

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7618023号  
(P7618023)

(45)発行日 令和7年1月20日(2025.1.20)

(24)登録日 令和7年1月9日(2025.1.9)

|                   |                               |          |                               |
|-------------------|-------------------------------|----------|-------------------------------|
| (51)国際特許分類        |                               | F I      |                               |
| G 0 6 F           | 16/11 (2019.01)               | G 0 6 F  | 16/11                         |
| G 0 6 F           | 11/20 (2006.01)               | G 0 6 F  | 11/20 6 5 6                   |
| G 0 6 F           | 3/06 (2006.01)                | G 0 6 F  | 3/06 3 0 4 F                  |
| G 0 6 F           | 13/10 (2006.01)               | G 0 6 F  | 3/06 3 0 1 X                  |
|                   |                               | G 0 6 F  | 13/10 3 4 0 A                 |
| 請求項の数 22 (全29頁)   |                               |          |                               |
| (21)出願番号          | 特願2023-516240(P2023-516240)   | (73)特許権者 | 503433420                     |
| (86)(22)出願日       | 令和3年9月11日(2021.9.11)          |          | 華為技術有限公司                      |
| (65)公表番号          | 特表2023-541069(P2023-541069 A) |          | HUAWEI TECHNOLOGIES CO., LTD. |
| (43)公表日           | 令和5年9月27日(2023.9.27)          |          | 中華人民共和國 5 1 8 1 2 9 広東省深      |
| (86)国際出願番号        | PCT/CN2021/117843             |          | ▲チェン▼市龍崗区坂田 華為總部▲ベ            |
| (87)国際公開番号        | WO2022/053033                 |          | ン▼公樓                          |
| (87)国際公開日         | 令和4年3月17日(2022.3.17)          |          | Huawei Administrat            |
| 審査請求日             | 令和5年4月20日(2023.4.20)          |          | ion Building, Banti           |
| (31)優先権主張番号       | 202010955301.5                |          | an, Longgang Distri           |
| (32)優先日           | 令和2年9月11日(2020.9.11)          |          | ct, Shenzhen, Guang           |
| (33)優先権主張国・地域又は機関 | 中国(CN)                        |          | dong 5 1 8 1 2 9, P. R. C     |
| (31)優先権主張番号       | 202011628940.7                | (74)代理人  | 100110364                     |
| (32)優先日           | 令和2年12月30日(2020.12.30)        |          | 弁理士 実広 信哉                     |
| 最終頁に続く            |                               | 最終頁に続く   |                               |

(54)【発明の名称】 アクティブーアクティブストレージシステムおよびそのデータ処理方法

(57)【特許請求の範囲】

【請求項1】

第1のストレージデバイス（100）および第2のストレージデバイス（200）を備えるアクティブーアクティブストレージシステムにおいて、前記第1のストレージデバイスに接続された第1のクライアントおよび前記第2のストレージデバイスに接続された第2のクライアントには、同じファイルシステムがマウントされ、

前記第1のストレージデバイス（100）は、クライアントクラスタによってファイルシステムに送信された第1のファイルのデータを受信し、前記第1のファイルの前記データを記憶し、前記第1のファイルの前記データの第1のコピーデータを前記第2のストレージデバイス（200）に送信するように構成されており、

前記第2のストレージデバイス（200）は、前記クライアントクラスタによって前記ファイルシステムに送信された第2のファイルのデータを受信し、前記第2のファイルの前記データを記憶し、前記第2のファイルの前記データの第2のコピーデータを前記第1のストレージデバイス（100）に送信するように構成されており、前記第1のストレージデバイスおよび前記第2のストレージデバイスの両方は、同じファイルシステムを使用することによってファイルデータを記憶することを特徴とする、アクティブーアクティブストレージシステム。

【請求項2】

前記アクティブーアクティブストレージシステムは、仮想ノードセットをさらに備え、前記仮想ノードセットは、複数の仮想ノードを備え、コンピューティングリソースは、前

記仮想ノードの各々に割り振られ、前記コンピューティングリソースは、前記第1のストレージデバイス（100）または前記第2のストレージデバイス（200）内の物理ノードに由来する、請求項1に記載のシステム。

【請求項3】

前記アクティブーアクティブストレージシステムは、管理デバイスをさらに備え、前記管理デバイスは、グローバルビューを作成するようにさらに構成されており、前記グローバルビューは、前記仮想ノードの各々と、前記仮想ノードに割り振られた前記コンピューティングリソースとの間の対応関係を記録するために使用され、

前記管理デバイスは、前記第1のストレージデバイス（100）および前記第2のストレージデバイス（200）に前記グローバルビューを送信するようにさらに構成されており、

前記第1のストレージデバイス（100）は、前記グローバルビューを記憶するようにさらに構成されており、

前記第2のストレージデバイス（200）は、前記グローバルビューを記憶するようにさらに構成されている、

請求項2に記載のシステム。

【請求項4】

前記第1のファイルの前記データを記憶するとき、前記第1のストレージデバイス（100）は、

前記第1のファイルの前記データのアドレスに基づいて、前記第1のファイルに対応する第1の仮想ノードを決定し、

前記第1の仮想ノードおよび前記グローバルビューに基づいて、前記第1の仮想ノードに割り振られたコンピューティングリソースを決定し、

前記コンピューティングリソースに対応する物理ノードが前記第1のファイルの前記データを前記物理ノードのメモリに記憶するようにするために、前記第1の仮想ノードに割り振られた前記コンピューティングリソースに基づいて、前記第1のファイルの前記データを前記物理ノードに送信する

ように構成されている、請求項3に記載のシステム。

【請求項5】

前記第1の仮想ノードは、少なくとも1つのバックアップ仮想ノードを有し、前記第1の仮想ノードに対応する物理ノードと、前記バックアップ仮想ノードに対応する物理ノードとは、異なるストレージデバイスに配置されており、

前記第1のストレージデバイス（100）は、

前記第1の仮想ノードに対応する前記バックアップ仮想ノードを決定し、

前記バックアップ仮想ノードおよび前記グローバルビューに基づいて、前記バックアップ仮想ノードに対応する前記物理ノードを決定し、

前記バックアップ仮想ノードに対応する前記物理ノードが前記物理ノードに前記第1のコピーデータを記憶するようにするために、前記バックアップ仮想ノードに対応する前記物理ノードに前記第1のコピーデータを送信する

ようにさらに構成されている、請求項4に記載のシステム。

【請求項6】

前記ファイルシステムに含まれるファイルおよびディレクトリは、前記仮想ノードセット内の前記複数の仮想ノードに対応する物理ノードに分散される、請求項2から5のいずれか一項に記載のシステム。

【請求項7】

前記仮想ノードセット内の前記仮想ノードの各々に1つ以上のシャード識別子が設定され、前記ファイルシステム内の前記ディレクトリおよび前記ファイルの各々に1つのシャード識別子が割り振られ、前記第1のストレージデバイス（100）および前記第2のストレージデバイス（200）内の前記物理ノードは、前記ディレクトリおよび前記ファイルの各々の前記シャード識別子に基づいて、前記シャード識別子が属する仮想ノードに対応する物理ノードに前記ディレクトリおよび前記ファイルを分散させる、請求項6に記載のシス

10

20

30

40

50

テム。

【請求項 8】

前記第1のストレージデバイス（100）内の第1の物理ノードは、前記第1のファイルの作成要求を受信し、前記第1の物理ノードに対応する仮想ノードに設定された1つ以上のシャード識別子の中から前記第1のファイルの1つのシャード識別子を選択し、前記第1のストレージデバイス（100）内に前記第1のファイルを作成するように構成されている、請求項7に記載のシステム。

【請求項 9】

前記第2のストレージデバイス（200）が障害を起こしたとき、または前記第1のストレージデバイス（100）と前記第2のストレージデバイス（200）との間のリンクが切断されたとき、前記第1のストレージデバイス（100）は、前記第2のファイルの前記データの前記第2のコピーデータに基づいて前記第2のファイルを復元し、前記クライアントクラスタによって前記第2のストレージデバイス（200）に送信されたサービスを引き継ぐようにさらに構成されている、請求項2から8のいずれか一項に記載のシステム。

【請求項 10】

前記第1のストレージデバイス（100）は、前記グローバルビューから、前記第2のストレージデバイス（200）のコンピューティングリソースに対応する仮想ノードを削除するようにさらに構成されている、請求項3に記載のシステム。

【請求項 11】

前記第1のストレージデバイス（100）は、第1のファイルシステムをさらに有し、前記第2のストレージデバイス（200）は、第2のファイルシステムをさらに有する、請求項1に記載のシステム。

【請求項 12】

アクティブ-アクティブストレージシステムによって実行されるデータ処理方法において、前記アクティブ-アクティブストレージシステムは、第1のストレージデバイスおよび第2のストレージデバイスを備え、前記第1のストレージデバイスに接続された第1のクライアントおよび前記第2のストレージデバイスに接続された第2のクライアントには、同じファイルシステムがマウントされ、前記方法は、

前記第1のストレージデバイスによって、クライアントクラスタによってファイルシステムに送信された第1のファイルのデータを受信し、前記第1のファイルの前記データを記憶し、前記第1のファイルの前記データの第1のコピーデータを前記第2のストレージデバイスに送信するステップと、

前記第2のストレージデバイスによって、前記クライアントクラスタによって前記ファイルシステムに送信された第2のファイルのデータを受信し、前記第2のファイルの前記データを記憶し、前記第2のファイルの前記データの第2のコピーデータを前記第1のストレージデバイスに送信するステップと

を含み、前記第1のストレージデバイスおよび前記第2のストレージデバイスの両方は、同じファイルシステムを使用することによってファイルデータを記憶することを特徴とする、データ処理方法。

【請求項 13】

前記アクティブ-アクティブストレージシステムは、仮想ノードセットをさらに備え、前記仮想ノードセットは、複数の仮想ノードを備え、コンピューティングリソースは、前記仮想ノードの各々に割り振られ、前記コンピューティングリソースは、前記第1のストレージデバイスまたは前記第2のストレージデバイス内の物理ノードに由来する、請求項12に記載の方法。

【請求項 14】

前記アクティブ-アクティブストレージシステムは管理デバイスをさらに備え、前記方法は、

前記管理デバイスによって、グローバルビューを作成するステップであって、前記グローバルビューは、前記仮想ノードの各々と、前記仮想ノードに割り振られた前記コンピュ

10

20

30

40

50

ーティングリソースとの間の対応関係を記録するために使用される、ステップと、

前記管理デバイスによって、前記第1のストレージデバイスおよび前記第2のストレージデバイスに前記グローバルビューを送信するステップと、

前記第1のストレージデバイスおよび前記第2のストレージデバイスによって、前記グローバルビューを記憶するステップと

をさらに含む、請求項13に記載の方法。

【請求項15】

前記第1のファイルの前記データを記憶するとき、前記第1のストレージデバイスは、

前記第1のファイルの前記データのアドレスに基づいて、前記第1のファイルに対応する第1の仮想ノードを決定し、

前記第1の仮想ノードおよび前記グローバルビューに基づいて、前記第1の仮想ノードに割り振られたコンピューティングリソースを決定し、

前記コンピューティングリソースに対応する物理ノードが前記第1のファイルの前記データを前記物理ノードのメモリに記憶するようにするために、前記第1の仮想ノードに割り振られた前記コンピューティングリソースに基づいて、前記第1のファイルの前記データを前記物理ノードに送信する

ように構成されている、請求項14に記載の方法。

【請求項16】

前記第1の仮想ノードは、少なくとも1つのバックアップ仮想ノードを有し、前記第1の仮想ノードに対応する物理ノードと、前記バックアップ仮想ノードに対応する物理ノードとは、異なるストレージデバイスに配置されており、

前記方法は、

前記第1のストレージデバイスによって、前記第1の仮想ノードに対応する前記バックアップ仮想ノードを決定するステップと、

前記第1のストレージデバイスによって、前記バックアップ仮想ノードおよび前記グローバルビューに基づいて、前記バックアップ仮想ノードに対応する前記物理ノードを決定するステップと、

