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

特許第5320678号
(P5320678)

(45)発行日　平成25年10月23日(2013.10.23)　(24)登録日　平成25年7月26日(2013.7.26)

(51)Int.Cl.　F I
G06F 12/00 (2006.01)　G06F 12/00 531D
G06F 3/06 (2006.01)　G06F 12/00 501B
G06F 3/06 540
G06F 3/06 304P
G06F 3/06 304F

請求項の数 39　(全 46 頁)

(21)出願番号　特願2007-40093(P2007-40093)
(22)出願日　平成19年2月20日(2007.2.20)
(65)公開番号　特開2008-204206(P2008-204206A)
(43)公開日　平成20年9月4日(2008.9.4)
審査請求日　平成22年1月19日(2010.1.19)

(73)特許権者　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100103090
弁理士　岩壁　冬樹
(74)代理人　100124501
弁理士　塩川　誠人
(72)発明者　大和　純一
東京都港区芝五丁目７番１号　日本電気株式会社内

審査官　池田　聡史

最終頁に続く

(54)【発明の名称】データ分散格納システム及びデータ分散方法、それに用いる装置並びにそのプログラム

(57)【特許請求の範囲】
【請求項１】
コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群における複数の記憶装置に格納されるデータ分散格納システムであって、
分割データの複製数を決定する複製数計画手段と、
コンテンツを複数の分割データに分割し、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する分割データ管理手段と、
前記分割データに対応する複製データの格納先を示す複製管理情報を記憶する複製管理情報記憶手段と、
前記複製管理情報に基づいて、分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定するアクセス先決定手段とを備え、
前記複製数計画手段は、記憶装置に対するアクセスが不能になったときの複製データの他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定し、
前記分割データ管理手段が、記憶装置が追加されたときに、前記記憶装置群における複製データの再配置を行う
ことを特徴とするデータ分散格納システム。
【請求項２】
アクセス先決定手段は、少なくともコンテンツの一部を読み出すホスト端末に対し、前

記少なくともコンテンツの一部を構成する分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定し、通知する
請求項１に記載のデータ分散格納システム。
【請求項３】
アクセス先決定手段は、少なくともコンテンツの一部を読み出すホスト端末で、前記少なくともコンテンツの一部を構成する分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定する
請求項１または請求項２に記載のデータ分散格納システム。
【請求項４】
アクセス先決定手段は、分割データに対応する複数の複製データが格納されている複数の記憶装置から、乱数に基づいてランダムに少なくとも１つの記憶装置を決定する
請求項１から請求項３のうちのいずれか１項に記載のデータ分散格納システム。
【請求項５】
アクセス先決定手段は、分割データに対応する複数の複製データが格納されている複数の記憶装置のうち負荷の低い記憶装置をアクセス先として決定する
請求項１から請求項３のうちのいずれか１項に記載のデータ分散格納システム。
【請求項６】
複製管理情報記憶手段は、前記分割データ管理手段によって決定された複製データの格納先を示す複製管理情報を記憶する
請求項１から請求項５のうちのいずれか１項に記載のデータ分散格納システム。
【請求項７】
分割データ管理手段は、分割データに対応する複数の複製データが各々記憶装置に均等化されるように、複製データの格納先を決定する
請求項１から請求項６のうちのいずれか１項に記載のデータ分散格納システム。
【請求項８】
分割データ管理手段は、コンテンツにおける連続する所定数の分割データに対応する複数の複製データが分散配置されるように、複製データの配置先を決定する
請求項１から請求項７のうちのいずれか１項に記載のデータ分散格納システム。
【請求項９】
分割データ管理手段は、２つの記憶装置間で共有する同じ分割データに対応する複製データの数を共有数とした場合に、各々の記憶装置の組み合わせにおける共有数が均等化されるように、複製データの配置先を決定する
請求項１から請求項８のうちのいずれか１項に記載のデータ分散格納システム。
【請求項１０】
分割データ管理手段は、複製データの配置先を決定する際に、該複製データと同じ分割データに対応する複製データの配置先が既に決定している場合に、配置先として決定済みの記憶装置を除いた記憶装置の中から、前記決定済み記憶装置との組み合わせにおける共有数が小さい記憶装置を該複製データの配置先として決定する
請求項９に記載のデータ分散格納システム。
【請求項１１】
分割データ管理手段は、記憶装置が使用不能となった場合に、複製管理情報記憶手段から、前記使用不能となった記憶装置を格納先として示している情報を削除する
請求項１から請求項１０のうちのいずれか１項に記載のデータ分散格納システム。
【請求項１２】
分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理および記憶装置からの複製データの削除処理を行う複製処理手段を備え、
分割データ管理手段は、使用不能となった記憶装置が配置先として割り当てられている複製データについて、新たな配置先を決定するとともに、該複製データと同じ分割データに対応する複製データを格納している記憶装置の中から１つの記憶装置をコピー元記憶装

置に決定して、前記複製処理手段に複製データのコピー処理を行わせる
請求項１から請求項１１のうちのいずれか１項に記載のデータ分散格納システム。
【請求項１３】
分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理および記憶装置からの複製データの削除処理を行う複製処理手段を備えたデータ分散格納システムであって、
分割データ管理手段は、所定のタイミングで、複製管理情報に基づき、複製データについて新たな配置先を決定して、前記複製処理手段に複製データのコピー処理および前の配置先からの削除処理を行わせる
請求項１から請求項１２のうちのいずれか１項に記載のデータ分散格納システム。
【請求項１４】
分割データ管理手段は、複製数計画手段によって決定された分割データの複製数に応じて、該分割データに対応する複製データの配置先を決定する
請求項１から請求項１３のうちのいずれか１項に記載のデータ分散格納システム。
【請求項１５】
複製数計画手段は、コンテンツにおける分割データの位置に基づいて、各分割データの複製数を決定する
請求項１４に記載のデータ分散格納システム。
【請求項１６】
複製数計画手段は、コンテンツの先頭に位置するデータを含む分割データ、およびコンテンツにチャプターが付与されている場合に、各チャプターの先頭に位置するデータを含む分割データの複製数を、該コンテンツにおける他の分割データの複製数よりも多くする
請求項１５に記載のデータ分散格納システム。
【請求項１７】
複製数計画手段は、コンテンツのアクセス要求予測量またはアクセス要求量と、各記憶装置の処理能力と、システムで規定した安全係数とに基づいて、分割データの複製数を決定する
請求項１から請求項１６のうちのいずれか１項に記載のデータ分散格納システム。
【請求項１８】
分割データ管理手段からの指示に従い記憶装置間の複製データのコピー処理および記憶装置からの複製データの削除処理を行う複製処理手段を備えたデータ分散格納システムであって、
分割データ管理手段は、複製数計画手段によって決定された分割データの複製数と、記憶装置に格納されている該分割データに対応する複製データの数とが一致するように、前記分割データに対応する複製データの新たな配置先または削除対象とする配置先を決定し、前記複製処理手段に該複製データのコピー処理または削除処理を行わせる
請求項１から請求項１７のうちのいずれか１項に記載のデータ分散格納システム。
【請求項１９】
コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群における複数の記憶装置に格納されるデータ分散格納システムに適用され、少なくともコンテンツの一部を読み出すホスト端末に対し、前記コンテンツを構成する分割データへのアクセス先を通知する分割データ管理装置であって、
分割データの複製数を決定する複製数計画手段と、
コンテンツを複数の分割データに分割し、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する分割データ管理手段と、
前記ホスト端末のコンテンツ読み出し範囲に含まれる分割データに対応する複製データの配置先を示す複製管理情報に基づいて、前記分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定するアクセス先決定手段とを備え、
前記複製数計画手段は、記憶装置に対するアクセスが不能になったときの複製データの

他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定し、
前記分割データ管理手段が、記憶装置が追加されたときに、前記記憶装置群における複製データの再配置を行う
ことを特徴とする分割データ管理装置。
【請求項２０】
コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群における複数の記憶装置に格納されるデータ分散格納システムに適用されるデータ分散方法であって、
分割データ管理手段が、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定し、
複製管理情報記憶手段が、前記分割データに対応する複製データの格納先を示す複製管理情報を記憶し、
アクセス先決定手段が、前記複製管理情報に基づいて、分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定し、
複製数計画手段が、記憶装置に対するアクセスが不能になったときの複製データの他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定し、
前記分割データ管理手段が、記憶装置が追加されたときに、前記記憶装置群における複製データの再配置を行う
ことを特徴とするデータ分散方法。
【請求項２１】
アクセス先決定手段が、少なくともコンテンツの一部を読み出すホスト端末に対し、前記少なくともコンテンツの一部を構成する分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定し、通知する
請求項２０に記載のデータ分散方法。
【請求項２２】
アクセス先決定手段が、少なくともコンテンツの一部を読み出すホスト端末で、前記少なくともコンテンツの一部を構成する分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定する
請求項２０または請求項２１に記載のデータ分散方法。
【請求項２３】
アクセス先決定手段が、分割データに対応する複数の複製データが格納されている複数の記憶装置から、乱数に基づいてランダムに少なくとも１つの記憶装置を決定する
請求項２０から請求項２２のうちのいずれか１項に記載のデータ分散方法。
【請求項２４】
アクセス先決定手段が、分割データに対応する複数の複製データが格納されている複数の記憶装置のうち負荷の低い記憶装置をアクセス先として決定する
請求項２０から請求項２３のうちのいずれか１項に記載のデータ分散方法。
【請求項２５】
複製管理情報記憶手段が、前記分割データ管理手段によって決定された複製データの格納先を示す複製管理情報を記憶する
請求項２０から請求項２４のうちのいずれか１項に記載のデータ分散方法。
【請求項２６】
分割データ管理手段が、分割データに対応する複数の複製データが各々記憶装置に均等化されるように、複製データの格納先を決定する
請求項２０から請求項２５のうちのいずれか１項に記載のデータ分散方法。
【請求項２７】
分割データ管理手段が、コンテンツにおける連続する所定数の分割データに対応する複数の複製データが分散配置されるように、複製データの配置先を決定する

