...	2G	FC
...	...	...	...	...	...	...	...	...	...	...

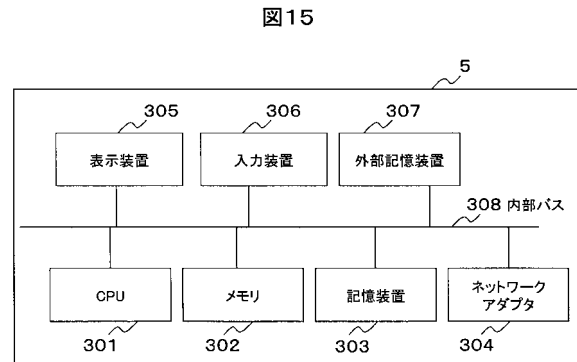
【図 14】

226

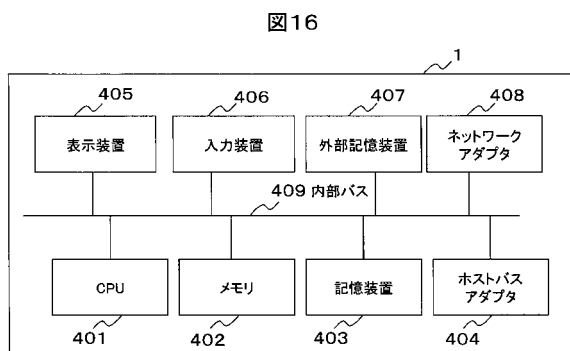
図14

プロセッサ番号 1451	状態 1452	稼働率 1453
00	ターゲット	...
01	ターゲット	...
02	ターゲット	...
...	...	...
10	イニシエータ	...
11	イニシエータ	...
...	...	...