前記バックアップ仮想ノードに対応する前記物理ノードが前記物理ノードに前記第1のコピーデータを記憶するようにするために、前記第1のストレージデバイスによって、前記バックアップ仮想ノードに対応する前記物理ノードに前記第1のコピーデータを送信するステップと

をさらに含む、請求項15に記載の方法。

【請求項17】

前記ファイルシステムに含まれるファイルおよびディレクトリは、前記仮想ノードセット内の前記複数の仮想ノードに対応する物理ノードに分散される、請求項13から16のいずれか一項に記載の方法。

【請求項18】

前記仮想ノードセット内の前記仮想ノードの各々に1つ以上のシャード識別子が設定され、前記ファイルシステム内の前記ディレクトリおよび前記ファイルの各々に1つのシャード識別子が割り振られ、前記第1のストレージデバイスおよび前記第2のストレージデバイス内の前記物理ノードは、前記ディレクトリおよび前記ファイルの各々の前記シャード識別子に基づいて、前記シャード識別子が属する仮想ノードに対応する物理ノードに前記ディレクトリおよび前記ファイルを分散させる、請求項17に記載の方法。

【請求項19】

前記方法は、

前記第1のストレージデバイス内の第1の物理ノードによって、前記第1のファイルの作成要求を受信するステップと、前記第1の物理ノードに対応する仮想ノードに設定された1つ以上のシャード識別子の中から前記第1のファイルの1つのシャード識別子を選択するステップと、前記第1のストレージデバイス内に前記第1のファイルを作成するステップと

をさらに含む、請求項18に記載の方法。

10

20

30

40

50

## 【請求項 20】

前記方法は、前記第2のストレージデバイスが障害を起こしたとき、または前記第1のストレージデバイスと前記第2のストレージデバイスとの間のリンクが切断されたときに、前記第1のストレージデバイスによって、前記第2のファイルの前記データの前記第2のコピーデータに基づいて前記第2のファイルを復元するステップと、前記クライアントクラスタによって前記第2のストレージデバイスに送信されたサービスを引き継ぐステップとをさらに含む、請求項13から19のいずれか一項に記載の方法。

## 【請求項 21】

前記方法は、前記第1のストレージデバイスによって、前記グローバルビューから、前記第2のストレージデバイスのコンピューティングリソースに対応する仮想ノードを削除するステップをさらに含む、請求項14に記載の方法。

10

## 【請求項 22】

前記第1のストレージデバイスは第1のファイルシステムをさらに有し、前記第2のストレージデバイスは第2のファイルシステムをさらに有する、請求項13から21のいずれか一項に記載の方法。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本出願は、ストレージ分野に関し、特に、アクティブ－アクティブストレージシステムおよびそのデータ処理方法に関する。

20

## 【背景技術】

## 【0002】

ネットワークストレージクラスタ、例えば、ネットワークアタッチトストレージ（Network Attached Storage、NAS）クラスタの場合、アクティブ性－アクティブ性の実施時、書き込みデータを受信したとき、第1のストレージデバイスは、受信された書き込みデータをローカルに書き込み、受信された書き込みデータをバックアップデータとしてピアストレージデバイスに同期させる。このようにして、第1のストレージデバイスが障害を起こしたとき、または第1のストレージデバイスが第2のストレージデバイスから切断されたとき、第2のストレージデバイスは、サービスが中断されないことを保証する、すなわち、アクティブ－パッシブモードのアクティブ性－アクティブ性を実施するために、バックアップデータを使用して第1のストレージデバイスのサービスを引き継ぐことができる。しかしながら、アクティブ－アクティブモードのアクティブ性－アクティブ性は実施されることができない。

30

## 【発明の概要】

## 【0003】

本出願は、アクティブ－アクティブストレージシステム内のストレージデバイスが同じファイルシステム内のデータにアクセスできるように、アクティブ－アクティブモードのアクティブ性－アクティブ性を実施するために、アクティブ－アクティブストレージシステムおよびアクティブ－アクティブストレージシステムを実施するための方法を提供する。

40

## 【課題を解決するための手段】

## 【0004】

本出願の第1の態様は、アクティブ－アクティブストレージシステムを提供する。アクティブ－アクティブストレージシステムは、第1のストレージデバイスおよび第2のストレージデバイスを含む。第1のストレージデバイスは、クライアントクラスタによってファイルシステムに送信された第1のファイルのデータを受信し、第1のファイルのデータを記憶し、第1のファイルのデータの第1のコピーデータを第2のストレージデバイスに送信するように構成される。第2のストレージデバイスは、クライアントクラスタによってファイルシステムに送信された第2のファイルのデータを受信し、第2のファイルのデータを記憶し、第2のファイルのデータの第2のコピーデータを第1のストレージデバイスに送信す

50

るように構成される。

【0005】

第1のストレージデバイスと第2のストレージデバイスとの両方は、アクティブーアクティブモードのアクティブーアクティブストレージシステムを実施するために、同じファイルシステムを使用してファイルデータを記憶することができ、ピアエンドのファイルデータをバックアップすることができる。従来のNASデバイスもファイルシステムを有する。しかしながら、アクティブーパッシブモードの2つのストレージデバイスはそれぞれ、独立したファイルシステムを有する。2つの独立したファイルシステムの両方が、ストレージデバイスのコンピューティング／ストレージリソースを占有する必要があるため、その結果、低いリソース利用率および複雑な管理をもたらす。これは、実際のアクティブ性ーアクティブ性ではない。本出願では、第1のストレージデバイスおよび第2のストレージデバイスは、リソース利用率を改善し、管理の複雑さを低減するために、同じファイルシステムを有する。加えて、クライアントがストレージデバイスにアクセス要求を送信するとき、クライアントも同じファイルシステムに要求を送信する。したがって、クライアントのアクセス効率も改善される。

10

【0006】

本出願の第1の態様の可能な実施態様では、アクティブーアクティブストレージシステムは仮想ノードセットをさらに含み、仮想ノードセットは複数の仮想ノードを含む。コンピューティングリソースは、各仮想ノードに割り振られ、コンピューティングリソースは、第1のストレージデバイスまたは第2のストレージデバイス内の物理ノードに由来する。

20

【0007】

物理ノードは、第1のストレージデバイスまたは第2のストレージデバイスの制御ノードであってもよいし、または制御ノード内のCPUまたはCPU内のコアであってもよい。仮想ノードは論理的概念であり、システム内のコンピューティングリソースを分離するためのリソース割り振り媒体として使用される。このリソース管理方法では、独立したコンピューティングリソースが各仮想ノードに割り振られ、異なる仮想ノードに対応するファイル／ディレクトリによって使用されるコンピューティングリソースも独立している。これは、アクティブーアクティブストレージシステムの容量拡張または容量削減を容易にし、コンピューティングリソース間のロックフリーメカニズムの実施も容易にし、その結果、複雑さを低減する。

30

【0008】

本出願の第1の態様の可能な実施態様では、アクティブーアクティブストレージシステムは管理デバイスをさらに含む。管理デバイスは、グローバルビューを作成するようにさらに構成される。グローバルビューは、各仮想ノードと、仮想ノードに割り振られたコンピューティングリソースとの間の対応関係を記録するために使用される。管理デバイスは、第1のストレージデバイスおよび第2のストレージデバイスにグローバルビューを送信するようにさらに構成される。第1のストレージデバイスおよび第2のストレージデバイスは、グローバルビューを記憶する。

【0009】

管理デバイスは、ソフトウェアモジュールとして使用され、第1のストレージデバイスもしくは第2のストレージデバイスにインストールされてもよいし、または独立したデバイスであってもよい。管理デバイスが、第1のストレージデバイスにインストールされたソフトウェアモジュールとして使用されるとき、グローバルビューを生成した後、管理デバイスは、ストレージデバイス内の他のモジュールと相互作用することによって、記憶のために第1のストレージデバイスおよび第2のストレージデバイスにグローバルビューを送信する。

40

【0010】

仮想ノードセット内の仮想ノードは、グローバルビューの方法で第1のストレージデバイス内のアプリケーションおよび第2のストレージデバイス内のアプリケーションに別々に提示され、第1のストレージデバイス内のアプリケーションおよび第2のストレージデバ

50

イス内のアプリケーションは、使用のためにピアエンドの物理ノードをローカルエンドのリソースとして使用し、これにより、ピアエンドの物理ノードとの相互作用がより便利になる。

【0011】

本出願の第1の態様の可能な実施態様では、第1のファイルのデータを記憶するとき、第1のストレージデバイスは、第1のファイルのデータのアドレスに基づいて、第1のファイルに対応する第1の仮想ノードを決定し、第1の仮想ノードおよびグローバルビューに基づいて、第1の仮想ノードに割り振られたコンピューティングリソースを決定し、コンピューティングリソースに対応する物理ノードが第1のファイルのデータを物理ノードのメモリに記憶するようにするために、第1の仮想ノードに割り振られたコンピューティングリソースに基づいて、第1のファイルのデータを物理ノードに送信する。

10

【0012】

グローバルビューによって提供される仮想ノードセットを使用して、第1のストレージデバイスは、仮想ノードセット内の任意の仮想ノードに対応する物理ノードに属するファイルのデータを受信し、処理のために、受信されたファイルのデータを、ファイルが属する物理ノードに転送することができる。このようにして、データを書き込むとき、ユーザは、ファイルの実際の記憶場所を知覚する必要がなく、任意のストレージデバイスを使用してファイルを操作することができる。

【0013】

本出願の第1の態様の可能な実施態様では、第1の仮想ノードは、少なくとも1つのバックアップ仮想ノードを有し、第1の仮想ノードに対応する物理ノードと、バックアップ仮想ノードに対応する物理ノードとは、異なるストレージデバイスに配置される。第1のファイルに対応する第1の仮想ノードを決定した後、第1のストレージデバイスは、第1の仮想ノードに対応するバックアップ仮想ノードをさらに決定し、バックアップ仮想ノードおよびグローバルビューに基づいて、バックアップ仮想ノードに対応する物理ノードを決定し、バックアップ仮想ノードに対応する物理ノードが物理ノードに第1のコピーデータを記憶するようにするために、バックアップ仮想ノードに対応する物理ノードに第1のコピーデータを送信する。

20

【0014】

第1のストレージデバイスに書き込まれた、ファイルのデータは、第2のストレージデバイスにバックアップされる。第1のストレージデバイスが障害を起こした後、または第2のストレージデバイスから切断された後、システムの信頼性を改善するために、第1のストレージデバイスのサービスは、バックアップデータを使用して引き継がれることができる。

30

【0015】

本出願の第1の態様の可能な実施態様では、ファイルシステムに含まれるファイルおよびディレクトリは、仮想ノードセット内の複数の仮想ノードに対応する物理ノードに分散される。

【0016】

ファイルシステムに含まれるファイルおよびディレクトリが、仮想ノードセット内の複数の仮想ノードに対応する物理ノードに分散されることは、具体的には、ファイルシステムに含まれるファイルおよびディレクトリが処理のために複数の物理ノードに分散されることを意味する。このようにして、第1のストレージデバイスおよび第2のストレージデバイスの物理リソースは、ファイル処理効率を改善するために完全に使用されることができる。

40

【0017】

本出願の第1の態様の可能な実施態様では、仮想ノードセット内の各仮想ノードに1つ以上のシャード識別子が設定される。ファイルシステム内の各ディレクトリおよびファイルに1つのシャード識別子が割り振られる。第1のストレージデバイスおよび第2のストレージデバイス内の物理ノードは、各ディレクトリおよびファイルのシャード識別子に基づいて、シャード識別子が属する仮想ノードに対応する物理ノードにディレクトリおよびファ

50

イルを分散させる。

【0018】

ファイルシステムに含まれるファイルおよびディレクトリは、シャード識別子を使用して、第1のストレージデバイスおよび第2のストレージデバイスのすべての物理ノードにより便利に分散されることができる。