請求項２０から請求項２６のうちのいずれか１項に記載のデータ分散方法。
【請求項２８】
分割データ管理手段が、２つの記憶装置間で共有する同じ分割データに対応する複製データの数を共有数とした場合に、各々の記憶装置の組み合わせにおける共有数が均等化されるように、複製データの配置先を決定する
請求項２０から請求項２７のうちのいずれか１項に記載のデータ分散方法。
【請求項２９】
分割データ管理手段が、複製データの配置先を決定する際に、該複製データと同じ分割データに対応する複製データの配置先が既に決定している場合に、配置先として決定済みの記憶装置を除いた記憶装置の中から、前記決定済み記憶装置との組み合わせにおける共有数が小さい記憶装置を該複製データの配置先として決定する
請求項２８に記載のデータ分散方法。
【請求項３０】
分割データ管理手段が、記憶装置が使用不能となった場合に、複製管理情報記憶手段から、前記使用不能となった記憶装置を格納先として示している情報を削除する
請求項２０から請求項２９のうちのいずれか１項に記載のデータ分散方法。
【請求項３１】
分割データ管理手段が、使用不能となった記憶装置が配置先として割り当てられている複製データについて、新たな配置先を決定するとともに、該複製データと同じ分割データに対応する複製データを格納している記憶装置の中から１つの記憶装置をコピー元記憶装置に決定し、
複製処理手段が、分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理を行う
請求項２０から請求項３０のうちのいずれか１項に記載のデータ分散方法。
【請求項３２】
分割データ管理手段が、所定のタイミングで、複製管理情報に基づき、複製データについて新たな配置先を決定し、
複製処理手段が、分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理および前の配置先からの削除処理を行う
請求項２０から請求項３１のうちのいずれか１項に記載のデータ分散方法。
【請求項３３】
分割データ管理手段が、複製数計画手段によって決定された分割データの複製数に応じて、該分割データに対応する複製データの配置先を決定する
請求項２０から請求項３２のうちのいずれか１項に記載のデータ分散方法。
【請求項３４】
複製数計画手段が、コンテンツにおける分割データの位置に基づいて、各分割データの複製数を決定する
請求項３３に記載のデータ分散方法。
【請求項３５】
複製数計画手段が、コンテンツの先頭に位置するデータを含む分割データ、およびコンテンツにチャプターが付与されている場合に、各チャプターの先頭に位置するデータを含む分割データの複製数を、該コンテンツにおける他の分割データの複製数よりも多くする
請求項３４に記載のデータ分散方法。
【請求項３６】
複製数計画手段が、コンテンツのアクセス要求予測量またはアクセス要求量と、各記憶装置の処理能力と、システムで規定した安全係数とに基づいて、分割データの複製数を決定する
請求項２０から請求項３５のうちのいずれか１項に記載のデータ分散方法。
【請求項３７】
分割データ管理手段が、複製数計画手段によって決定された分割データの複製数と、記

憶装置に格納されている該分割データに対応する複製データの数とが一致するように、前記分割データに対応する複製データの新たな配置先または削除対象とする配置先を決定し、

複製処理手段が、分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理および記憶装置からの複製データの削除処理を行う

請求項２０から請求項３６のうちのいずれか１項に記載のデータ分散方法。

【請求項３８】

コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群における複数の記憶装置に格納されるデータ分散格納システムにおいて、少なくともコンテンツの一部を読み出すホスト端末に対し、前記コンテンツを構成する分割データへのアクセス先を通知する分割データ管理装置に適用されるデータ分散用プログラムであって、

コンピュータに、

コンテンツを複数の分割データに分割し、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する処理、

前記ホスト端末のコンテンツ読み出し範囲に含まれる分割データに対応する複製データの配置先を示す複製管理情報に基づいて、前記分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定する処理、

記憶装置に対するアクセスが不能になったときの複製データの他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定する処理、および、

記憶装置が追加されたときに、前記記憶装置群において複製データを再配置する処理

を実行させるためのデータ分散用プログラム。

【請求項３９】

コンピュータに、

コンテンツを複数の分割データに分割し、少なくとも１つの分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する処理

を実行させる請求項３８に記載のデータ分散用プログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、データ分散格納システム、データ分散方法、分割データ管理装置、およびデータ分散用プログラムに関し、特に、複数の記憶装置にデータを分散させて格納するデータ分散格納システム、データ分散方法、分割データ管理装置、およびデータ分散用プログラムに関する。

【背景技術】

【０００２】

あるデータを複数に分割し、複数の記憶装置に分散させるデータ分散方法に関し、ストリーミングを再生するために複数のコンテンツを形成する各データを分割して複数の記憶装置に格納させる方法がある。例えば、特許文献１には、一連のデータストリームとして処理されて意味をなすデータを、複数のファイル（対応する磁気ディスク装置）に跨り連続的に分割して書き込むデータ制御方法が記載されている。また、例えば、特許文献２には、ファイル上の連続するファイルブロックをそれぞれ別の物理ブロック群に割り付けられるよう配置制御するディスクアレイ装置が記載されている。また、例えば、特許文献３には、予め仮想アドレスの指定順序が明らかになっている場合に、それらが各記憶装置に対する均一なアクセスとなり、複数のクライアントによる記憶装置群へのアクセスが同一の記憶装置に重ならないように、分割したデータの配置先を決定するディスクアレイ装置が記載されている。

【０００３】

また、非特許文献１に記載されているように、データを分散させるだけでなく、データ

の冗長化を行う方法も考えられている。非特許文献１には、ＲＡＩＤ１＋０やＲＡＩＤ０＋１として、ブロックストレージのレベルでデータの分散および冗長化を行う方法が開示されている。また、特許文献４には、複製したデータブロックをアドレスに応じて均等に他の記憶装置に分散させて格納する方法が開示されている。

【０００４】

また、特許文献５に記載されているように、アクセス負荷が限界値に達する前に、コンテンツの複製の作成やコンテンツとその他のコンテンツとで所在の再配置を行う方法も考えられている。

【０００５】

【特許文献１】特開２００２－２４４８９３号公報

【特許文献２】特開平０９－２２３０４９号公報

【特許文献３】特許第３０５２８７７号公報

【特許文献４】特許第２８５３６２４号公報

【特許文献５】特開平１１－０８５６０４号公報

【非特許文献１】John L. Hennessy, David A. Paterson, "Computer Architecture:A Quantitative Approach", 3rd Edition, Morgan Kaufmann Pub, 2001, pp.707

【発明の開示】

【発明が解決しようとする課題】

【０００６】

しかしながら、特許文献１～３に記載されている方法は、データを複数に分割して複数の記憶装置に分散させて格納することによってスループットを上げることはできるが、障害時の対応については何ら考慮がされていない。なお、特許文献２に記載されている方法では、論理ブロックと物理ブロックの対応関係を予め設定しておかなければならず、一度ディスクアレイを構成するとハードディスクの追加が容易ではないという問題もある。

【０００７】

また、非特許文献１に記載されているようなブロックベースの分散では、冗長化することで障害時の耐久性（信頼性）の向上という点では効果があるが、スループットが向上するとは限らない。その理由は、ブロックベースの分散では、データがコンテンツとしては認識されていないため、同一コンテンツに関するアクセスの並列性を生かそうということが考えられていないためである。

【０００８】

例えば、ブロックベースの分散を行うストレージ上にファイルシステムを構築した場合、ファイルシステムは仮想的なブロックデバイスのどこが実際のブロックデバイスに割り当てられているかを認識していない。このため、一つのコンテンツ（ファイル）を複数の仮想的なブロックデバイスに割り当てたとしても、複数の物理的なブロックデバイスに割り当てられる保障はない。

【０００９】

また、特許文献４に記載されている方法は、障害時の対応についても考慮されているが、通常用の記憶装置と障害用の記憶装置とを分けて扱っているため、記憶装置の容量効率がよくない。また、ある記憶装置に対応するデータ（すなわち、その記憶装置が記憶しているデータと同一内容のデータ）が、他の一つの記憶装置に記憶されているため、障害時やその回復時にアクセス性能が維持できないという問題がある。例えば、ある記憶装置が故障している間、その記憶装置が記憶していたデータのアクセス負荷が、同じデータを記憶している一つの記憶装置にだけかかることになる。また、故障した記憶装置の代わりに新しい記憶装置を追加した際には、コピー処理が完了するまでの間、コピー処理の読み出し元における負荷がその記憶装置にだけかかることになり、スループットを制限しつつコピー処理を行うとしても、コピー処理の時間がかかる分、その記憶装置のアクセス性能は長い時間低下することとなる。

【００１０】

さらに、特許文献４に記載されている方法では、アドレスを元にデータ（複製されたデ

ータも含む）の配置を決定しているので、一度ディスクアレイを構成するとハードディスクの追加が容易ではないという問題もある。

【００１１】

また、特許文献５に記載されている方法は、デマンドに対応させることで、再配置後にスループットを向上させることはできるが、複製の追加や再配置の処理の際にアクセス性能を維持しようという点は、何ら考慮がされていない。

【００１２】

すなわち、従来のデータ分散方法の問題点は、第１に、容量効率を下げることなく信頼性の向上とスループットの向上とを同時に充足することができないことである。第２に、障害時や複製の追加時や再配置時等、通常時以外でのアクセス性能が維持できないことである。第３に、記憶装置の追加等、スケーラビリティの向上が図れられていないという点である。