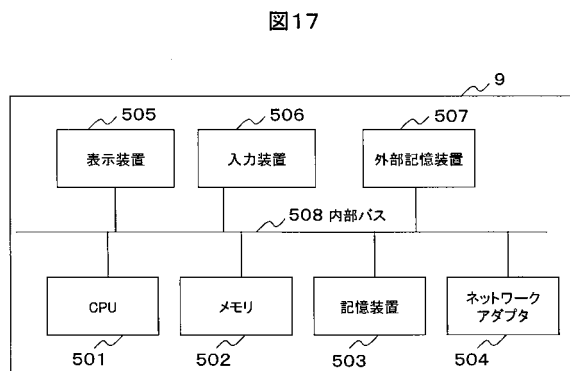
【図 15】



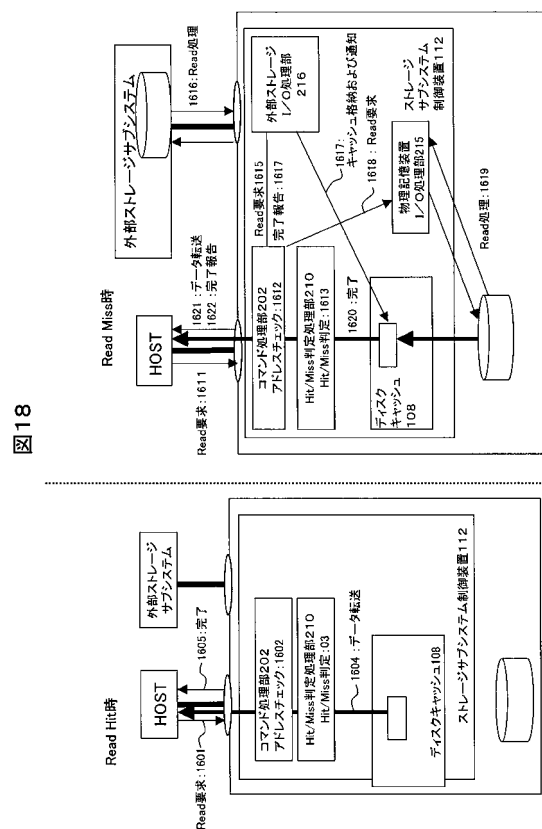
【図 16】



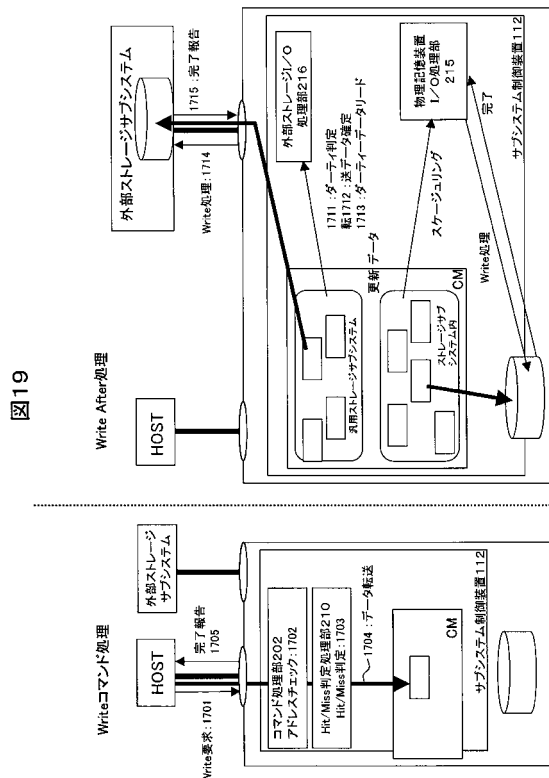
【図 17】



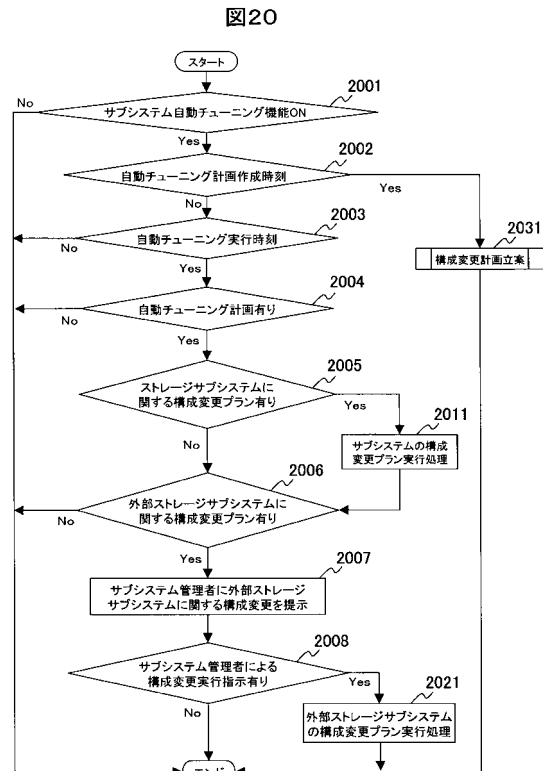
【図 18】



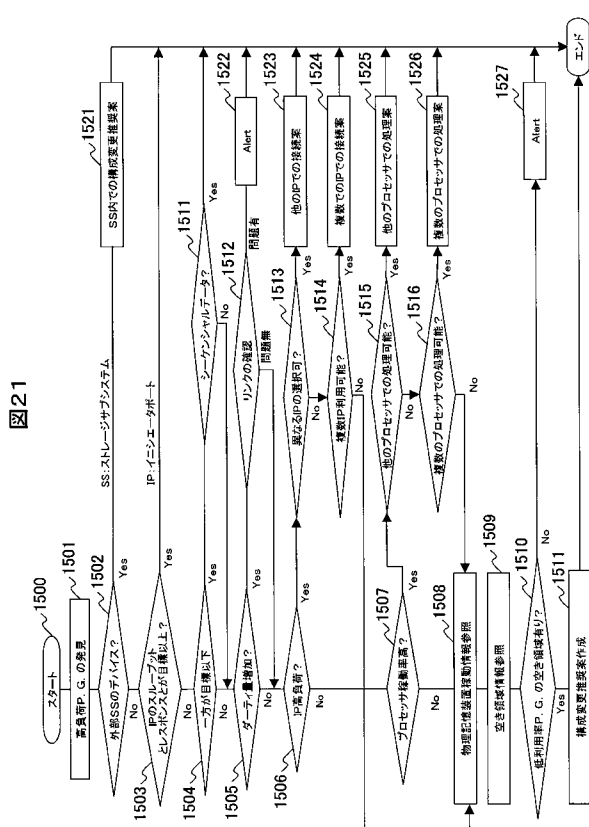
【 図 1 9 】



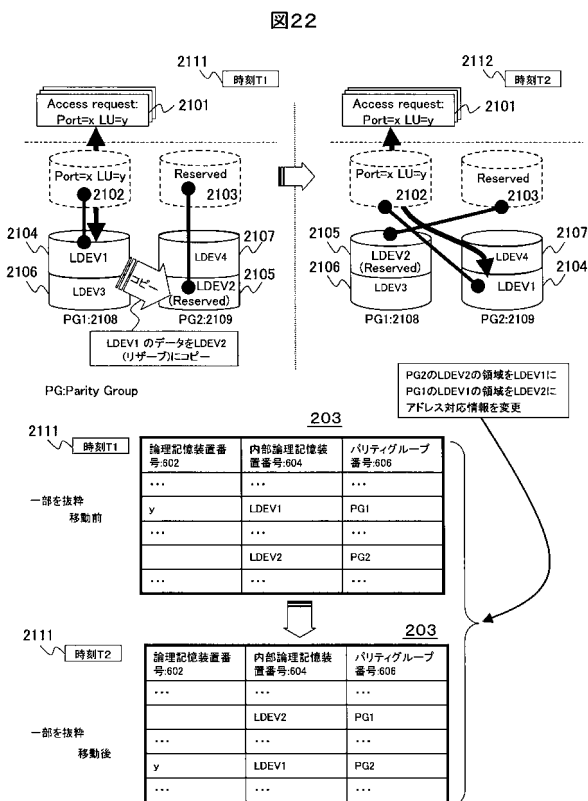
【 図 2 0 】



【 図 2 1 】