【0019】

本出願の第1の態様の可能な実施態様では、第1のストレージデバイス内の第1の物理ノードは、第1のファイルの作成要求を受信し、第1の物理ノードに対応する仮想ノードに設定された1つ以上のシャード識別子の中から第1のファイルの1つのシャード識別子を選択し、第1のストレージデバイス内に第1のファイルを作成するように構成される。

10

【0020】

ファイルが作成されるとき、処理効率を改善するために、ファイル作成要求が別の物理ノードに転送されないように、ファイル作成要求を受信した対応する物理ノードの仮想ノードのシャード識別子がファイルに割り振られる。

【0021】

本出願の第1の態様の可能な実施態様では、第2のストレージデバイスが障害を起こしたとき、または第1のストレージデバイスと第2のストレージデバイスとの間のリンクが切断されたとき、第1のストレージデバイスは、第2のファイルのデータの第2のコピーデータに基づいて第2のファイルを復元し、クライアントクラスタによって第2のストレージデバイスに送信されたサービスを引き継ぐようにさらに構成される。

20

【0022】

第1のストレージデバイスが障害を起こした後、または第2のストレージデバイスから切断された後、システムの信頼性を改善するために、第1のストレージデバイスのサービスは、バックアップデータを使用して引き継がれることができる。

【0023】

本出願の第1の態様の可能な実施態様では、第1のストレージデバイスは、グローバルビューから、第2のストレージデバイスのコンピューティングリソースに対応する仮想ノードを削除するようにさらに構成される。

【0024】

本出願の第1の態様の可能な実施態様では、第1のストレージデバイスは第1のファイルシステムをさらに有し、第2のストレージデバイスは第2のファイルシステムをさらに有する。

30

【0025】

ユーザがストレージデバイス内のデータにアクセスするための複数の方法を提供するために、ローカルファイルシステムおよびクラスタファイルシステムは、同じストレージデバイス上で同時に動作する。

【0026】

本出願の第2の態様は、アクティブ-アクティブファイルシステムを実施するための方法を提供する。本方法に含まれるステップは、本出願の第1の態様で提供されるアクティブ-アクティブストレージシステムにおいて第1のストレージデバイスおよび第2のストレージデバイスによって実行されるすべての機能を実施するために使用される。

40

【0027】

本出願の第3の態様は管理デバイスを提供する。管理デバイスは、グローバルビューを作成するように構成される。グローバルビューは、各仮想ノードと、仮想ノードに割り振られたコンピューティングリソースとの間の対応関係を記録するために使用される。管理デバイスは、記憶のために第1のストレージデバイスおよび第2のストレージデバイスにグローバルビューを送信するようにさらに構成される。

【0028】

管理デバイスは、第1のストレージデバイスおよび第2のストレージデバイス内の仮想ノードの変化を監視し、仮想セットに新しい仮想ノードが追加されたことを検出したとき、

50

または仮想ノードが削除されたとき、例えば、仮想ノードに対応する物理ノードが障害を起こしたときに、グローバルビューを更新するように構成される。

【0029】

監視モジュールは、グローバルビューを適時に更新するために、仮想ノードセット内の仮想ノードの変化をリアルタイムで監視することができる。

【0030】

本出願の第4の態様は、プログラム命令を記憶するように構成されたストレージ媒体を提供する。プログラム命令は、第3の態様で提供される管理デバイスによって提供される機能を実施するために使用される。

【0031】

本出願の実施形態における技術的解決策をより明確に説明するために、以下は、実施形態または従来技術を説明する際に使用される添付の図面について簡単に説明する。

【図面の簡単な説明】

【0032】

【図1】アクティブ—パッシブモードのアクティブ—アクティブストレージシステムのアーキテクチャ図である。

【図2】本出願の一実施形態によるアクティブ—アクティブモードのアクティブ—アクティブストレージシステムのアーキテクチャ図である。

【図3A】本出願の一実施形態によるアクティブ—アクティブストレージシステムを確立するための方法のフローチャートである。

【図3B】本出願の一実施形態によるアクティブ—アクティブストレージシステムを構築するプロセスで生成されるパラメータの概略図である。

【図4A】本出願の一実施形態によるアクティブ—アクティブストレージシステムのファイルシステムを確立するフローチャートである。

【図4B】本出願の一実施形態による構築されたアクティブ—アクティブシステムの概略図である。

【図5】本出願の一実施形態によるファイルシステムにおいてディレクトリを作成するための方法のフローチャートである。

【図6】本出願の一実施形態によるファイルシステムにおいてディレクトリを問い合わせるための方法のフローチャートである。

【図7】本出願の一実施形態によるファイルシステムにおいてファイルを作成するための方法のフローチャートである。

【図8】本出願の一実施形態によるファイルシステムにおいてファイルにデータを書き込むための方法のフローチャートである。

【図9】本出願の一実施形態によるファイルシステムにおいてファイルにデータを書き込むための方法のフローチャートである。

【図10】本出願の一実施形態による、第1のストレージデバイスが第2のストレージデバイスのサービスを引き継ぐ概略図である。

【図11】本出願の一実施形態による、第1のストレージデバイスによって第2のストレージデバイスのサービスを引き継ぐための方法のフローチャートである。

【発明を実施するための形態】

【0033】

以下は、本出願の実施形態において添付の図面を参照して、本出願の実施形態における技術的解決策について明確かつ完全に説明する。説明されている実施形態は、本出願の実施形態の全部ではなく一部にすぎないことは明らかである。

【0034】

図1は、Active—passiveモードのアクティブ—アクティブシステムの概略アーキテクチャ図である。システム10は、第1のストレージデバイス100および第2のストレージデバイス200を含む。第1のストレージデバイス100の制御ノード101には第1のファイルシステム102が配置され（第1のストレージデバイスは、複数の制御ノードを含んでもよく

10

20

30

40

50

、説明を容易にするために、1つの制御ノードのみが説明のための例として使用される)  
、第2のストレージデバイス200の制御ノード201には第2のファイルシステム202が配置される(第2のストレージデバイスも、複数の制御ノードを含んでもよく、説明を容易にするために、1つの制御ノードのみが説明のための例として使用される)。第1のクライアント300が第1のストレージデバイス100に接続された後、第1のストレージデバイス100は、第1のファイルシステム102を第1のクライアント300にマウントする。第2のクライアント400が第2のストレージデバイス200に接続された後、第2のストレージデバイス200は、第2のファイルシステム202を第2のクライアント400にマウントする。各ファイルシステムは、ルートディレクトリを有する。ストレージデバイスがファイルシステムをクライアントにマウントすることは、ストレージデバイスがファイルシステムのルートディレクトリをクライアントに提供し、クライアントがファイルシステムのルートディレクトリをクライアントのファイルシステムに設定することを意味し、これにより、クライアントは、ストレージデバイスのファイルシステムのルートディレクトリに基づいてストレージデバイスのファイルシステムにアクセスするために、ストレージデバイスのファイルシステムのルートディレクトリを取得することができる。このようにして、第1のファイルシステム102が第1のクライアント300にマウントされた後、第1のクライアント300は、第1のファイルシステム102を使用してデータを読み書きし、書き込みデータはローカルデータ103として記憶される。加えて、第1のストレージデバイス100は、第2のストレージデバイス200のバックアップデータ、すなわちピアバックアップデータ104をさらに記憶する。同様に、第2のクライアント400は、第2のファイルシステム202を使用してデータを

読み書きし、書き込みデータはローカルデータ203として記憶される。加えて、第2のストレージデバイス200は、第1のストレージデバイス100のバックアップデータ、すなわちピアバックアップデータ204をさらに記憶する。このようにして、第1のストレージデバイス100が障害を起こした後、または第1のストレージデバイス100と第2のクライアントとの間のリンクが切断された後、第2のクライアントは、ピアバックアップデータ204を使用して第1のクライアント300のサービスを引き継ぐことができる。言い換えれば、Active-passiveモードのアクティブ性-アクティブ性が実施される。しかしながら、Active-passiveモードのアクティブ-アクティブシステム10では、第1のストレージデバイス100と第2のストレージデバイス200との両方が正常に動作しているとき、第1のクライアント300は、第1のファイルシステムを使用して第1のストレージデバイス100内のデータにしかアクセスできず、第2のストレージデバイス200内のデータにはアクセスできず、第2のクライアント400は、第2のファイルシステムを使用して第2のストレージデバイス200内のデータにしかアクセスできず、第1のストレージデバイス100内のデータにはアクセスできない。言い換えれば、Active-Activeモードのアクティブ性-アクティブ性は実施されることができない。

#### 【0035】

本出願の実施形態で提供される技術的解決策では、グローバルビューが設定される。グローバルビューは、仮想ノードのセットであり、グローバルビュー内の各仮想ノードにコンピューティングリソースが割り振られる。コンピューティングリソースは、第1のストレージデバイスの物理ノードおよび第2のストレージデバイスの物理ノードに由来し、物理ノードは、第1のストレージデバイス内のコントローラおよび第2のストレージデバイス内のコントローラであってもよいし、またはコントローラ内のCPU、もしくはCPU内のコアであってもよいし、または分散ストレージシステム内のサーバであってもよい。本出願の実施形態では、各物理ノードは、グローバルビューを取得することができる。加えて、各物理ノードは、同じファイルシステムをさらに使用する。このようにして、第1のストレージデバイスに接続された第1のクライアントおよび第2のストレージデバイスに接続された第2のクライアントには、同じファイルシステムがマウントされる。結果として、第1のクライアントは、ファイルシステムおよびグローバルビューを使用して、第2のストレージデバイス内の、ファイルシステムに属するデータにアクセスすることができる。以下は、添付の図面を参照して、本出願の実施形態で提供される解決策について詳細に説明す

る。

#### 【0036】

図2は、本出願の一実施形態によるアクティブ-アクティブモードのアクティブ-アクティブシステム500のアーキテクチャ図である。システム500は、第1のストレージデバイス600および第2のストレージデバイス700を含む。第1のストレージデバイス600は、物理ノードAおよび物理ノードBを含む。第2のストレージデバイス700は、物理ノードCおよび物理ノードDを含む。実際の適用時、第1のストレージデバイス600および第2のストレージデバイス700はそれぞれ、より多くの物理ノードを含んでもよい。説明を容易にするために、この実施形態では、各ストレージデバイスが2つの物理ノードを含む例のみが説明のために使用される。第1のストレージデバイス600および第2のストレージデバイス700は、データを永続的に記憶するように構成された、複数のストレージディスクを含む永続ストレージデバイス601および永続ストレージデバイス701をそれぞれ含む。永続ストレージデバイス601および永続ストレージデバイス701のストレージディスクによって提供される物理記憶空間に基づいて、第1のストレージデバイス600および第2のストレージデバイス700は、それぞれ第1のボリューム609および第2のボリューム703を作成する。第1のストレージデバイス600および第2のストレージデバイス700は、第1のボリューム609および第2のボリューム703に基づいて、永続ストレージデバイス601および永続ストレージデバイス701にデータをそれぞれ記憶し得る。例えば、ストレージディスクは、永続ストレージ媒体、例えばソリッドステートディスク（Solid State Disk、SSD）またはハードディスクドライブ（Hard Disk Drive、HDD）であってもよい。