【００１３】

そこで、本発明は、以上の問題を解決すべく、データ分散における特性・性能の向上をさらに図ることを目的とする。具体的には、信頼性の向上とスループットの向上とを同時に充足できるデータ分散格納システム及びデータ分散方法、それに用いる装置並びにそのプログラムを提供することを目的とする。また、通常時以外であってもアクセス性能が維持できるデータ分散格納システム及びデータ分散方法、それに用いる装置並びにそのプログラムを提供することを目的とする。また、記憶装置の追加等、スケーラビリティの向上が図れるようなデータ分散格納システム及びデータ分散方法、それに用いる装置並びにそのプログラムを提供することを目的とする。

【課題を解決するための手段】

【００１４】

本発明によるデータ分散格納システムは、コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群における複数の記憶装置に格納されるデータ分散格納システムであって、分割データの複製数を決定する複製数計画手段と、コンテンツを複数の分割データに分割し、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する分割データ管理手段と、分割データに対応する複製データの格納先を示す複製管理情報を記憶する複製管理情報記憶手段（例えば、分割データ管理ＤＢ７）と、複製管理情報に基づいて、分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定するアクセス先決定手段（例えば、分割データ管理部３やホスト１のアクセス先決定機能）とを備え、複製数計画手段は、記憶装置に対するアクセスが不能になったときの複製データの他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定し、分割データ管理手段が、記憶装置が追加されたときに、記憶装置群における複製データの再配置を行うことを特徴とする。

【００１５】

また、アクセス先決定手段は、少なくともコンテンツの一部を読み出すホスト端末に対し、ホスト端末が読み出す少なくともコンテンツの一部を構成する分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定し、通知してもよい。このアクセス先決定手段は、例えば、分割データ管理部３のアクセス先決定機能によって実現される。

【００１６】

また、アクセス先決定手段は、少なくともコンテンツの一部を読み出すホスト端末で、ホスト端末が読み出す少なくともコンテンツの一部を構成する分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定してもよい。このアクセス先決定手段は、例えば、ホスト１のアクセス先決定機能によって実現される。

【００１７】

また、アクセス先決定手段は、分割データに対応する複数の複製データが格納されてい

る複数の記憶装置から、乱数に基づいてランダムに少なくとも１つの記憶装置を決定してもよい。

【００１８】

また、アクセス先決定手段は、分割データに対応する複数の複製データが格納されている複数の記憶装置のうち負荷の低い記憶装置をアクセス先として決定してもよい。

【００１９】

また、複製管理情報記憶手段は、分割データ管理手段によって決定された複製データの格納先を示す複製管理情報を記憶してもよい。

【００２０】

また、分割データ管理手段は、分割データに対応する複数の複製データが各々記憶装置に均等化されるように、複製データの格納先を決定してもよい。

【００２１】

また、分割データ管理手段は、コンテンツにおける連続する所定数の分割データに対応する複数の複製データが分散配置されるように、複製データの配置先を決定してもよい。

【００２２】

また、分割データ管理手段は、２つの記憶装置間で共有する同じ分割データに対応する複製データの数を共有数とした場合に、各々の記憶装置の組み合わせにおける共有数が均等化されるように、複製データの配置先を決定してもよい。

【００２３】

また、分割データ管理手段は、複製データの配置先を決定する際に、その複製データと同じ分割データに対応する複製データの配置先が既に決定している場合に、配置先として決定済みの記憶装置を除いた記憶装置の中から、決定済み記憶装置との組み合わせにおける共有数が小さい記憶装置を複製データの配置先として決定してもよい。

【００２４】

分割データ管理手段は、記憶装置が使用不能となった場合に、複製管理情報記憶手段から、使用不能となった記憶装置を格納先として示している情報を削除してもよい。

【００２５】

また、データ分割可能システムは、分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理および記憶装置からの複製データの削除処理を行う複製処理手段（例えば、複製処理部５）を備え、分割データ管理手段は、使用不能となった記憶装置が配置先として割り当てられている複製データについて、新たな配置先を決定するとともに、該複製データと同じ分割データに対応する複製データを格納している記憶装置の中から１つの記憶装置をコピー元記憶装置に決定して、複製処理手段に複製データのコピー処理を行わせてもよい。

【００２６】

また、分割データ管理手段は、所定のタイミングで、複製管理情報に基づき、複製データについて新たな配置先を決定して、複製処理手段に複製データのコピー処理および前の配置先からの削除処理を行わせてもよい。

【００２７】

また、分割データ管理手段は、複製数計画手段によって決定された分割データの複製数に応じて、該分割データに対応する複製データの配置先を決定してもよい。

【００２８】

また、複製数計画手段は、コンテンツにおける分割データの位置に基づいて、各分割データの複製数を決定してもよい。

【００２９】

また、複製数計画手段は、コンテンツの先頭に位置するデータを含む分割データ、およびコンテンツにチャプターが付与されている場合に、各チャプターの先頭に位置するデータを含む分割データの複製数を、そのコンテンツにおける他の分割データの複製数よりも多くしてもよい。

【００３１】

また、複製数計画手段は、コンテンツのアクセス要求予測量またはアクセス要求量と、各記憶装置の処理能力と、システムで規定した安全係数とに基づいて、分割データの複製数を決定してもよい。

【００３２】

また、分割データ管理手段は、複製数計画手段によって決定された分割データの複製数と、記憶装置に格納されているその分割データに対応する複製データの数とが一致するように、その分割データに対応する複製データの新たな配置先または削除対象とする配置先を決定し、複製処理手段に該複製データのコピー処理または削除処理を行わせてもよい。

【００３３】

また、本発明による分割データ管理装置は、コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群における複数の記憶装置に格納されるデータ分散格納システムに適用され、少なくともコンテンツの一部を読み出すホスト端末に対し、そのコンテンツを構成する分割データへのアクセス先を通知する分割データ管理装置であって、分割データの複製数を決定する複製数計画手段と、コンテンツを複数の分割データに分割し、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する分割データ管理手段と、ホスト端末のコンテンツ読み出し範囲に含まれる分割データに対応する複製データの配置先を示す複製管理情報に基づいて、その分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定するアクセス先決定手段（例えば、分割データ管理部３のアクセス先決定機能）とを備え、複製数計画手段は、記憶装置に対するアクセスが不能になったときの複製データの他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定し、分割データ管理手段が、記憶装置が追加されたときに、記憶装置群における複製データの再配置を行うことを特徴とする。

【００３５】

また、本発明によるデータ分散方法は、コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群に複数の記憶装置に格納されるデータ分散格納システムに適用されるデータ分散方法であって、分割データ管理手段が、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定し、複製管理情報記憶手段が、分割データに対応する複製データの格納先を示す複製管理情報を記憶し、アクセス先決定手段が、複製管理情報に基づいて、分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定し、複製数計画手段が、記憶装置に対するアクセスが不能になったときの複製データの他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定し、分割データ管理手段が、記憶装置が追加されたときに、記憶装置群における複製データの再配置を行うことを特徴とする。

【００３６】

また、データ分散方法は、アクセス先決定手段が、少なくともコンテンツの一部を読み出すホスト端末に対し、ホスト端末が読み出す少なくともコンテンツの一部を構成する分割データへのアクセス先として該分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定し、通知してもよい。

【００３７】

また、データ分散方法は、アクセス先決定手段が、少なくともコンテンツの一部を読み出すホスト端末で、ホスト端末が読み出す少なくともコンテンツの一部を構成する分割データへのアクセス先として分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定してもよい。

【００３８】

また、データ分散方法は、アクセス先決定手段が、分割データに対応する複数の複製データが格納されている複数の記憶装置から、乱数に基づいてランダムに少なくとも１つの記憶装置を決定してもよい。

【００３９】

また、データ分散方法は、アクセス先決定手段が、分割データに対応する複数の複製データが格納されている複数の記憶装置のうち負荷の低い記憶装置をアクセス先として決定してもよい。

【００４０】

また、データ分散方法は、複製管理情報記憶手段が、分割データ管理手段によって決定された複製データの格納先を示す複製管理情報を記憶してもよい。

【００４１】

また、データ分散方法は、分割データ管理手段が、分割データに対応する複数の複製データが各々記憶装置に均等化されるように、複製データの格納先を決定してもよい。

【００４２】

また、データ分散方法は、分割データ管理手段が、コンテンツにおける連続する所定数の分割データに対応する複数の複製データが分散配置されるように、複製データの配置先を決定してもよい。

【００４３】

また、データ分散方法は、分割データ管理手段が、２つの記憶装置間で共有する同じ分割データに対応する複製データの数を共有数とした場合に、各々の記憶装置の組み合わせにおける共有数が均等化されるように、複製データの配置先を決定してもよい。

【００４４】

また、データ分散方法は、分割データ管理手段が、複製データの配置先を決定する際に、その複製データと同じ分割データに対応する複製データの配置先が既に決定している場合に、配置先として決定済みの記憶装置を除いた記憶装置の中から、決定済み記憶装置との組み合わせにおける共有数が小さい記憶装置を該複製データの配置先として決定してもよい。

【００４５】

また、データ分散方法は、分割データ管理手段が、記憶装置が使用不能となった場合に、複製管理情報記憶手段から、使用不能となった記憶装置を格納先として示している情報を削除してもよい。

【００４６】

また、データ分散方法は、分割データ管理手段が、使用不能となった記憶装置が配置先として割り当てられている複製データについて、新たな配置先を決定するとともに、その複製データと同じ分割データに対応する複製データを格納している記憶装置の中から１つの記憶装置をコピー元記憶装置に決定し、複製処理手段が、分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理を行ってもよい。

【００４７】

また、データ分散方法は、分割データ管理手段が、所定のタイミングで、複製管理情報に基づき、複製データについて新たな配置先を決定し、複製処理手段が、分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理および前の配置先からの削除処理を行ってもよい。