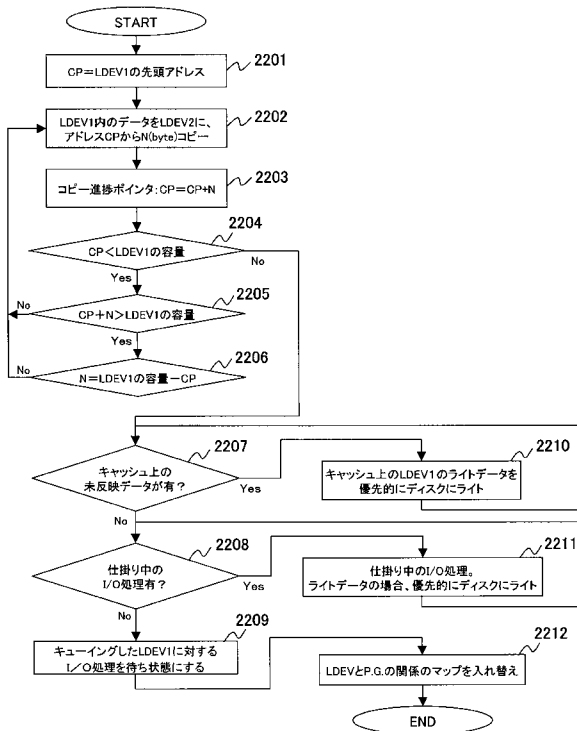


【 図 2 2 】



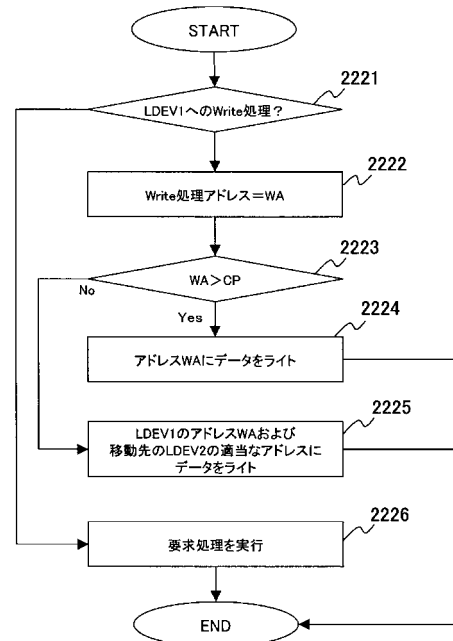
【図23】

図23



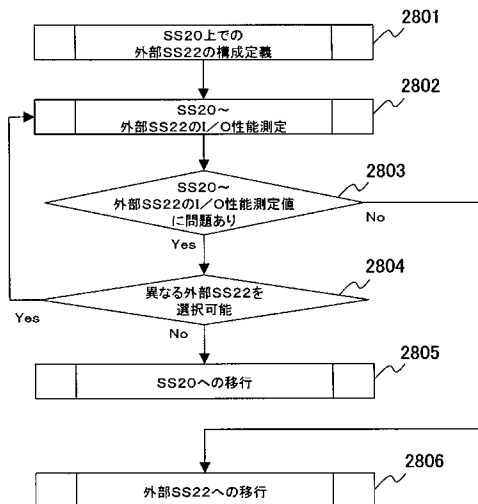
【図24】

図24



【図25】

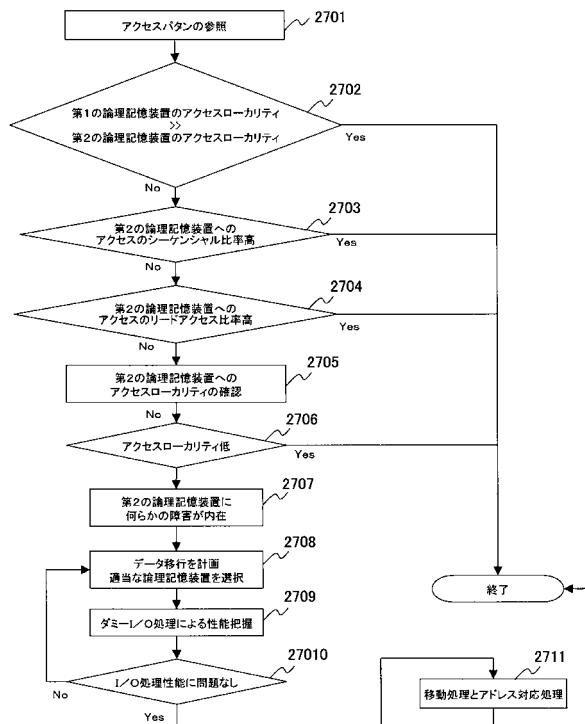
図25



SS:ストレージサブシステム

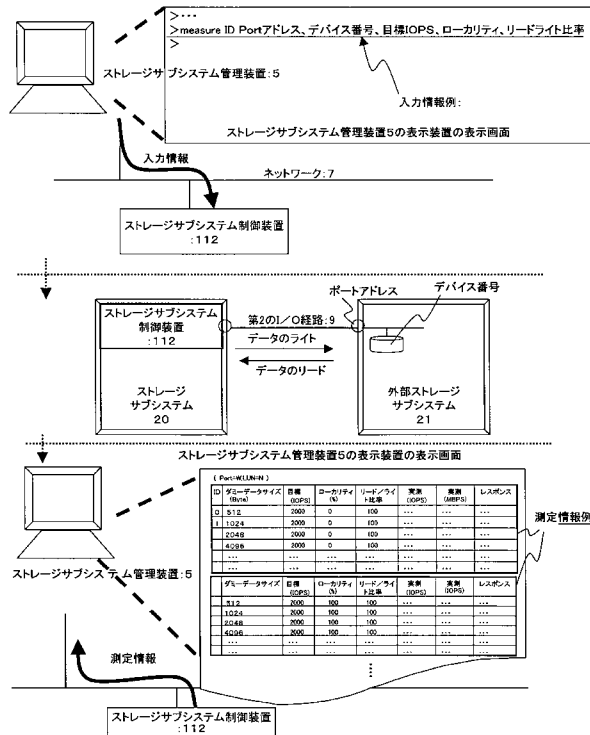
【図26】

図26



【 図 2 7 】

図27