#### 【0037】

物理ノードA、物理ノードB、物理ノードC、および物理ノードDの構造は同じである。本出願のこの実施形態では、ノードAの構造のみが説明のための例として使用される。物理ノードAは、プロセッサ602およびメモリ603を含む。メモリ603は、アプリケーションプログラム命令（図示せず）と、プロセッサの動作プロセスで生成されたデータとを記憶する。プロセッサ602は、本出願のこの実施形態で提供されるActive-Activeモードのアクティブ-アクティブ機能を実施するためにアプリケーションプログラム命令を実行する。第1のファイルシステム608に加えて、メモリ603は、グローバルビュー604、ファイルシステム605、キャッシュデータ606、およびバックアップデータ607をさらに記憶する。第1のファイルシステム608の機能は、図1における第1のファイルシステム102の機能と同じである。ここでは詳細は再び説明されない。言い換えれば、本出願のこの実施形態では、各物理ノードは2つのファイルシステムを含み、一方は、すべての物理ノードによって共有されるファイルシステムであり、他方は、各物理ノードのファイルシステムである。メモリ603内の他のデータの詳細な説明は、アクティブ性-アクティブ性を実施するための方法、例えば、図5から図9に示されているフローチャートを参照して提供される。第1のクライアント800は、第1のストレージデバイス600内のデータにアクセスするために第1のストレージデバイス600に接続され、第2のクライアント900は、第2のストレージデバイス700内のデータにアクセスするために第2のストレージデバイス700に接続される。

#### 【0038】

以下は、図3A、図4A、および図5から図9のフローチャートを参照して、本出願の実施形態によるアクティブ-アクティブモードのアクティブ性-アクティブ性を実施するための方法について説明する。

#### 【0039】

最初に、図3Aは、本出願の一実施形態による、グローバルビューを確立するための方法のフローチャートである。

#### 【0040】

ステップS301：第1のストレージデバイス600の物理ノードAは、クライアントによって送信された仮想クラスタ確立要求を受信する。

#### 【0041】

アクティブ・アクティブシステムが構築される必要があるとき、グローバルビューが確立され、ユーザは、クライアントを使用してグローバルビュー確立要求を第1のストレージデバイス600に送信し得る。第1のストレージデバイスはプライマリアレイであり、第1のストレージデバイス600内の物理ノードAはプライマリノードである。この場合、物理ノードAが要求を処理する。

【0042】

ステップS302：物理ノードAは、グローバルビュー604を確立し、確立されたグローバルビュー604を、グローバルビュー内の別の仮想ノードに対応する物理ノードに同期させる。

【0043】

第1のストレージデバイス600が第2のストレージデバイス700とのネットワーク接続を確立した後、第1のストレージデバイス600は、第2のストレージデバイス700内の各物理ノードの識別子と、各物理ノードのIPアドレスとを取得する。グローバルビュー604が確立されるとき、ノードAは、仮想ノードを識別するために、第1のストレージデバイス600および第2のストレージデバイス700内の各物理ノードに仮想識別子を割り振り、仮想ノードの仮想識別子を記録するためにグローバルビューを確立する。各物理ノードのコンピューティングリソース、例えばプロセッサリソースまたはメモリリソースは、仮想ノードに割り振られたコンピューティングリソースである。別の実施形態では、コンピューティングリソースに加えて、別の物理リソース、例えば帯域幅が、各仮想ノードにさらに割り振られてもよい。本出願のこの実施形態では、すべての仮想ノードに割り振られた物理リソースは互いに独立している。このようにして、容量拡張が、ストレージデバイスにとってより便利になることができる。例えば、新しい物理リソースがストレージデバイスに追加されるとき、仮想ノードの数を増加させるために、新しい物理リソースに基づいて新しい仮想ノードが生成され、新しく追加された仮想ノードはグローバルビューに追加される。分散ストレージでは、追加されたサーバが新しい物理リソースとして使用され、グローバルビュー内の仮想ノードの数を増加させるために、追加されたサーバに基づいて仮想ノードが確立される。確立されたグローバルビューは、図3BではVclusterとして示されている。例えば、第1のストレージデバイス600内の物理ノードAおよび物理ノードBには仮想識別子Vnode AおよびVnode Bが割り振られ、第2のストレージデバイス700内の物理ノードCおよび物理ノードDには仮想識別子Vnode CおよびVnode Dが割り振られる。グローバルビュー604が生成された後、ノードAは、メモリ603および永続ストレージデバイス601にグローバルビュー604を記憶し、次に、他の仮想ノードに対応する物理ノード（物理ノードB、C、およびD）ならびに第2のストレージデバイス700の永続ストレージ媒体701にノードセットテーブル604を同期させる。

【0044】

ステップS303：物理ノードAは、ノードセットに基づいてシャード（shard）ビューを生成し、仮想ノードクラスタ内の別の仮想ノードに対応する物理ノードにshardビューを同期させる。

【0045】

本出願のこの実施形態では、予め設定された数のShard、例えば4096個のShardが仮想クラスタに設定され、これらのshardは、グローバルビュー604内のすべての仮想ノードに均等に割り振られ、すなわち、shardビューが生成される。生成されたshardビューは、図3Bではshardビューとして示されている。shardは、ファイルシステム605のディレクトリおよびファイルを、グローバルビュー604内のすべての仮想ノードに対応する物理ノードに分散して記憶するように構成される。shardビューの具体的な機能は、以下で詳細に説明される。shardビューが生成された後、物理ノードAは、ローカルメモリ603および永続ストレージ媒体601にshardビューを記憶し、他の仮想ノードに対応する物理ノード（物理ノードB、C、およびD）ならびに第2のストレージデバイス700の永続ストレージ媒体701にshardビューを同期させる。

【0046】

ステップS304：物理ノードAは、データバックアップポリシーを生成し、仮想ノードクラスタ内の別の仮想ノードに対応する物理ノードにデータバックアップポリシーを同期させる。

【0047】

データの信頼性を保証し、デバイス障害後のデータ損失を防止するために、本出願のこの実施形態では、データバックアップポリシーが設定され得、すなわち、生成されたデータは複数のノードにバックアップされる。本出願のこの実施形態におけるバックアップポリシーは、データの3つのコピーをバックアップすることであり、2つのコピーは2つのローカル物理ノードに記憶され、残りの1つのコピーはリモートストレージデバイスの物理ノードに記憶される。具体的には、図3Bに示されているバックアップポリシーでは、各仮想ノードにバックアップノードのグループが設定されている。例えば、仮想ノードVnodeAに対応するバックアップノードは仮想ノードVnodeBおよびVnodeCに設定され、仮想ノードVnodeBに対応する仮想ノードはVnodeAおよびVnodeDに設定され、仮想ノードVnodeCに対応する仮想ノードはVnodeAおよびVnodeDに設定され、仮想ノードVnodeDに対応する仮想ノードはVnodeCおよびVnodeBに設定される。バックアップポリシーが生成された後、ノードAは、ローカルメモリ603および永続ストレージデバイス601にバックアップポリシーを記憶し、第2のストレージデバイス700の永続ストレージデバイス701およびグローバルビュー内の別の仮想ノードに対応する物理ノードにバックアップポリシーを同期させる。

【0048】

図3Aでは、仮想クラスタの確立は管理モジュールによって実行される。図3Aおよび図4Aでは、管理モジュールが第1のストレージデバイスに配置されている例が説明のために使用される。ファイルシステムおよびグローバルビューを生成した後、管理モジュールは、生成されたファイルシステムおよびグローバルビューを記憶のために第1のストレージデバイスおよび第2のストレージデバイスに送信し得る。別の実施形態では、管理モジュールは、代わりに、独立した第三者管理デバイスに配置されてもよい。ファイルシステムおよびグローバルビューを生成した後、第三者管理デバイスは、ファイルシステムおよびグローバルビューを記憶のために第1のストレージデバイスおよび第2のストレージデバイスに送信し、これにより、各物理ノードはグローバルビューを取得することができる。

【0049】

確立された仮想クラスタの動作時、監視モジュールは、第1のストレージデバイスおよび第2のストレージデバイス内の仮想ノードの変化を監視する。監視モジュールは、新しい仮想ノードが仮想クラスタに追加されたことを検出したとき、または仮想ノードが削除されたとき、例えば、仮想ノードに対応する物理ノードが障害を起こしたとき、管理モジュールにグローバルビューを更新するように通知する。監視モジュールは、第三者管理デバイスに配置されてもよいし、または第1のストレージデバイスもしくは第2のストレージデバイスに配置されてもよい。第1のストレージデバイスはプライマリストレージデバイスとして機能し、第2のストレージデバイスは監視された変化を第1のストレージデバイスに送信し、第1のストレージデバイス内の管理モジュールはグローバルビューを更新する。このようにして、仮想ノードクラスタの確立が遂行されることができる。仮想ノードクラスタが確立された後、第1のストレージデバイス600および第2のストレージデバイス700は、クライアントの要求に基づいてファイルシステムを確立し得る。詳細は図4Aのフローチャートに示されている。

【0050】

ステップS401：物理ノードAは、ファイルシステム作成要求を受信する。

【0051】

第1のクライアント800は、ファイルシステム作成要求を第1のストレージデバイス600に送信してもよいし、またはファイルシステム作成要求を第2のストレージデバイス700に送信してもよい。第1のストレージデバイス600がファイルシステム作成要求を受信した場合、物理ノードAがファイルシステム作成要求を処理する。第2のストレージデバイス

700がファイルシステム作成要求を受信した場合、第2のストレージデバイス700は、ファイルシステム作成要求を処理のために第1のストレージデバイス600の物理ノードAに転送する。

【0052】

ステップS402：物理ノードAは、ファイルシステムのルートディレクトリを設定する。

【0053】

ルートディレクトリを設定するとき、プライマリノードは最初にルートディレクトリのマークを生成する。一般に、ルートディレクトリのデフォルトマークは「/」である。次に、識別情報およびshard IDがルートディレクトリに割り振られる。プライマリノードによって作成されたshardビューはすべてのノードに同期されるため、プライマリノードは、プライマリノードのメモリからshardビューを取得し、shardビューからルートディレクトリのshard IDを選択する。図3Bに示されているように、shardビューにおいて、複数のshard IDが各仮想ノードに割り振られる。したがって、クロスネットワークおよびクロスノードアクセスを低減するために、物理ノードAに対応する仮想ノードVnode Aに含まれるshard IDのうちのshard IDが、好ましくは、ルートディレクトリに割り振られる。ルートディレクトリには、shard IDが割り振られていない。したがって、例えば、shard 0が、ルートディレクトリのshard IDとして選択されてもよい。

【0054】

ステップS403：物理ノードAは、ファイルシステムのマウントコマンドを第1のクライアント800に送信する。

【0055】

クラスタファイルシステムのルートディレクトリが生成された後、第1のクライアント800がファイルシステムにアクセスすることを可能にするために、物理ノードAは、ファイルシステムを第1のクライアント800のファイルシステムにマウントする。例えば、物理ノードAは、mountコマンドを使用して、ファイルシステムのルートディレクトリを第1のクライアント800に提供する。mountコマンドを送信するとき、物理ノードAは、ルートディレクトリのパラメータ情報を搬送する。ルートディレクトリのパラメータ情報は、ルートディレクトリのハンドル情報であり、ハンドル情報は、shard IDおよびルートディレクトリの識別情報を搬送する。

【0056】

ステップS404：第1のクライアント800は、マウントコマンドに従って、クラスタファイルシステムを第1のクライアント800のファイルシステムにマウントする。

【0057】

ファイルシステムのルートディレクトリのパラメータ情報を受信した後、第1のクライアント800は、第1のクライアントのファイルシステム上にマウントポイントを生成し、マウントポイントにファイルシステムのルートディレクトリのパラメータ情報を記録する。マウントポイントは、記憶空間のセグメントである。

【0058】

このようにして、第1のファイルシステム608を使用して第1のストレージデバイス600とデータ伝送を実行することに加えて、第1のクライアント800は、代わりに、ファイルシステム605を使用して第1のストレージデバイス600とデータ伝送を実行することができる。ユーザは、実際の要件に基づいて、ユーザがアクセスする必要があるファイルシステムを選択することができる。

【0059】

ステップS405：物理ノードAは、仮想ボリュームをファイルシステムに割り振る。

【0060】

仮想ボリュームVvolume 0は、新たに作成された各ファイルシステムに割り振られ、第1のクライアントまたは第2のクライアントによって書き込まれたデータをファイルシステムに書き込むために使用される。ステップS406：物理ノードAは、仮想ボリュームのミラーボリュームペアを作成する。