【００４８】

また、データ分散方法は、分割データ管理手段が、複製数計画手段によって決定された分割データの複製数に応じて、分割データに対応する複製データの配置先を決定してもよい。

【００４９】

また、データ分散方法は、複製数計画手段が、コンテンツにおける分割データの位置に基づいて、各分割データの複製数を決定してもよい。

【００５０】

また、データ分散方法は、複製数計画手段が、コンテンツの先頭に位置するデータを含む分割データ、およびコンテンツにチャプターが付与されている場合に、各チャプターの先頭に位置するデータを含む分割データの複製数を、そのコンテンツにおける他の分割データの複製数よりも多くしてもよい。

【００５２】
また、データ分散方法は、複製数計画手段が、コンテンツのアクセス要求予測量またはアクセス要求量と、各記憶装置の処理能力と、システムで規定した安全係数とに基づいて、分割データの複製数を決定してもよい。
【００５３】
また、データ分散方法は、分割データ管理手段が、複製数計画手段によって決定された分割データの複製数と、記憶装置に格納されているその分割データに対応する複製データの数とが一致するように、分割データに対応する複製データの新たな配置先または削除対象とする配置先を決定し、複製処理手段が、分割データ管理手段からの指示に従い、記憶装置間の複製データのコピー処理および記憶装置からの複製データの削除処理を行ってもよい。
【００５４】
また、本発明によるデータ分散用プログラムは、コンテンツが複数の分割データに分割され、少なくとも１つの分割データに対応する複数の複製データが記憶装置群における複数の記憶装置に格納されるデータ分散格納システムにおいて、少なくともコンテンツの一部を読み出すホスト端末に対し、そのコンテンツを構成する分割データへのアクセス先を通知する分割データ管理装置に適用されるデータ分散用プログラムであって、コンピュータに、コンテンツを複数の分割データに分割し、分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する処理、ホスト端末のコンテンツ読み出し範囲に含まれる分割データに対応する複製データの配置先を示す複製管理情報に基づいて、分割データへのアクセス先としてその分割データに対応する複製データが格納されている記憶装置のうち少なくとも１つの記憶装置を決定する処理、記憶装置に対するアクセスが不能になったときの複製データの他の記憶装置からのアクセスを保障できる数に応じて、分割データの複製数を決定する処理、および、記憶装置が追加されたときに、記憶装置群において複製データを再配置する処理を実行させることを特徴とする。
【００５５】
また、コンピュータに、コンテンツを複数の分割データに分割し、少なくとも１つの分割データに対応する複数の複製データの格納先として、複数の記憶装置を決定する処理を実行させてもよい。
【発明の効果】
【００５７】
本発明によれば、データ分散における特性・性能の向上を図ることができる。その理由は、少なくとも１つの分割データに対応する複数の複製データが複数の記憶装置に格納された各分割データの配置先を複製管理情報記憶手段が複製管理情報として記憶し、アクセス先決定手段が、その複製管理情報に基づいて、通常用と障害用とを区別することなく複数の記憶装置のうち少なくとも１つの記憶装置をアクセス先として決定するからである。
【発明を実施するための最良の形態】
【００５８】
まず、本発明の概要について説明する。本発明は、コンテンツを複数のデータに分割して記憶するデータ分散格納システムにおいて、個々の分割データをそれぞれ２以上に複製した上で、通常用と障害用等を区別することなく、使用可能な全記憶装置を配置先候補にして、同一内容の分割データがそれぞれ異なる記憶装置に格納されるように配置決定を行うことを特徴とする。さらに、同一内容の分割データが異なる記憶装置に格納されるだけでなく、各記憶装置間の相関（ある記憶装置と他のいずれの記憶装置との間における、２記憶装置間で共有する分割データの数）が低くなるように配置決定を行うことを特徴とする。さらに、ストリーミング再生用のコンテンツなど、データの連続性が認められるようなコンテンツを対象とする場合には、１度にアクセスされる可能性の高いデータサイズ（例えば、アクセスに用いるバッファ量相当）分を目安に、連続する分割データが同一の記憶装置に格納されないように配置決定を行うことを特徴とする。
【００５９】

また、配置決定に基づき各記憶装置に分散配置させた上で、同一内容の分割データに対するアクセス先の決定処理や、障害時の回復処理、および複製数の変更処理を行うことを特徴とする。

【００６０】

ここで、本発明を説明するために使用する用語を以下に定義する。本発明において、コンテンツとは、データ分散格納における分散格納の対象とするデータのまとまりをいう。例えば、コンテンツは、１つのファイルとしてまとめられたデータである。また、分割データとは、コンテンツを複数に分割したもの（データ）をいう。なお、各分割データのサイズは固定でなくてもよい。コンテンツ毎に異なるサイズであってもよいし、分割データ毎に異なるサイズであってもよい。例えば、音声や映像のようなエンコーディングされているコンテンツを対象とする場合には、そのエンコーディングの単位やその単位を複数まとめた単位を分割データのサイズとしてもよい。

【００６１】

以下、本発明の実施の形態について図面を参照して説明する。図１は、本発明によるデータ分散格納システムの構成例を示すブロック図である。図１に示すデータ分散格納システムは、ホスト１と、複数の記憶装置２（記憶装置２－１～ｎ。ｎは自然数）と、分割データ管理部３と、障害通知部４と、複製処理部５と、複製数計画部６と、分割データ管理データベース（分割データ管理ＤＢ）７とを備える。なお、図１では１つのホスト１しか示していないが、ホスト１は、本システムを利用するユーザの数に応じて複数存在する。

【００６２】

ホスト１は、複数の記憶装置２に分散格納されたコンテンツにアクセスするためのユーザ端末である。記憶装置２は、コンテンツを分散格納するための記憶装置である。各記憶装置２は、コンテンツを分割してできる各分割データおよびその複製を分散して記憶する。

【００６３】

分割データ管理部３は、コンテンツを構成する各分割データに対応する複数の複製データの配置先を管理する。なお、分割データ管理部３は、コンテンツ内におけるある１つの分割データに対し、その複製数に応じて複数の複製データの配置先を管理する。複製データとは、コンテンツを分割した分割データと同一内容のデータであって、実際に記憶装置に格納されるデータのことをいう。本発明において、”分割データ”をコンテンツ内における論理的な分割データとして表現するのに対し、”複製データ”を実際に記憶装置に格納する物理的な分割データとして表現するために用いている。なお、分割データ管理部３は、コンテンツを構成する各分割データに対応する複製データを複数の記憶装置に分散配置するために配置先を決定してその配置先を管理するだけでなく、既に分散配置されている状態からその配置先を管理してもよい。配置先の管理として、具体的には、分割データへのアクセス先としての記憶装置を決定したり、複製数の変更や記憶装置の追加、使用不能に伴う複製数の維持処理（複製データの追加、削除、再配置等）を行う。

【００６４】

なお、分割データへのアクセス先としての記憶装置の決定機能（以下、単にアクセス先決定機能という。）は、ホスト１が有していてもよい。その場合、分割データ管理部３は、ホスト１からの要求に応じて、所望の分割データに対応する複製データの配置先を示す情報をホスト１に通知してもよい。

【００６５】

障害通知部４は、記憶装置２の障害（具体的には、使用可否）を分割データ管理部３に通知する。複製処理部５は、記憶装置２への分割データの追加処理や削除処理やコピー処理など、記憶装置２に対するアクセス処理を行う。複製数計画部６は、各分割データの複製数を決定する。分割データ管理ＤＢ７は、分割データ管理部３が各分割データに対応する複製データの配置先を管理するための情報を記憶するための記憶装置である。分割データ管理ＤＢ７は、コンテンツと分割データの関連性に関する情報や、各複製データの配置先に関する情報や、配置先対象となる記憶装置に関する情報を記憶する。

【００６６】
以下、本実施の形態では、分割データのことを”チャンク”と言う。なお、”チャンク”という表現には、一般に意味的にまとまりのある分割データを指す場合があるが、本実施の形態においては、意味的にまとまりのある分割データに限定されず、意味的にまとまりのない分割データを含めてチャンクという。
【００６７】
図２は、本実施の形態によるデータ分散格納システムの構成例を示すブロック図である。図２に示すように、本実施の形態によるデータ分散格納システムは、ホスト１と、複数の記憶装置２（記憶装置２－１～ｎ。ｎは自然数）と、チャンク管理部３と、障害通知部４と、複製処理部５と、複製数計画部６とを備える。なお、図２では１つのホスト１しか示していないが、ホスト１は、本システムを利用するユーザの数に応じて複数存在する。
【００６８】
ホスト１は、具体的には、複数の記憶装置２に分散格納されたコンテンツにアクセスするデータ処理装置である。本実施の形態では、ホスト１は、チャンク管理部３から取得した、そのコンテンツを構成するチャンクを記憶している記憶装置２を示すチャンク配置情報に基づいて、必要なチャンクにアクセスする。
【００６９】
記憶装置２－１～ｎは、各チャンクを格納する記憶装置である。記憶装置２－１～ｎは、ブロックベースの記憶装置に限らず、例えば、ＮＡＳ等のファイルベースやＯＳＤ（Object based storage device ）等のオブジェクトベースの記憶装置システムであってもよい。そのような場合には、チャンクは、ファイルやオブジェクトとして格納されることとなる。なお、記憶装置２－１～ｎは、記憶装置を制御する制御装置を含む。例えば、記憶装置２－１～ｎは、磁気記憶装置、光磁気記憶装置、不揮発性の半導体記憶装置、およびそれらのアレイ装置、並びにそれらを制御する制御装置を備えたサーバ装置によって実現される。また、記憶装置２－１～ｎには、それぞれを識別するための記憶装置ＩＤが割り振られているものとする。
【００７０】
なお、本実施の形態において、記憶装置に記憶させるチャンクが複製かそうでないかを区別する必要はなく、単に、あるコンテンツを構成するチャンクと同一内容のチャンクが複数格納されている、として認識すれば足る。以下、”各チャンク”という表現には、コンテンツ内における各々のチャンクを示す場合と、その複製も含め記憶装置内における各々のチャンクを示す場合とを含んでいるものとする。なお、コンテンツ内において識別されるチャンクと、実際に複数の記憶装置に記憶させるチャンクとを区別するために、記憶装置に記憶させるチャンクを”複製チャンク”と表現する場合がある。
【００７１】
チャンク管理部３は、コンテンツを構成する各チャンクに対応する複数の複製チャンクの配置先を管理する。本実施の形態では、チャンク管理部３は、コンテンツを複数のチャンクに分割して、各チャンクに対しそれぞれ複数の配置先を決定して、その配置先を管理する。なお、チャンク管理部３は、コンテンツを構成する各チャンクに対しそれぞれ複数の配置先を決定してその配置先を管理するのではなく、既に分散配置されている状態からその配置先を管理してもよい。チャンク管理部３は、各チャンクの配置先を管理するためのチャンク管理データベース７（以下、チャンク管理ＤＢ７という。）を有する。なお、チャンク管理ＤＢ７は、独立したデータベースシステムとして存在していてもよい。その場合、少なくともチャンク管理部３と複製処理部５とがアクセス可能に接続されていればよい。
【００７２】
チャンク管理部３は、配置先の管理として、例えば、チャンクの複製数の変更が行われたときや障害時、他の処理部からの通知等、所定のタイミングでチャンクの削除や追加やコピー（再配置）を行う。なお、本実施の形態では、チャンク管理部３は各チャンクの配置先や参照元を決定するに留まり、実際の記憶装置への反映は複製処理部５が行う。また