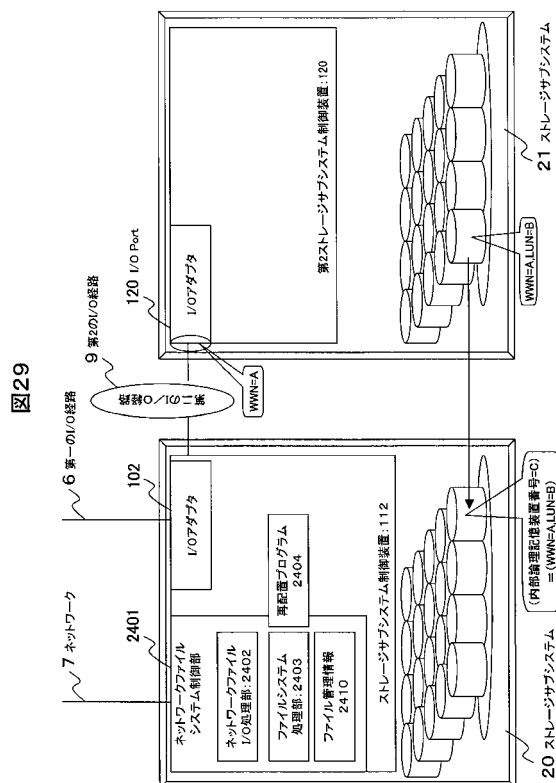
【 図 2 8 】

图28

3000 (Port=W:LUN=N)							
ID	ゲームデータサイズ (Byte)	目標 (IOPS)	ローカルリティ (%)	リード/ライト 比率	実測(IOPS)	実測(MBPS)	レスポンス
0	512	2000	0	100	***	***	***
1	1024	2000	0	100	***	***	***
	2048	2000	0	100	***	***	***
	4096	2000	0	100	***	***	***
	***	***	***	***	***	***	***
	***	***	***	***	***	***	***