## 【0061】

仮想ボリュームVvloume 0が確立された後、物理ノードAは、最初に、永続ストレージ媒体601、例えば図2の第1のボリュームに基づいてローカルボリュームを作成し、次に、第2のストレージデバイス700、例えば図2の第2のボリュームのファイルシステム内に第1のボリュームのミラーボリュームを作成するように第2のストレージデバイス700に要求する。

## 【0062】

ステップS407：物理ノードAは、仮想ボリュームおよび対応するミラーボリュームペアを記録することによってフラッシングポリシーを生成する。

## 【0063】

生成されたフラッシングポリシーは、図3Bに示されているフラッシングポリシーとして示されており、ファイルシステムの仮想ボリュームは、ミラーボリュームペア（第1のボリュームおよび第2のボリューム）に対応する。図3Bに示されているフラッシングポリシーに従って、メモリにキャッシュされている、ファイルシステムのデータは、データの信頼性を保証するために、第1のストレージデバイス600の永続ストレージ媒体601および第2のストレージデバイス700の永続ストレージ媒体701に別々に記憶され得る。具体的には、フラッシングポリシーに従って永続ストレージ媒体601および永続ストレージ媒体701にメモリ内のデータをどのように書き込むかは、図9で詳細に説明される。

## 【0064】

フラッシングポリシーが生成された後、物理ノードAは、ファイルシステムのフラッシングポリシーをローカルメモリ603および永続ストレージデバイス601に記憶し、ファイルシステムのフラッシングポリシーを、第2のストレージデバイス700の永続ストレージデバイス701およびグローバルビュー内の別の仮想ノードに対応する物理ノードに同期させる。

## 【0065】

図3Aおよび図4Aの方法を実行することによって、Active-Activeモードのアクティブ-アクティブファイルシステムの作成が遂行されることができ。ファイルシステムが作成された、すなわち、第1のストレージデバイスおよび第2のストレージデバイス上にクロスデバイスファイルシステム、仮想ボリューム、Shardビュー、およびグローバルビューが生成されたアクティブ-アクティブストレージシステムの概略図が図4Bに示されている。

## 【0066】

アクティブ-アクティブモードのアクティブ-アクティブストレージシステムが作成された後、ファイルシステムに基づいて、ディレクトリおよびファイルは作成およびアクセスされ得る。

## 【0067】

最初に、図5に示されているフローチャートを参照して、ファイルシステムの下にディレクトリを作成するプロセスが説明される。以下は、説明のために、ルートディレクトリを親ディレクトリとして使用し、作成されるディレクトリを親ディレクトリのサブディレクトリとして使用する。本出願のこの実施形態では、ユーザは、サブディレクトリを作成するために、第1のクライアントを使用して第1のストレージデバイスにアクセスし得るし、またはサブディレクトリを作成するために、第2のクライアントを使用して第2のストレージデバイスにアクセスし得る。第1のストレージデバイスがファイルシステムを第1のクライアントにマウントするとき、第1のクライアントがファイルシステムにアクセスするためのパスが確立される。例えば、第1のストレージデバイスが、物理ノードAを使用してファイルシステムを第1のクライアントにマウントする場合、第1のクライアントは、物理ノードAを使用してファイルシステムにアクセスする。アクティブ-アクティブモードのアクティブ-アクティブアクセスを実施するために、第2のストレージデバイスも、ファイルシステムを第2のクライアントのファイルシステムにマウントする。このようにして、第2のクライアントがファイルシステムにアクセスするためのパスが確立され、第2のク

10

20

30

40

50

クライアントによってファイルシステムにアクセスするための要求が、ファイルシステムをマウントする物理ノード、例えば物理ノードCに送信される。以下は、サブディレクトリ作成要求が第2のクライアントを使用して第2のストレージデバイスに送信される例を使用して、サブディレクトリ作成プロセスについて説明する。

【0068】

具体的な作成プロセスは図5のフローチャートに示されている。

【0069】

ステップS501：第2のクライアントは、サブディレクトリ作成要求を物理ノードCに送信する。

【0070】

物理ノードCは、第2のストレージデバイス700のプライマリノード、すなわち、ファイルシステムを第2のクライアントにマウントするノードである。サブディレクトリ作成要求は、親ディレクトリのパラメータ情報およびサブディレクトリの名称を含む。

【0071】

ステップS502：物理ノードCは、第2のクライアントによって送信された作成要求を受信し、作成要求に基づいてサブディレクトリのパラメータ情報を生成する。

【0072】

パラメータ情報は、親ディレクトリの識別情報と、shard IDとを含む。識別情報は、サブディレクトリを一意に識別するために使用され、識別情報は、例えば、NFSファイルシステムにおけるオブジェクトIDである。Shard IDを生成するとき、物理ノードCは、shardビューを検索し、shardビューに記録されているShard IDのうちのshard IDをサブディレクトリに割り振り、次に、そのshard IDが属する仮想ノードに対応する物理ノードにサブディレクトリを作成する。各ディレクトリには1つのshard IDが割り振られてもよいが、複数のディレクトリに1つのshard IDが割り振られてもよいことに留意されたい。本出願のこの実施形態では、データ転送を低減するために、サブディレクトリ要求を受信する物理ノードに対応する仮想ノードのshard IDのうちのshard IDがサブディレクトリに割り振られる。具体的には、物理ノードCに対応する仮想ノードVnode Cに対応するshard ID [2048, 3071] のうちのshard IDがサブディレクトリに割り振られる。しかしながら、仮想ノードVnode Cのshard IDに対応するディレクトリの数が予め設定された閾値を超えるときには、別の仮想ノードに対応するshard IDがサブディレクトリに割り振られる。

【0073】

ステップS503：物理ノードCはサブディレクトリを作成する。

【0074】

サブディレクトリを作成することは、サブディレクトリのディレクトリエントリテーブル (directory entry table、DET) およびInodeテーブルを生成することを含む。ディレクトリエントリテーブルは、サブディレクトリが正常に作成された後にサブディレクトリが親ディレクトリとして機能するときに、サブディレクトリまたはサブディレクトリの下に作成されたファイルのパラメータ情報を記録するために使用される。パラメータ情報は、例えば、サブディレクトリの名称、ならびにディレクトリまたはファイルの識別情報およびshard IDを含む。

【0075】

Inodeテーブルは、その後サブディレクトリに作成されるファイルに関する詳細情報、例えば、ファイルのファイル長、ファイルに対するユーザの操作許可、およびファイルの修正時点などの情報を記録するために使用される。

【0076】

ステップS504：物理ノードCは、親ディレクトリのパラメータ情報に基づいて、第1のストレージデバイス内の、親ディレクトリが属する物理ノードBを決定する。

【0077】

親ディレクトリのパラメータ情報は、Shard IDを含む。Shardビューにおいて、Shard IDに対応する仮想ノードは仮想ノードVnode Bであると決定され得、次に、仮想ノードV

10

20

30

40

50

node Bに基づいて、仮想ノードVnode Bに対応する物理ノードは第1のストレージデバイス内の物理ノードBであるとさらに決定される。

【0078】

ステップS505：物理ノードCは、サブディレクトリのパラメータ情報および親ディレクトリのパラメータ情報を物理ノードBに送信する。

【0079】

ステップS506：物理ノードBは、親ディレクトリのパラメータ情報に基づいて、親ディレクトリのディレクトリエントリテーブルを見つける。

【0080】

具体的には、親ディレクトリは、親ディレクトリのパラメータ情報内のshard IDおよび親ディレクトリの名称に基づいて見つけれ得る。

【0081】

ステップS507：物理ノードBは、サブディレクトリのパラメータ情報を親ディレクトリのディレクトリエントリテーブルに記録する。

【0082】

ステップS508：物理ノードBは、最初にサブディレクトリのパラメータ情報を物理ノードCに返し、次に物理ノードCは、サブディレクトリのパラメータ情報を第2のクライアントに返す。

【0083】

ファイルシステムにおいて、ファイルにアクセスするプロセス、例えば、ファイルを読み出すプロセスまたはファイルを作成するプロセスでは、ファイルはディレクトリの下に作成されるため、ディレクトリの下にファイルがさらにアクセスされる前に、ディレクトリが最初に見つけられる必要がある。アクセスされるファイルがマルチレベルディレクトリの下にある場合、ディレクトリは、最下位レベルのディレクトリが見つけれられるまでレベルごとに問い合わせられる必要がある。例えば、マルチレベルディレクトリfilesystem1/user1/favoriteの場合、ルートディレクトリのパラメータ情報が第1のクライアントのファイルシステムに記録されているため、クライアントは、最初に、ルートディレクトリfilesystem1のパラメータ情報に基づいてサブディレクトリuser1のパラメータ情報を問い合わせ、すなわち、「user1」を問い合わせるための要求を生成し、「user1」のパラメータ情報が問い合わせされた後、「user1」のパラメータ情報に基づいて「favorite」のパラメータ情報を問い合わせる、すなわち、「favorite」を問い合わせるための要求を生成する。各レベルのディレクトリのパラメータ情報を問い合わせるための方法は同じである。以下は、ディレクトリ問い合わせプロセスを説明するために、上位層ディレクトリが親ディレクトリであり、問い合わせされるディレクトリがサブディレクトリである例を使用する。本出願のこの実施形態では、第2のストレージデバイスの物理ノードCが問い合わせ要求を受信する例が、説明のために引き続き使用される。

【0084】

ステップS601：第2のクライアントは、サブディレクトリの問い合わせ要求を物理ノードCに送信する。

【0085】

問い合わせ要求は、親ディレクトリのパラメータ情報およびサブディレクトリの名称を搬送する。親ディレクトリのパラメータ情報は、例えば、親ディレクトリのハンドルである。親ディレクトリがルートディレクトリであるとき、クライアントのファイルシステムからルートディレクトリのハンドルが取得される。親ディレクトリがルートディレクトリでないとき、親ディレクトリのハンドルは、親ディレクトリを問い合わせるための問い合わせ要求を使用して見つけれ得る。

【0086】

親ディレクトリのハンドルは、親ディレクトリの識別情報およびshard IDを含む。

【0087】

ステップS602：物理ノードCは、第2のクライアントによって送信された問い合わせ要

10

20

30

40

50

求を受信し、問い合わせ要求に基づいて、親ディレクトリが属する物理ノードBを決定する。

【0088】

物理ノードCは、ルートディレクトリのパラメータ情報からルートディレクトリのshard IDを取得し、そのshard IDに基づいて、親ディレクトリが属する仮想ノードを取得する。

【0089】

物理ノードAは、作成されたshardビューをすべてのノードに同期させるため、物理ノードCは、物理ノードCのメモリからshardビューを取得し、親ディレクトリのShard IDに基づいて、親ディレクトリが属する仮想ノードを決定し、次に、仮想ノードに対応する物理ノードを決定する。

【0090】

ステップS603：物理ノードCは、親ディレクトリのパラメータ情報およびサブディレクトリの名称を、親ディレクトリが配置されている物理ノードBに送信する。

【0091】

ステップS604：物理ノードBは、親ディレクトリのパラメータ情報に基づいて、親ディレクトリのディレクトリエントリテーブルを決定する。

【0092】

図5の説明を参照されたい。親ディレクトリを作成するとき、物理ノードBは、親ディレクトリのディレクトリエントリテーブルを作成し、ディレクトリエントリテーブルは、親ディレクトリの下に作成されるすべてのサブディレクトリのパラメータ情報を記録する。

【0093】

ステップS605：物理ノードBは、親ディレクトリのディレクトリエントリテーブルからサブディレクトリのパラメータ情報を取得する。

【0094】

ステップS606：物理ノードBは、サブディレクトリのパラメータ情報を物理ノードCに返す。

【0095】

ステップS607：物理ノードCは、サブディレクトリのパラメータ情報を第2のクライアントに返す。

【0096】

図5および図6では、第2のクライアントが第2のストレージデバイスにアクセスし、ファイルシステム内のサブディレクトリを作成し問い合わせる例が説明のために使用される。しかしながら、実際の適用では、第1のクライアントも、第1のストレージデバイスにアクセスすることによってサブディレクトリを作成し問い合わせし得る。