、例えば、ホスト１からのコンテンツのアクセス要求を受けて、そのコンテンツを構成するチャンクのアクセス先を決定する。なお、このアクセス先決定機能は、ホスト１が有していてもよい。その場合、チャンク管理部３は、コンテンツを構成する各チャンクに対応する複製チャンクが格納されている記憶装置の情報をホスト１に通知してもよい。

【００７３】

障害通知部４は、障害が発生した等によって記憶装置２－１～ｎの使用できなくなった旨をチャンク管理部３に通知する。なお、障害通知部４は、自ら記憶装置の障害を検出する障害検出手段であってもよいし、あるいはシステム管理者の操作に応じて、障害対応動作指示を示す情報を入力する入力手段（例えば、キー入力手段）であってもよい。

【００７４】

複製処理部５は、チャンク管理部３からの指示に応じて、記憶装置２－１～ｎへのチャンクの追加処理、記憶装置２－１～ｎからのチャンクの削除処理、または記憶装置２－１～２－ｎ間のチャンクのコピー処理を行う。

【００７５】

複製数計画部６は、システムの状況に応じて、各チャンクの複製数を決定する。複製数画部６は、例えば、各コンテンツの要求量、要求予測量、および可用化数（障害等でも同一内容のチャンクが失われないことを目安に定められた数など）に基づいて、各チャンクの複製数を決定し、複製数が変わった場合にチャンク管理部３にチャンクの管理情報の変更を行わせる手段である。

【００７６】

チャンク管理ＤＢ７は、コンテンツとチャンクの関連性に関する情報、各チャンクの配置先に関する情報、および配置先とする記憶装置に関する情報を記憶する。具体的には、本システムが扱う各コンテンツの分割管理情報、チャンク毎の複製管理情報、および記憶装置情報を記憶する。

【００７７】

なお、チャンク管理部３，複製処理部５，複製数計画部６は、具体的には、ＣＰＵ等のプログラムに従って動作する情報処理装置によって実現される。チャンク管理部３，複製処理部５，複製数計画部６は、それぞれ異なる情報処理装置によって実現されてもよいし、同一の情報処理装置によって実現されてもよい。なお、通常アクセスに与える影響をなるべく無くすという点で、少なくともチャンク管理部３と他の処理部（複製処理部５や，複製数計画部６）とは、異なる情報処理装置によって実現されることが好ましい。

【００７８】

ここで、分割管理情報とは、各コンテンツについて、そのコンテンツをどのようなチャンクに分割したかを示す情報であって、例えば、あるコンテンツを構成する各チャンク（ここでは複製を含まず。）についてのチャンク参照情報を含む。チャンク参照情報は、具体的には、コンテンツ内においてそのチャンクを識別するためのチャンクコンテンツＩＤと、そのチャンクがコンテンツのどこのデータに該当するかを示す情報（例えば、先頭からのオフセットやチャンクサイズ等）とを含む情報である。なお、コンテンツ内の順序に従ってチャンク参照情報を登録する場合など、チャンク参照情報のエントリの順番によってチャンクコンテンツＩＤが特定できる場合には、チャンク参照情報はチャンクコンテンツＩＤを含んでいなくてもよい。また、チャンクサイズが全てのチャンクで固定である場合には、各チャンクのチャンク参照情報としてチャンクサイズを登録するのではなく、分割管理情報として１つの共通するチャンクサイズを含んでいればよい。

【００７９】

また、複製管理情報とは、各コンテンツ内における各チャンクについて、そのチャンクと同一内容のチャンク（すなわち、そのチャンクの複製チャンク）が、少なくともどの記憶装置に記憶されているかを示す情報であって、例えば、コンテンツ内における各チャンクに対して割り当てたチャンクコンテンツＩＤと、そのチャンクの複製数と、そのチャンクの各複製チャンクについての配置先情報とを対応づけた情報である。チャンク（複製チャンク）の配置先情報は、例えば、そのチャンクを記憶している記憶装置を識別するため

の記憶装置ＩＤとその記憶装置内でそのチャンクにアクセスするためのアクセス情報（ブロック番号やオフセット、または各記憶装置内においてチャンクを識別するためのチャンクアクセスＩＤなど）とを含む情報である。

【００８０】

また、記憶装置情報とは、本システムにおいて使用可能な記憶装置を示す情報であって、例えば、本システムが備える記憶装置を識別するための記憶装置ＩＤに対応づけて、その記憶装置の稼働状況（稼働中または停止中）と、記憶容量とを記憶する。

【００８１】

次に、チャンク管理部３でのコンテンツの管理方法について説明する。チャンク管理部３において、各コンテンツは、システム内においてそのコンテンツを識別するために定められるコンテンツＩＤにより一意に識別される。コンテンツ内における各チャンクは、そのコンテンツ内においてそのチャンクを識別するために定められるチャンクコンテンツＩＤにより一意に識別される。

【００８２】

また、記憶装置内における各チャンク（複製チャンク）は、その記憶装置におけるファイル名やパス名やｉｎｏｄｅ番号やｏｂｊｅｃｔＩＤ等のチャンクアクセスＩＤにより一意に識別される。また、本発明では、同一内容のチャンクはそれぞれ異なる記憶装置に配置されるので、各複製チャンクを記憶装置ＩＤとチャンクアクセスＩＤとによって一意に識別することも可能である。ここで、あるチャンクと同一内容の複製チャンクに対して全記憶装置で共通のチャンクアクセスＩＤが割り当て可能であるならば、チャンクコンテンツＩＤがチャンクアクセスＩＤを兼ねることができる。その場合、あるチャンクコンテンツＩＤで識別されるチャンクの複製チャンクを記憶装置ＩＤにより一意に識別することもできる。

【００８３】

まず、コンテンツＩＤとチャンクコンテンツＩＤとの間に相関をもたせない場合の管理方法を説明する。チャンク管理部３は、管理対象となるコンテンツについて、そのコンテンツの分割管理情報として、図３に示すような分割管理情報をコンテンツＩＤと対応づけてチャンク管理ＤＢ７に記憶させることによって、そのコンテンツをチャンクとして管理する。

【００８４】

図３は、分割管理情報の例を示す説明図である。図３に示すように、チャンク管理部３は、あるコンテンツの分割管理情報として、そのコンテンツのコンテンツＩＤと対応づけて、そのコンテンツをチャンクに分割した順に、そのチャンクのチャンク参照情報を記憶してもよい。例えば、コンテンツ内において先頭から１番目のチャンクのチャンク参照情報は、その分割管理情報において１番目のエントリとして登録される。なお、チャンク管理部３は、このような分割管理情報を、コンテンツ毎にチャンク管理ＤＢ７に記憶すればよい。

【００８５】

ここで、コンテンツ内における各チャンクのサイズが固定長である場合、各チャンクのチャンク参照情報は、そのチャンクのチャンクコンテンツＩＤを含んでいればよい。例えば、チャンクのサイズがｃであるとすると、コンテンツ上のオフセットアドレスがａであるデータが格納されるチャンクは、分割管理情報においてａ÷ｃ番目（小数点以下切捨て）のエントリとして登録されているチャンク参照情報に含まれるチャンクコンテンツＩＤで示されるチャンクであることがわかる。なお、実際のデータの格納先は、そのチャンクコンテンツＩＤに基づき、そのチャンクの複製管理情報を参照することによって得られる。なお、チャンクサイズは、１つの共通する値として分割管理情報に登録すればよい。

【００８６】

また、コンテンツ内において各チャンクのサイズがそれぞれ異なる場合、各チャンクのチャンク参照情報は、そのチャンクのチャンクコンテンツＩＤとチャンクサイズとを含んでいればよい。例えば、コンテンツ内においてｊ番目のチャンクのサイズがc_jであると

すると、コンテンツ上のオフセットアドレスがａであるデータが格納されるチャンクは、以下に示す式（１）となるｉ番目のエントリとして登録されているチャンク参照情報に含まれるチャンクコンテンツＩＤで示されるチャンクであることがわかる。