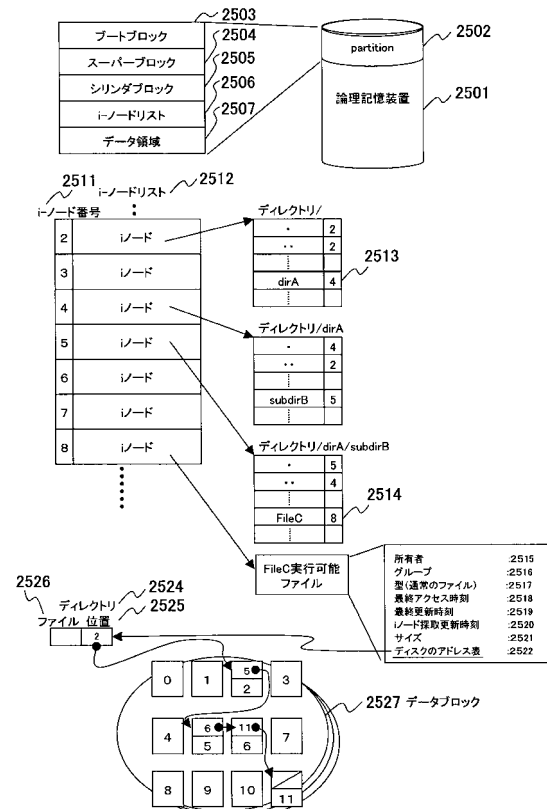
ゲームデータサイズ (Byte)	目標 (IOPS)	ローカルリティ (%)	リード/ライト 比率	実測(IOPS)	実測(IOPS)	レスポンス
512	2000	100	100	***	***	***
1024	2000	100	100	***	***	***
2048	2000	100	100	***	***	***
4096	2000	100	100	***	***	***
***	***	***	***	***	***	***
***	***	***	***	***	***	***

【 図 2 9 】



【 図 3 0 】

図30



【図 3 1】

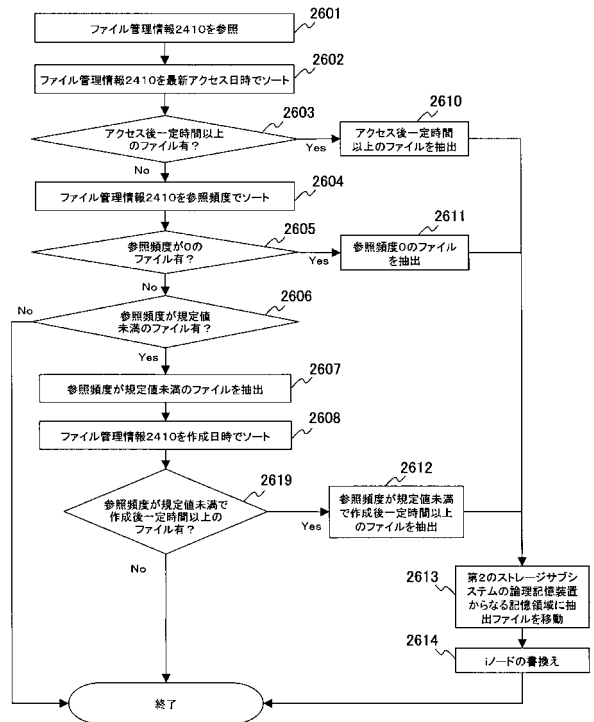
図31

2410

2411	2412	2413	2414	2415	2516	2417	2418	2419
ファイル名	ファイルID	ファイルサイズ	ファイルの種類	作成時間	最終アクセス時刻	最終更新時刻	アクセス頻度	重要度

【図 3 2】

図32



フロントページの続き

(51)Int.Cl.

F I

G 0 6 F 3/06 5 4 0

G 0 6 F 13/10 3 4 0 A

(72)発明者 里山 愛

神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

(72)発明者 島田 朗伸

神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I D システム事業部内

合議体

審判長 清水 稔

審判官 甲斐 哲雄

審判官 山田 正文

(56)参考文献 特開 2 0 0 3 - 2 1 6 3 4 8 (J P , A)

特開 2 0 0 1 - 1 2 5 8 7 9 (J P , A)

特開平 7 - 1 2 9 3 3 1 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G06F 12/00, G06F 3/06, G06F 13/10