【0097】

サブディレクトリが見つけれられた後、または新しいサブディレクトリが作成された後、第1のクライアントまたは第2のクライアントは、サブディレクトリのパラメータ情報を取得し得、次に、サブディレクトリのパラメータ情報に基づいてサブディレクトリ内にファイルを作成し得る。以下は、ユーザが第1のクライアントを使用して第1のストレージデバイスにアクセスし、サブディレクトリ内にファイルを作成するプロセスについて説明する。詳細は図7に示されている。

【0098】

ステップS701：クライアントは、ファイル生成要求を物理ノードAに送信する。

【0099】

ファイル生成要求は、サブディレクトリのパラメータ情報およびファイル名を搬送する。

【0100】

図5または図6に示されているように、物理ノードAは、サブディレクトリのパラメータ情報をクライアントに送信している。したがって、サブディレクトリ内にファイルを作成する必要があるとき、クライアントは、サブディレクトリのパラメータ情報およびファイルのファイル名をファイル生成要求に追加し得る。

10

20

30

40

50

## 【0101】

ステップS702：ファイル生成要求を受信した後、物理ノードAは、サブディレクトリのパラメータ情報に基づいて、サブディレクトリが属する物理ノードDを決定する。

## 【0102】

サブディレクトリが属する物理ノードDを決定する方法は、図6のステップS602と同じである。ここでは詳細は再び説明されない。

## 【0103】

ステップS703：物理ノードAは、サブディレクトリのパラメータ情報およびファイル名を物理ノードDに送信する。

## 【0104】

ステップS704：物理ノードDは、ファイルが作成されているかどうかを決定する。

## 【0105】

物理ノードDは、サブディレクトリのパラメータ情報内のShard IDおよびサブディレクトリ名に基づいてサブディレクトリを見つけ、次に、サブディレクトリに対応するDETを見つけ、ファイル名を求めてDETを検索する。ファイル名が存在する場合、それは、同じファイル名を有するファイルが作成されていることを示し、ステップS705が実行される。ファイル名が存在しない場合、それは、サブディレクトリ内にファイルが作成されることができることを示し、ステップS706が実行される。

## 【0106】

ステップS705：ノードDは、ファイル名が作成されていることを示すフィードバックをノードAに送信し、ノードAは、フィードバックを第1のクライアントにさらにフィードバックする。

## 【0107】

フィードバックメッセージを受信した後、第1のクライアントは、通知メッセージを使用して、同じファイル名を有するファイルが既に存在することをユーザにさらに通知し得、ユーザは、通知情報に基づいてさらなる操作を実行し得る、例えば、ファイル名を変更し得る。

## 【0108】

ステップS706：ノードDはファイルを作成する。

## 【0109】

ファイルを作成するとき、ノードDは、ファイルのパラメータ情報を設定し、例えば、shard IDを割り振り、ファイル識別情報を割り振り、shard IDおよびファイル識別情報をサブディレクトリのDETに追加する。図5のステップS503で説明されたように、サブディレクトリが作成されるとき、サブディレクトリに対してInodeテーブルが生成され、Inodeテーブルは、サブディレクトリの下に生成されたファイルに関する情報を記録するために使用される。したがって、このステップでは、ノードDがファイルを作成した後、ファイルに関する情報がサブディレクトリのInodeテーブルに追加される。ファイル情報は、ファイル長、ファイルに対するユーザの操作許可、およびファイルの修正時点などの情報を含む。

## 【0110】

ステップS707：物理ノードDは、ファイルのパラメータ情報をフィードバックする。

## 【0111】

物理ノードDは、最初にフィードバック情報をノードAに送信し、ノードAは、フィードバック情報を第1のクライアントにさらにフィードバックする。

## 【0112】

ステップS702において、物理ノードAが、サブディレクトリのホームノードは物理ノードAであると決定したとき、物理ノードAは、ステップS704からS707が実行する。

## 【0113】

ここで、図5で生成されたサブディレクトリおよび図7で生成されたファイルは、図3Bで設定されたバックアップポリシーに従って、対応するバックアップノードにバックアップ

10

20

30

40

50

プされることに留意されたい。

【0114】

ファイルが作成された後、ユーザは、ファイルにデータを書き込み得る。ユーザは、第1のストレージデバイスに接続された第1のクライアントと、第2のストレージデバイスに接続された第2のクライアントとを使用して、ファイルにデータを書き込み得る。以下は、説明のための例として、ユーザが第1のクライアントを使用して第1のストレージデバイスにアクセスし、ファイルにデータを書き込むプロセスを使用する。詳細は図8に示されている。

【0115】

ステップS801：物理ノードAは、ファイルの書き込み要求を受信する。

10

【0116】

本出願のこの実施形態では、任意のノードがファイルシステムを記憶するため、ユーザは、任意のノードに接続されたクライアントを使用してファイルシステム内のファイルにアクセスし得る。

【0117】

書き込み要求はファイルのアドレス情報を搬送し、アドレス情報は、ファイルのパラメータ情報、オフセットアドレス、および書き込まれるデータを含む。本出願のこの実施形態では、ファイルのパラメータ情報はファイルのハンドルであり、ファイルシステム識別子、ファイル識別子、およびshard IDを含む。

【0118】

20

ステップS802：物理ノードAは、書き込み要求に基づいてファイルのホームノードDを決定する。

【0119】

ファイルのshard IDに基づいて、ファイルを記録するホームノードDを決定する方法については、図6のステップS602を参照されたい。ここでは詳細は再び説明されない。

【0120】

ステップS803：物理ノードAは、書き込み要求を物理ノードDに転送する。

【0121】

ステップS804：物理ノードDは、ファイルシステムへのアクセスを、ファイルシステムに対応する仮想ボリュームへのアクセスに変換する。

30

【0122】

各物理ノードは、ファイルシステムのために作成された仮想ボリュームを記録するため、物理ノードDは、書き込み要求内のファイルシステム識別子を仮想ボリュームの識別子に置き換える。

【0123】

ステップS805：物理ノードDは、書き込み要求内のファイル識別子およびshard IDに基づいてファイルを見つけ、ファイルに関する情報を更新する。

【0124】

ファイル識別子およびshard IDに基づいてファイルを見つけた後、物理ノードDは、ファイル識別子に含まれるファイルのInode番号に基づいてInodeテーブル内のファイルに対応するinodeエントリを見つけ、例えば、書き込み要求で搬送された書き込まれるデータの長さおよびオフセットアドレスに基づいて、inodeエントリ内のファイルに関する情報を記録し、ファイルの長さおよびオフセットアドレスを更新し、現在の時点をファイルの更新時点として記録する。

40

【0125】

ステップS806：物理ノードDは、予め設定されたバックアップポリシーに基づいて、書き込まれるデータの複数のコピーを書き込む。

【0126】

仮想ファイルクラスタが確立されるとき、ファイルシステムに対してバックアップポリシーが確立される。バックアップポリシーでは、各ノードにバックアップノードが設定さ

50

れる。例えば、図3Bで設定されたバックアップポリシーに従って、物理ノードDのバックアップノードは物理ノードCおよび物理ノードBであると決定され得る。この場合、書き込まれるデータをローカルメモリに書き込むとき、物理ノードDは、書き込まれるデータを物理ノードCおよび物理ノードBに送信し、物理ノードCおよび物理ノードBは、書き込まれるデータを物理ノードCおよび物理ノードBのメモリに書き込む。

【0127】

ステップS807：複数のコピーの書き込みが遂行されたと決定した後、物理ノードDは、書き込み要求が遂行されたことを示すメッセージを第1のクライアントに返す。

【0128】

ステップS808：物理ノードDは、書き込まれるデータを永続的に記憶する。

10

【0129】

図3Bに示されているフラッシングポリシーによれば、ファイルシステムの仮想ボリュームは、ミラーボリュームペア：第1のストレージデバイス内の第1のボリュームおよび第2のストレージデバイス内の第2のボリュームに対応する。予め設定されたメモリエビクションアルゴリズムに基づいて、書き込まれるデータが永続ストレージデバイスにエビクションされる必要がある、すなわちディスクにフラッシュされる必要があると決定したとき、物理ノードDは、最初に、書き込まれるデータ内のアドレスに記録された仮想ボリュームに基づいてフラッシングポリシーから、仮想ボリュームに対応するミラーボリュームペア、すなわち、第1のストレージデバイス内の第1のボリュームおよび第2のストレージデバイス内の第2のボリュームを取得し、次に、第2のストレージデバイスのメモリ内の書き込まれるデータを永続ストレージデバイス701内の第2のボリュームに対応する物理空間に書き込み、次に、書き込まれるデータのメモリアドレスを、第1のストレージデバイス内の物理ノードDに対応するバックアップノードBに送信する。物理ノードBは、メモリアドレスに基づいて、物理ノードBのメモリに記憶された書き込まれるデータを、第1のストレージデバイスの永続ストレージデバイス601内の第1のボリュームに対応する物理空間に書き込む。

20

【0130】

図9は、本出願の一実施形態によるファイル読み出し方法のフローチャートである。

【0131】

本出願のこの実施形態では、ユーザはまた、任意のクライアントを使用してファイルシステム内のファイルにアクセスし得る。この実施形態では、ユーザが第2のクライアントを使用してファイルを読み出す例が説明のために使用される。

30

【0132】

ステップS901：物理ノードCは、ファイルの読み出し要求を受信する。

【0133】

読み出し要求は、ファイルのアドレス情報を搬送する。アドレス情報は、ファイルのパラメータ情報およびオフセットアドレスを含み、パラメータ情報は、ファイルのハンドルであり、ファイルシステム識別子、ファイル識別子、およびshard IDを含む。第2のクライアントが読み出し要求を送信するとき、ファイルのパラメータ情報は、図6に示されている方法に従って取得されている。

40

【0134】

ステップS902：物理ノードCは、読み出し要求に基づいてファイルのホームノードBを決定する。

【0135】

ファイルのホームノードBを決定する方法については、図6のステップS602の説明を参照されたい。ここでは詳細は再び説明されない。

【0136】

ステップS903：物理ノードCは、読み出し要求をホームノードBに転送する。

【0137】

ステップS904：物理ノードBは、ファイルシステムへの読み出し要求のアクセスを、フ

50

ファイルシステムの仮想ボリュームへのアクセスに変換する。

【0138】

ステップS905：物理ノードBは、読み出し要求内のアドレスに基づいて、物理ノードBのメモリからファイルを読み出す。

【0139】

ステップS906：物理ノードBはファイルを返す。

【0140】

ステップS907：ファイルがメモリ内にないとき、物理ノードBは、フラッシングポリシー内の仮想ボリュームに対応する第1のストレージデバイス内の第1のボリュームに基づいて永続ストレージデバイス601からファイルを読み出し、ファイルを物理ノードCに返し、次に、物理ノードCは、ファイルを第2のクライアントに返す。

10

【0141】

本出願の実施形態では、ファイルおよびディレクトリにアクセスするとき、第1のストレージデバイスおよび第2のストレージデバイスは、shard IDに基づいてファイルおよびディレクトリのホームノードにアクセス要求を転送する。これは、クロスデバイスデータアクセスをもたらし、アクセス効率にさらに影響を及ぼす。本出願のこの実施形態で提供される可能な実施態様では、第1のストレージデバイスと第2のストレージデバイスとの両方がピアエンドのデータをバックアップするため、ピアエンドのデータにアクセスするためのアクセス要求が受信されたとき、アクセスされる必要があるデータは、ローカルエンドによってバックアップされた、ピアエンドのバックアップデータから取得され得、アクセスされるデータはピアエンドから取得される必要がない。これはデータアクセス効率を改善する。