【００８７】

【数１】

$$\sum_{j=0}^{i} c_j - c_i \leq a < \sum_{j=0}^{i} c_j \qquad \cdots \text{式(1)}$$

【００８８】

次に、コンテンツＩＤとチャンクコンテンツＩＤとの間に相関をもたせる場合の管理方法を説明する。この方法では、コンテンツＩＤとコンテンツ内におけるチャンクの位置とによりそのチャンクのチャンクコンテンツＩＤを特定する。従って、各チャンクのチャンク参照情報には、チャンクコンテンツＩＤを含めなくてもよい。チャンクコンテンツＩＤの構成として、例えば、上位数ｂｉｔをコンテンツＩＤとし、残りの下位ｂｉｔをそのチャンクがコンテンツ内において何番目のチャンクかを示す番号とすればよい。

【００８９】

例えば、コンテンツ内における各チャンクのサイズが固定長である場合、各チャンクのチャンク参照情報は省略してもよい。すなわち、分割管理情報を、そのコンテンツにおけるチャンクサイズを格納する表として実現すればよい。なお、全コンテンツでチャンクのサイズが固定長である場合には、分割管理情報をコンテンツ毎に持たせる必要はなく、システムにおけるチャンクサイズを格納する１つのデータとして実現すればよい。また、コンテンツ内における各チャンクのサイズが可変長である場合には、各チャンクのチャンク参照情報として、チャンクサイズだけを分割管理情報に登録されればよい。すなわち、分割管理情報を、各チャンクのサイズを格納する表として実現すればよい。なお、コンテンツ上のオフセットアドレスで示されるデータが格納されるチャンクの把握方法は、上記で示したエントリの位置をチャンクの位置としてチャンクコンテンツＩＤを特定すればよい。

【００９０】

次に、チャンク管理部３でのチャンクの配置先の管理方法について説明する。チャンク管理部３は、管理対象となるコンテンツを構成する各チャンクについて、図４に示すような複製管理情報をチャンクコンテンツＩＤと対応づけてチャンク管理ＤＢ７に記憶させることによって各チャンクの配置先を管理する。

【００９１】

図４は、複製管理情報の一例を示す説明図である。図４に示すように、チャンク管理部３は、あるチャンクの複製管理情報として、そのチャンクのチャンクコンテンツＩＤと対応づけて、そのチャンクと同一内容のチャンク（すなわち、複製チャンク）を記憶している記憶装置の記憶装置ＩＤを記憶すればよい。なお、記憶装置ＩＤだけでなく、その記憶装置内でそのチャンクにアクセスするための情報を含んでいてもよい。例えば、記憶装置内におけるアドレス情報や記憶装置固有の識別子名（チャンクアクセスＩＤ）などである。また、複製管理情報は、そのチャンクの複製数を含んでいてもよい。なお、チャンク管理部３は、このような複製管理情報を、各コンテンツを構成する各チャンク毎にチャンク管理ＤＢ７に記憶すればよい。

【００９２】

チャンク管理部３は、コンテンツ内における複数のチャンクを連結して１つのチャンクとして管理することも可能である。チャンクの連結方法としては、記憶装置内においてチャンクを連結する方法と、コンテンツ内においてチャンクを連結する方法とがある。

【００９３】

記憶装置内においてチャンクを連結する方法とは、同じコンテンツを構成するチャンク

のうち同じ記憶装置に記憶される複数のチャンクに対し、１つのチャンクアクセスＩＤを割り当てる方法である。例えば、コンテンツをチャンクに分割した上で、そのチャンクを記憶装置に記憶させる際に、複製毎に異なるパターンで配置決定した後で同じ記憶装置に入るチャンクをまとめればよい。図５は、チャンクの連結結果の一例を示す説明図である。図５では、あるコンテンツを１２個に分割してできるチャンク（チャンク０～１１）を複製数＝２として各記憶装置に記憶させる場合において、記憶装置ＩＤ＝０～３に対して１通りの複製チャンクを割り当て、記憶装置ＩＤ＝１０～１３に対してもう１通りの複製チャンクを割り当てた場合の例を示している。図５に示す例では、複製毎にチャンクの配置先とする記憶装置と割り当てパターンを替えることで、同一内容のチャンクが異なる記憶装置に格納されるように、かつ各記憶装置間の相関が低くなるように配置先を決定している。このように配置決定された場合、チャンク管理部３は、例えば、記憶装置ＩＤ＝０を割当先とするチャンク０，４，８を連結した連結チャンクを記憶装置内における１つのチャンクとして管理する。ここで、連結チャンクとは、コンテンツを分割してできたチャンクの中から、２以上のチャンクを連結してできたチャンクをいう。同様に、例えば、記憶装置ＩＤ＝１０を割当先とするチャンク０，７，１０を連結した連結チャンクを記憶装置内における１つのチャンクとして管理する。

【００９４】

具体的には、複製管理情報に含まれるチャンクの配置先情報として、そのチャンクと同一内容のチャンクについて、そのチャンクを含む連結チャンクを記憶している記憶装置の記憶装置ＩＤと、その記憶装置内でその連結チャンクにアクセスするための情報（チャンクアクセスＩＤ）と、そのチャンクがその連結チャンクのどこに位置するかを示す情報（例えば、連結順序を示す情報やオフセットアドレス等）とを記憶すればよい。このように、記憶装置に記憶させる際にチャンクを連結する方法を用いれば、記憶装置内におけるチャンク数とコンテンツの分割数とを切り離して管理することができ、管理すべきチャンクアクセスＩＤの量を減らすことができる。

【００９５】

また、コンテンツ内においてチャンクを連結する方法とは、コンテンツ内における非連続なチャンクに対し、１つのチャンクコンテンツＩＤを割り当てる方法である。例えば、コンテンツをチャンクに分割した後で、巡回的なパターンで配置決定された場合に同じ記憶装置に入るチャンクを１つにまとめてもよい。図６は、チャンクの連結結果の一例を示す説明図である。図６では、あるコンテンツを１２個に分割してできるチャンク（チャンク０～１１）に対し、３を区切りに巡回させた場合の例を示している。このような場合、チャンク管理部３は、例えば、記憶装置ＩＤ＝１が配置先となるチャンク０，３，９を連結した連結チャンクをコンテンツ内における１つのチャンクとして管理する。同様に、例えば、記憶装置ＩＤ＝２が配置先となるチャンク１，４，１０を連結した連結チャンクをコンテンツ内における１つのチャンクとして管理する。

【００９６】

具体的には、分割管理情報として、コンテンツ内における連結チャンク数分のチャンク参照情報を記憶するようにし、各連結チャンクのチャンク参照情報に、各連結チャンクに対し割り当てたチャンクコンテンツＩＤと、その連結チャンクを構成する各チャンクの位置情報（コンテンツ上のオフセットアドレスとサイズ等）とを含めればよい。このように、コンテンツ内においてチャンクを連結する方法を用いれば、チャンクサイズの決定に際しデータの連続性を考慮しなくてもよいので、例えば、そのコンテンツに対し１度にアクセスされる可能性の高いデータサイズ分よりも大きいサイズを最終的なチャンクサイズとすることも可能である。また、システムで管理するチャンクコンテンツＩＤおよびチャンクアクセスＩＤの量を減らすこともできる。

【００９７】

次に、本実施の形態によるデータ分割管理システムの動作について説明する。まず、ホスト１からコンテンツを読み出す動作について説明する。図７は、本実施の形態によるデータ分割管理システムのコンテンツ読み出し動作の一例を示すフローチャートである。本

実施の形態では、読み出しを行うホスト１が、チャンク管理部３に対し、読み出したいコンテンツのコンテンツＩＤとコンテンツ内の読み出し開始位置とサイズとを指定してアドレス変換を要求し、その応答として、そのコンテンツの該当部分を構成するチャンクの格納先を示す情報を得て、その情報に基づき記憶装置にアクセスする場合を例にとって説明する。

【００９８】

ホスト１で読み込みを行う際には、ホスト１は、チャンク管理部３に、コンテンツＩＤとコンテンツ内の読み出し開始位置を示すオフセットアドレスとサイズと読み出しであることを指定して、アドレス変換を依頼する（ステップＳ１００）。ホスト１は、例えば、通信ネットワークを介してチャンク管理部３を備えたサーバ装置に接続されている場合には、そのサーバ装置に、コンテンツＩＤとコンテンツ内オフセットアドレスと読み出しサイズとを含む読み出し用アドレス変換要求メッセージを送信すればよい。そして、チャンク管理部３からの応答を待つ（ステップＳ１０１）。なお、チャンク管理部３におけるアドレス変換動作については後述する。

【００９９】

応答が帰ってきたら、ホスト１は、その応答の先頭のエントリを読み出し対象エントリとする（ステップＳ１０２）。ここで、読み出し用のアドレス変換要求に対するチャンク管理部からの応答のフォーマットは、読み出し対象のデータを構成するチャンク毎に１つのエントリとして格納されているものとする。また、各エントリは、例えば、そのチャンクを格納している記憶装置の記憶装置ＩＤと、記憶装置内でのアクセス情報（例えば、チャンクアクセスＩＤ）と、そのチャンク内での読み出し開始位置と、そのチャンク内での読み出しサイズとで構成される。また、ここでは、各エントリが、応答のデータの一部に、そのコンテンツにおけるオフセットアドレスの順に格納されている場合を例にとって説明する。なお、複数のチャンクを連結して１つのチャンクとする場合には、連結前のチャンク毎に１つのエントリとしてもよい。