20

【0142】

アクティブ-アクティブモードのアクティブ-アクティブストレージシステム内の1つのストレージデバイスが障害を起こしたとき、障害を起こしたストレージデバイスのサービスは、バックアップデータを使用して引き継がれ得る。図10に示されているように、第1のストレージデバイスと第2のストレージデバイスとの間のリンクが切断された後、または第2のストレージデバイスが障害を起こした後、第2のストレージデバイスのサービスは、第1のストレージデバイスに記憶された第2のストレージデバイスのバックアップデータを使用して引き継がれ得る。以下は、説明のために、第1のストレージデバイスと第2のストレージデバイスとの間のリンクが切断される例を使用する。詳細は、図11に示されているフローチャートに示されている。

30

【0143】

ステップS111：第1のストレージデバイスおよび第2のストレージデバイスは、ピアエンドのハートビートを同時に検出する。

【0144】

ステップS112：ピアエンドのハートビートが検出されないとき、第1のストレージデバイスおよび第2のストレージデバイスはそれぞれ、実行されているサービスを中断する。

【0145】

サービスを中断することは、実行されているアクセス要求を停止することを指す。

40

【0146】

ステップS113：第1のストレージデバイスおよび第2のストレージデバイスは、グローバルビューおよびファイルシステムを修正する。

【0147】

ピアエンドのハートビートが検出されないとき、第1のストレージデバイスおよび第2のストレージデバイスは、ピアエンドのサービスを引き継ぐ準備をし、グローバルビューおよびファイルシステムを修正し、グローバルビュー内のピアエンドの物理ノードに対応する仮想ノードをグローバルビューから削除し、バックアップポリシー内のピアエンドのバックアップノードを削除する必要がある。例えば、第1のストレージデバイスは、グローバルビューを（Vnode A, Vnode B）に修正し、第2のストレージデバイスは、グローバ

50

ルビューを（Vnode C，Vnode D）に修正する。加えて、ファイルシステムのshardビュー内のピアノードに対応する仮想ノードのshardは、ローカルノードに対応する仮想ノードに対応するshardに修正される。例えば、第1のストレージデバイス内のshardビューは、Vnode A [0，2047] およびVnode B [2048，4095] に修正され、第2のストレージデバイス内のshardビューは、Vnode C [0，2047] およびVnode D [2048，4095] に修正され、フラッシングポリシー内のピアノードのボリュームは削除される。

【0148】

ステップS114：第1のストレージデバイスおよび第2のストレージデバイスはそれぞれ、調停要求をクォーラムデバイスに送信する。

【0149】

ステップS115：クォーラムデバイスは、調停によって、第1のストレージデバイスがサービスを引き継ぐと決定する。

【0150】

クォーラムデバイスは、調停要求を受信する順序に基づいて、サービスを引き継ぐデバイスを決定し得る。例えば、最初に受信されたクォーラム要求に対応するストレージデバイスが、サービスを引き継ぐデバイスとして機能する。

【0151】

ステップS116：クォーラムデバイスは、調停結果を第1のストレージデバイスおよび第2のストレージデバイスに別々に通知する。

【0152】

ステップS117：通知を受信した後、第2のストレージデバイスは、第2のクライアントから切断される、すなわち、サービスの実行を停止する。

【0153】

ステップS118：通知を受信した後、第1のストレージデバイスは、第2のストレージアレイのIPアドレスを第1のストレージデバイスにドリフトさせ、第2のクライアントとの接続を確立する。

【0154】

ステップS119：第2のストレージアレイのバックアップデータを使用して第2のストレージアレイのサービスを引き継ぐ。

【0155】

第2のストレージアレイのバックアップデータは第1のストレージデバイスに記憶されているため、第2のストレージデバイス内のデータへのアクセスを受信したとき、第1のストレージデバイスは、第1のクライアントおよび第2のクライアントがリンクの中断を知覚しないように、アクセス要求内のshard IDを使用して、第2のストレージデバイス内のデータに対する第1のクライアントおよび第2のクライアントのアクセス要求を特定して、バックアップデータにアクセスすることができる。

【0156】

データアクセスプロセスでは、バックアップポリシーおよびフラッシングポリシーが変更されているため、書き込みデータは、第1のストレージデバイスのノードのメモリにのみ書き込まれ、第1のストレージデバイスのボリュームにのみ記憶される。

【0157】

本出願の実施形態で提供される解決策は上記で説明されている。本出願の原理および実施態様は、本明細書の具体的な例を通して説明されている。本出願の実施形態の説明は、本出願の方法および中核となるアイデアの理解を助けるために提供されているにすぎない。加えて、当業者は、本出願のアイデアに従って、具体的な実施態様および適用範囲に関して本出願に変形および修正を加えることができる。したがって、本明細書の内容は、本出願に対する限定として解釈されてはならない。

【符号の説明】

【0158】

100 第1のストレージデバイス

10

20

30

40

50

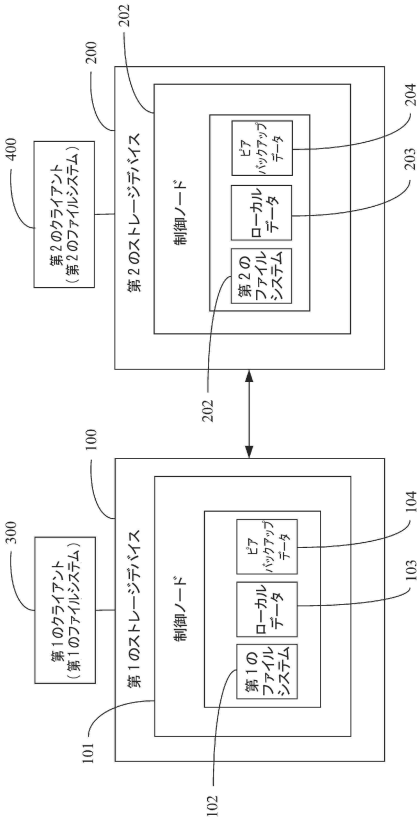
|     |                 |    |
|-----|-----------------|----|
| 101 | 制御ノード           |    |
| 102 | 第1のファイルシステム     |    |
| 103 | ローカルデータ         |    |
| 104 | ピアバックアップデータ     |    |
| 200 | 第2のストレージデバイス    |    |
| 201 | 制御ノード           |    |
| 202 | 第2のファイルシステム     |    |
| 203 | ローカルデータ         |    |
| 204 | ピアバックアップデータ     |    |
| 300 | 第1のクライアント       | 10 |
| 400 | 第2のクライアント       |    |
| 500 | アクティブ-アクティブシステム |    |
| 600 | 第1のストレージデバイス    |    |
| 601 | 永続ストレージデバイス     |    |
| 602 | プロセッサ           |    |
| 603 | メモリ             |    |
| 604 | グローバルビュー        |    |
| 605 | ファイルシステム        |    |
| 606 | キャッシュデータ        |    |
| 607 | バックアップデータ       | 20 |
| 608 | 第1のファイルシステム     |    |
| 609 | 第1のボリューム        |    |
| 700 | 第2のストレージデバイス    |    |
| 701 | 永続ストレージデバイス     |    |
| 702 | 第2のファイルシステム     |    |
| 703 | 第2のボリューム        |    |
| 800 | 第1のクライアント       |    |
| 900 | 第2のクライアント       |    |