【０１００】

図８は、読み出し用アドレス変換要求に対する応答フォーマットの一例を示す説明図である。図８に示す例では、読み出し対象のデータを構成するチャンク毎に、１つのエントリとして、そのチャンクと同一内容のチャンク（すなわち、複製チャンク）のいずれかの格納先である記憶装置の記憶装置ＩＤと、その記憶装置内でのチャンクアクセスＩＤと、そのチャンク内での読み出し開始位置（チャンク内オフセットアドレス）とそのチャンク内での読み出しサイズとを格納する例を示している。図８に示す例では、まず、エントリ１が、読み出し対象エントリとしてホスト１によって処理される。

【０１０１】

ホスト１は、読み出し対象エントリが示すチャンク（以下、単に対象チャンクという。）が未割り当ての場合にはステップＳ１０６に移行し、割り当てられていた場合にはステップＳ１０４に移行する（ステップＳ１０３）。ここで、チャンクが未割り当てとは、そのチャンクが記憶装置に書き込まれていないことを示している。なお、対象チャンクが未割り当てか否かは、例えば、記憶装置ＩＤやチャンクアクセスＩＤに無効な値が格納されているか否かで判断すればよい。

【０１０２】

対象チャンクが割り当てられていた場合、ホスト１は、読み出し対象エントリ中の記憶装置ＩＤで特定される記憶装置２に対して、読み出し対象エントリ中のチャンクアクセスＩＤ、チャンク内オフセットアドレス、読み出しサイズを指定し読み出しを要求する（ステップＳ１０４）。そして、記憶装置２からの応答を待つ（ステップＳ１０５）。

【０１０３】

ホスト１は、チャンク管理部３から通知された全エントリ（全チャンク）について、上記読み出し処理を行う。すなわち、記憶装置２から応答が帰ってきたら、チャンク管理部３からのアドレス変換応答に含まれる全エントリで示されるチャンクの読み出しが全て完了したか否かを判定し、全て完了していたら処理を終了し、完了していなければステップ

Ｓ１０７に移行する（ステップＳ１０６）。ステップＳ１０７では、まだ読み出しを行っていないチャンクを読み出すために、次のエントリを読み出し対象エントリとし、ステップＳ１０３に移行する。

【０１０４】

図８に示す例では、エントリ１の次に、エントリ２が読み出し対象エントリとされ、最終的にエントリｎまでが、読み出し対象エントリとしてホスト１によって処理されることとなる。本例では、各エントリがそのコンテンツにおけるオフセットアドレスの順に格納されているので、ホスト１は、読み出し対象エントリに従い、各チャンクの該当データを読み出して順次結合していけば、所望のコンテンツを得ることができる。なお、各エントリを、そのコンテンツにおけるオフセットアドレスの順に格納しない場合には、各エントリに、そのチャンクがコンテンツにおけるどの位置のデータであるかを示す情報を含めればよい。

【０１０５】

ここで、各チャンクをシーケンシャルに読み込む例を示したが、応答を待たずに並行に処理することも可能である。このようにすると処理時間を短縮することができる。なお、後述の配置先決定動作において説明するように、連続したチャンクをそれぞれ異なる記憶装置に配置させている場合には、ホスト１側で特別に意識しなくても、各記憶装置に並列にアクセスできるので、容易に処理時間を短縮することができる。

【０１０６】

次に、ホスト１からコンテンツを書き込む動作について説明する。図９は、本データ分割管理システムのコンテンツ書き込み動作の一例を示すフローチャートである。本データ分割管理システムでは、書き込みを行うホスト１が、チャンク管理部３に対し、書き込みたいコンテンツとコンテンツ内の書き込み開始位置とサイズとを指定してアドレス変換を要求し、その応答として、そのコンテンツの該当部分を構成するチャンク（およびその複製）の格納先を示す情報を得て、その情報に基づき記憶装置にアクセスする場合を例にとって説明する。

【０１０７】

ホスト１で書き込みを行う際には、ホスト１は、チャンク管理部３に、コンテンツＩＤとコンテンツ内の書き込み開始位置を示すオフセットアドレスと書き込みサイズと書き込みであることを指定して、アドレス変換を依頼する（ステップＳ２００）。ホスト１は、例えば、通信ネットワークを介してチャンク管理部３を備えたサーバ装置に接続されている場合には、そのサーバ装置に、コンテンツＩＤとコンテンツ内オフセットアドレスと書き込みサイズとを含む書き込み用アドレス変換要求メッセージを送信すればよい。そして、チャンク管理部３からの応答を待つ（ステップＳ２０１）。なお、チャンク管理部３が、新規なコンテンツに対するコンテンツＩＤの割り当てを行ってもよい。そのような場合、ホスト１は未割り当てを示すコンテンツＩＤを指定すればよい。また、チャンク管理部３におけるアドレス変換動作については後述する。なお、本発明において、同一内容のチャンクは、チャンク管理部３によって異なる記憶装置に配置されるべく配置先決定される。

【０１０８】

応答が帰ってきたら、ホスト１は、その応答の先頭のエントリを書き込み対象エントリとする（ステップＳ２０２）。ここで、書き込み用のアドレス変換要求に対するチャンク管理部３からの応答のフォーマットは、書き込み対象のデータを構成するチャンク毎に１つのエントリとして格納されているものとする。また、各エントリは、例えば、そのチャンク内での書き込みサイズと、そのチャンク内での書き込み開始位置と、そのチャンクの複製数に応じた各複製チャンクの格納先を示す情報とで格納される。複製チャンクの格納先を示す情報は、具体的には、その複製チャンクを格納する記憶装置の記憶装置ＩＤと、その記憶装置内でのチャンクアクセスＩＤでよい。また、本例では、各エントリが、応答のデータの一部に、そのコンテンツにおけるオフセットアドレスの順に格納されている場合を例にとって説明する。なお、複数のチャンクを連結して１つのチャンクとする場合に

は、連結前のチャンク毎に１つのエントリとしてもよい。

【０１０９】

図１０は、書き込み用アドレス変換要求に対する応答フォーマットの一例を示す説明図である。図１０に示す例では、書き込み対象のデータを構成するチャンク毎に１つのエントリとし、そのチャンク内での書き込みサイズと、そのチャンク内での書き込み開始位置と、そのチャンクの複製数と、そのチャンクの配置先情報として、そのチャンクの複製数に応じた、各複製チャンクの格納先となる記憶装置の記憶装置ＩＤとその記憶装置内でのアクセス情報（チャンクアクセスＩＤ）とを格納する例を示している。図１０に示す例では、まず、エントリ１が、書き込み対象エントリとしてホスト１によって処理される。なお、全記憶装置で共通のチャンクアクセスＩＤが定義されている場合には１つのエントリにつき１つのチャンクアクセスＩＤが格納されていればよい。また、同じ内容のチャンクの複製チャンクであっても、各複製チャンクでチャンク内での書き込み開始位置が異なる場合（例えば、記憶装置内においてチャンクを連結する場合）には、複製チャンク毎にその複製チャンク内での書き込み開始位置（チャンク内オフセットアドレス）を格納すればよい。

【０１１０】

次に、ホスト１は、書き込み対象エントリ中に、複製チャンクの書き込み先として先頭に示された記憶装置ＩＤで特定される記憶装置２を書き込み対象記憶装置として選ぶ（ステップＳ２０３）。以下、書き込み対象記憶装置に書き込むチャンクのことを対象複製チャンクという。

【０１１１】

書き込み対象記憶装置を特定すると、ホスト１は、書き込み対象エントリ中の書き込みサイズと、チャンク内オフセットアドレスと、対象複製チャンクについてのチャンクアクセスＩＤとを指定し、その対象複製チャンクとして書き込むべきデータとともに、書き込み対象記憶装置に対して書き込みを要求する（ステップＳ２０４）。なお、本例では、各エントリがそのコンテンツにおけるオフセットアドレスの順に格納されているので、ホスト１は、書き込むべきデータとして、エントリ毎に、コンテンツ内における書き込み開始位置からそのチャンク内での書き込みサイズ分のデータを順次指定していけばよい。そして、記憶装置２からの応答を待つ（ステップＳ２０５）。

【０１１２】

ホスト１は、記憶装置２から応答が帰ってきたら、書き込み対象エントリ中に各複製チャンクの格納先として示された全記憶装置に対し書き込みが完了したか否かを判定し、全て完了していればステップＳ２０８へ、完了していなければステップＳ２０７へ移行する（ステップＳ２０６）。ステップＳ２０７では、まだ書き込みを行っていない記憶装置に書き込むために、現書き込み対象エントリ中に、複製チャンクの格納先として次に示された記憶装置ＩＤで特定される記憶装置２を書き込み対象記憶装置とし、ステップＳ２０４に移行する。

【０１１３】

また、ステップＳ２０８では、チャンク管理部３からのアドレス変換応答に含まれる全エントリについて書き込み処理が完了していたか否かを判定し、全て完了していたら処理を終了し、完了していなければステップＳ２０９に移行する。ステップＳ２０９では、まだ書き込みを行っていないエントリについて書き込み処理を行うために、次のエントリを書き込み対象エントリとし、ステップＳ２０３に移行する。図１０に示す例では、エントリ１で示される全複製チャンクの書き込み処理が完了した後に、エントリ２が書き込み対象チャンクとされ、エントリ２で示される各複製チャンクの書き込み処理が行われることとなる。最終的に、エントリｎまでが書き込み対象エントリとされ、エントリｎで示される各複製チャンクの書き込み処理が行われることとなる。

【０１１４】

このようにして、同じ内容のチャンクを複数の記憶装置に書き込むことができるので、データが冗長化され、データの可用性および保全性が向上する。また、書き込み動作につ

いても、各チャンクを、応答を待たずに並列に処理することが可能である。このようにすると処理時間を短縮することができる。また、ホスト側で特別に意識しなくても、各記憶装置に並列にアクセスできるので、容易に処理時間を短縮することができる。なお、後述の配置先決定動作において説明するように、同一内容のチャンクが異なる記憶装置に格納されるよう配置先決定されるので、その配置先に基づいて各記憶装置にチャンクを書き込むことにより、読み出す際にアクセスする記憶装置を分散させることができ、同時に読み出す際のアクセス性（スループットや応答時間）も向上する。