30

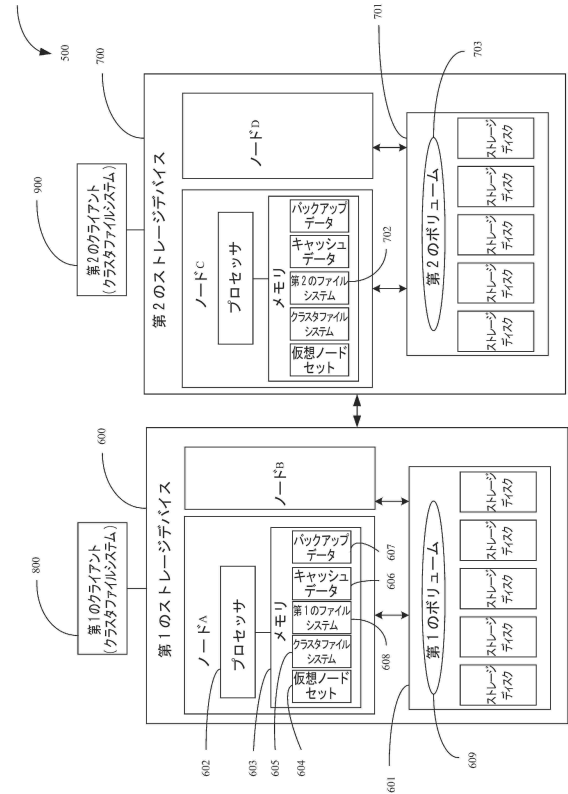
40

50

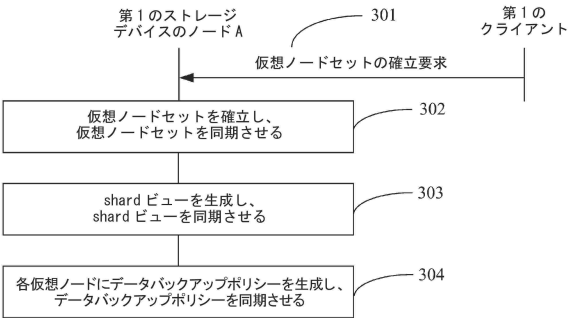
【図面】  
【図 1】



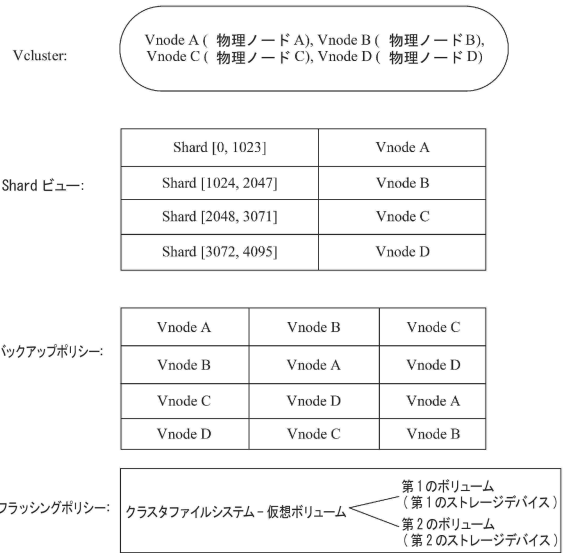
【図 2】



【図 3 A】



【図 3 B】



10

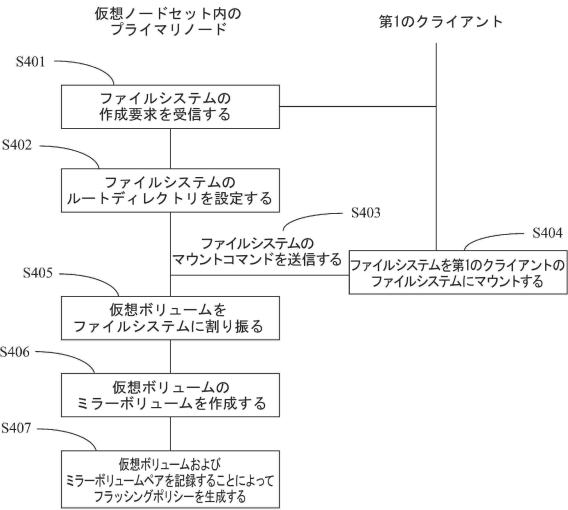
20

30

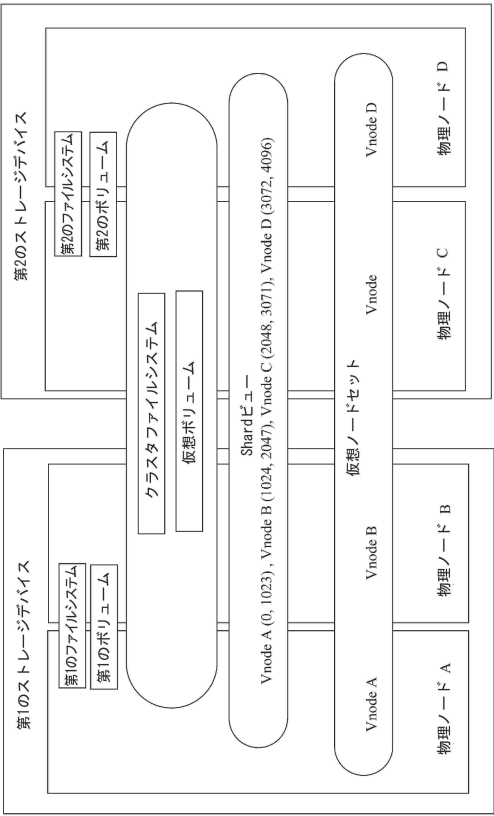
40

50

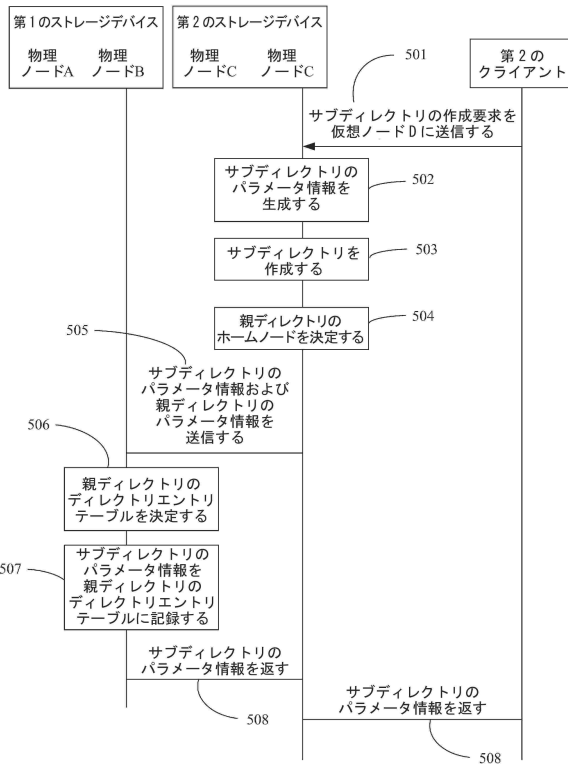
【図 4 A】



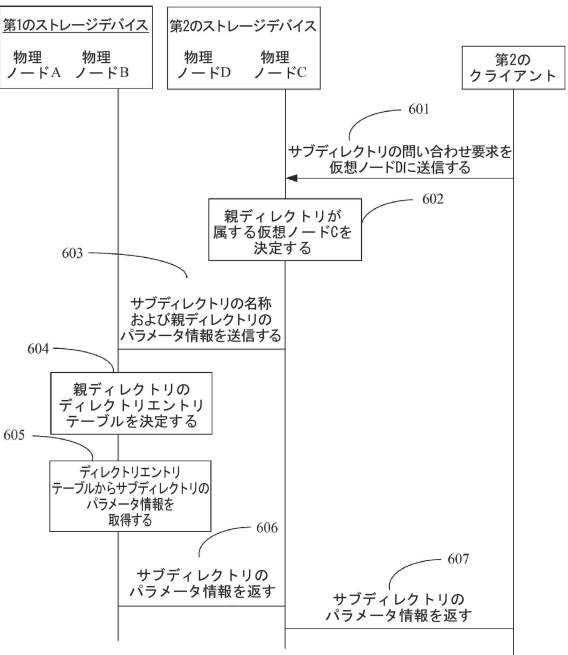
【図 4 B】



【図 5】



【図 6】



10

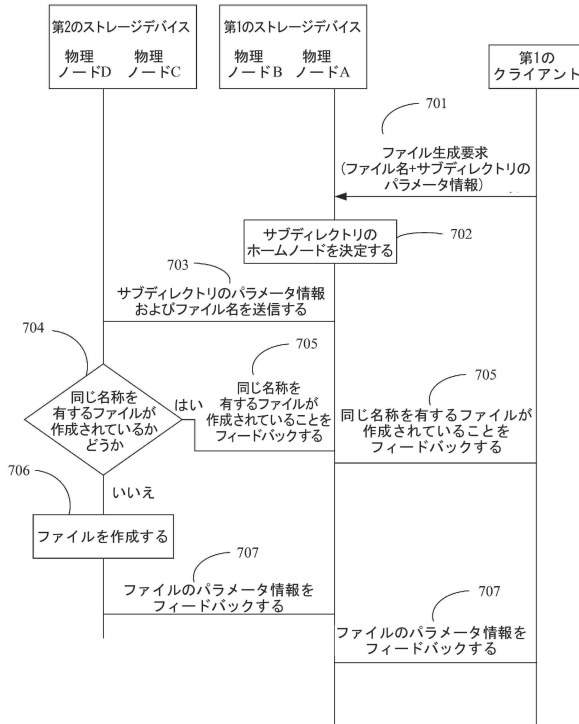
20

30

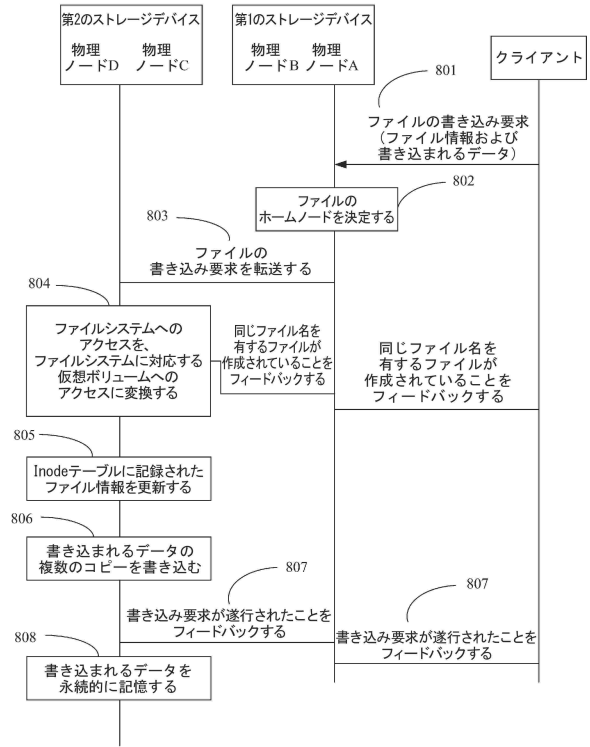
40

50

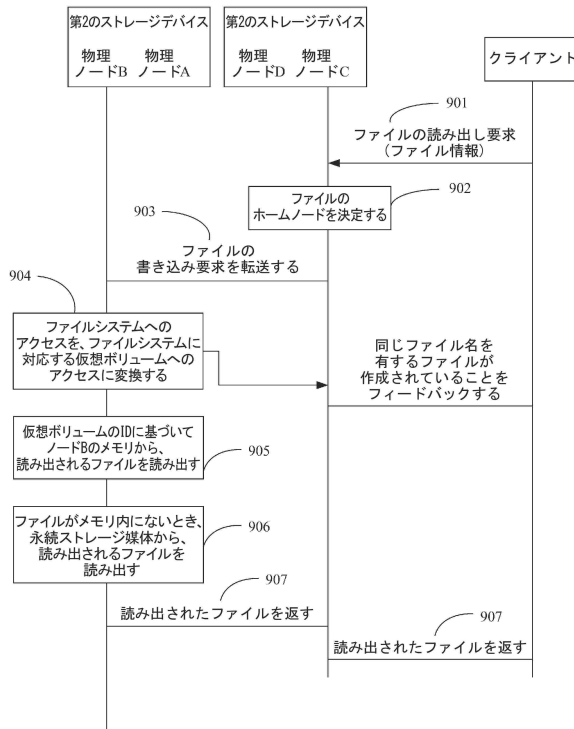
【図 7】



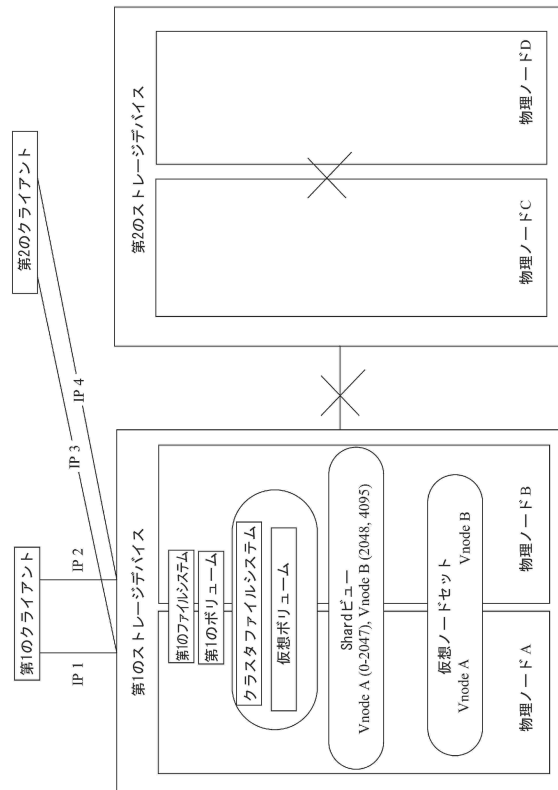
【図 8】



【図 9】



【図 10】



10

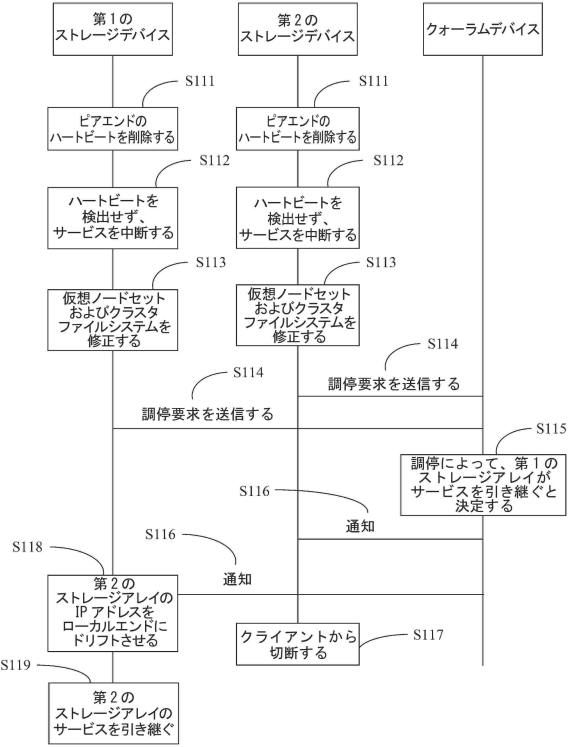
20

30

40

50

【図 11】



10

20

30

40

50

## フロントページの続き

(33)優先権主張国・地域又は機関

中国(CN)

(74)代理人 100133569

弁理士 野村 進

(72)発明者 杜 翔

中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍岗区坂田 華為総部▲ベン▼公楼

(72)発明者 翁 宇佳

中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍岗区坂田 華為総部▲ベン▼公楼

(72)発明者 李 小▲華▼

中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍岗区坂田 華為総部▲ベン▼公楼

(72)発明者 ▲張▼ ▲鵬▼

中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍岗区坂田 華為総部▲ベン▼公楼

審査官 田名網 忠雄

(56)参考文献 米国特許第09880777 (US, B1)

米国特許出願公開第2014/0310278 (US, A1)

特開2007-328778 (JP, A)

国際公開第2015/198449 (WO, A1)

(58)調査した分野 (Int.Cl., DB名)

G 0 6 F 1 6 / 1 0 - 1 6 / 1 8 8

G 0 6 F 1 1 / 2 0

G 0 6 F 3 / 0 6 - 3 / 0 8

G 0 6 F 1 3 / 1 0 - 1 3 / 1 4