【０１１５】

なお、上記例では、コンテンツ内におけるチャンク毎に１つのエントリとして、その中に各複製チャンクの格納先についての情報を格納する応答フォーマットの例を示したが、これに限らず、例えば、複製チャンク毎に１つのエントリとして、その中に、格納先についての情報と、その複製チャンクがコンテンツにおけるどの位置のデータであるかを示す情報とを格納させてもよい。そのような場合には、ホスト１は、各エントリで示される複製チャンクについての書き込み処理を順次行っていけばよい。

【０１１６】

記憶装置に対するチャンクのアクセス処理は、ブロックデバイスへのアクセス処理や、ＮＡＳにおけるファイルへのアクセス処理や、ＯＳＤにおけるオブジェクトへのアクセス処理など一般的なアクセス処理である。例えば、記憶装置がブロックデバイスによって実現される場合には、記憶装置内のチャンクに対応する、ブロックデバイス中の開始オフセットアドレスとサイズを指定してアクセスすればよい。記憶装置内のチャンクに対応する開始オフセットアドレスは、チャンクが固定長であれば、何番目のチャンクであるかによって算出すればよい。また、可変長の場合には、ブロックデバイス内のチャンクごとのオフセットアドレスやサイズのリストを定義しておき、これらに基づき算出すればよい。なお、チャンクの一部に対しアクセスする場合には、ブロックデバイス内のオフセットアドレスにチャンク内オフセットアドレスを加算すればよい。

【０１１７】

また、例えば、記憶装置がＮＡＳ等のファイルベースの記憶システムによって実現される場合には、記憶装置内のチャンクに対応するファイルを指定してアクセスすればよい。なお、チャンクの一部に対しアクセスする場合には、チャンクに対応するファイル内のオフセットアドレスにチャンク内オフセットアドレスを指定してアクセスすればよい。また、例えば、記憶装置がＯＳＤによって実現される場合には、記憶装置内のチャンクに対応するオブジェクトのリードまたはライト処理を実行すればよい。チャンクの一部に対しアクセスする場合には、オブジェクト内のＬＢＡにチャンク内オフセットアドレスを指定したリードまたはライト処理を実行すればよい。なおＮＡＳやＯＳＤにおいて書き込み時に対象チャンクが存在していなかった場合には、ファイルやオブジェクトの作成処理も同時に行う。

【０１１８】

次に、チャンク管理部３でのアドレス変換処理について説明する。既に説明したように、アドレス変換処理は、コンテンツＩＤとコンテンツ内の読み出し／書き込み開始位置と読み出し／書き込みサイズを指定して依頼される。

【０１１９】

図１１は、チャンク管理部３におけるアドレス変換処理の処理フローの一例を示すフローチャートである。チャンク管理部３は、まず、書き込みか読み出しのどちらが指示されたか調べ（ステップＳ３０１）、書き込みであれば書き込みアドレス変換処理（ステップＳ３０２）を行い、読み出しであれば読み出しアドレス変換処理（ステップＳ３０３）を行い、応答を返し（ステップＳ３０４）、終了する。

【０１２０】

図１２は、読み出しアドレス変換処理の処理フローの一例を示すフローチャートである。図１２に示すように、チャンク管理部３は、まず、アドレス変換の対象として指定されたコンテンツＩＤと対応づけられた分割管理情報を読み込む（ステップＳ４０１）。チャ

ンク管理部３は、例えば、指定されたコンテンツＩＤと対応づけられた分割管理情報をチャンク管理ＤＢ７から検索して読み込めばよい。ここで、チャンク管理部３は、指定されたコンテンツＩＤと対応づけられた分割管理情報がチャンク管理ＤＢ７中に存在していればステップＳ４０４に移行し、存在していなければステップＳ４０３に移行する（ステップＳ４０２）。

【０１２１】

ステップＳ４０３では、チャンク管理部３は、チャンクが未割り当てであることを示す値を応答のエントリに設定し、処理を終了する。

【０１２２】

また、ステップＳ４０４では、チャンク管理部３は、指定された読み出し開始位置と読み出しサイズ、およびステップＳ４０１で読み込んだ分割管理情報によって特定される各チャンクのコンテンツ内におけるオフセットアドレスとチャンクサイズとから、アドレス変換対象領域をチャンクに分割する。チャンク管理部３は、例えば、読み込んだ分割管理情報に含まれる各チャンクのチャンク参照情報群に基づいて、各チャンクのオフセットアドレスとサイズとを特定し、指定された読み出し開始位置と読み出しサイズ、および特定した各チャンクのコンテンツ内におけるオフセットアドレスとチャンクサイズから、読み出し対象として指定されたデータがそのコンテンツ内においてどのチャンクに該当するかを算出することによって、アドレス変換の対象とするデータに該当するチャンクの範囲をチャンクコンテンツＩＤにより特定する。例えば、読み出し対象データが、コンテンツ内のオフセットアドレスの順でみた場合に、何番目のチャンクから何番目のチャンクまでの情報によって構成されているかを計算することによって、変換対象とするチャンクの範囲を特定すればよい。

【０１２３】

変換対象とするチャンクの範囲の特定方法として、チャンクが固定長である場合には、指定された読み出し開始位置ａと読み出しサイズｓ、およびチャンクサイズｃとに基づいて、先頭のチャンクおよび末尾のチャンクが何番目（０基準）かは、先頭チャンクをｅ番目、末尾チャンクをｆ番目とすると、それぞれ次のように求まる。

【０１２４】

ｅ＝ａ÷ｃ（小数点以下切り捨て）　・・・式（２）

ｆ＝（ａ＋ｓ）÷ｃ（小数点以下切り捨て）　・・・式（３）

【０１２５】

また、ｅ番目の先頭チャンク中の読み出し開始位置ｇ、およびｆ番目の末尾チャンク中の読み出し終了位置ｈは、次のように求まる。

【０１２６】

ｇ＝ａ－ｃ（ｅ－１）　・・・式（４）

ｈ＝（ａ＋ｓ）－ｃ（ｆ－１）　・・・式（５）

【０１２７】

また、チャンクサイズが可変長の場合、ｊ番目のチャンクサイズがc_jであるとすると、先頭のチャンクが何番目かは以下の式（６）を満たすｅによって求まる。また、末尾のチャンクが何番目かは以下の式（７）を満たすｆによって求まる。

【０１２８】

【数２】

$$\sum_{j=0}^{e} c_j - c_e \le a < \sum_{j=0}^{e} c_j \qquad \cdots \text{式(6)}$$

$$\sum_{j=0}^{f} c_j - c_f \le a + s < \sum_{j=0}^{f} c_j \qquad \cdots \text{式(7)}$$

【０１２９】
また、ｅ番目の先頭チャンク中の読み出し開始位置ｇ、およびｆ番目の末尾チャンク中の読み出し終了位置ｈは、以下の式（８）および式（９）によって求まる。
【０１３０】
【数３】

$$g = a - \sum_{j=0}^{e} c_j - c_e \qquad \cdots\text{式(8)}$$

$$h = a + s - \sum_{j=0}^{f} c_j - c_f \qquad \cdots\text{式(9)}$$

【０１３１】
変換対象とするチャンクの範囲を特定すると、チャンク管理部３は、先頭のチャンクを対象チャンクとして設定するとともに、その対象チャンクについてのエントリを応答の先頭のエントリに設定する（ステップＳ４０５）。そして、対象チャンクの複製管理情報を読み込む（ステップＳ４０６）。チャンク管理部３は、例えば、対象チャンクのチャンクコンテンツＩＤと対応づけられた複製管理情報をチャンク管理ＤＢ７中から検索して読み込めばよい。ここで、チャンク管理部３は、分割管理情報において対象チャンクにチャンクコンテンツＩＤが割り当てられており、さらに複製管理情報において、対象チャンクの複製チャンク（対象チャンクと同一内容の記憶装置におけるチャンク）が記憶装置に割り当てられていればステップＳ４０９に移行し、そうでなければステップＳ４０８に移行する（ステップＳ４０７）。チャンク管理部３は、例えば、取得した複製管理情報において、対象チャンクのチャンクコンテンツＩＤに対し、少なくとも１つ以上の記憶装置ＩＤやチャンクアクセスＩＤが割り当てられているかを確認すればよい。
【０１３２】
ステップＳ４０８では、対象チャンクについての応答のエントリに、チャンクが未割り当てであることを示す情報を設定し、ステップＳ４１２に移行する。
【０１３３】
また、ステップＳ４０９では、対象チャンクの複製管理情報に含まれるチャンクの配置先情報を参照し、そして、ステップＳ４０９で配置先情報に基づいて、ホスト１にアクセスさせる記憶装置を決定する（ステップＳ４１０）。
【０１３４】
アクセス先の決定としては、例えば、乱数により決定する方法がある。例えば、所定の乱数を発生させ、チャンクの格納先の数（複製数）で除算した余りに応じてアクセス先を決定してもよい。また、例えば、複製チャンクを記憶している各記憶装置を順番に使用させるラウンドロビン方式によってアクセス先を決定してもよい。また、例えば、複製チャンクを記憶している各記憶装置のその時点の負荷から、最も負荷が低い記憶装置をアクセス先として決定してもよい。なお、各記憶装置の負荷は、例えば、チャンク管理部３や、各記憶装置が有する制御装置が、単位時間当たりの記憶装置に対するＩ／Ｏ数や転送データ量、ＣＰＵ利用率を検出することによって、判断すればよい。
【０１３５】
アクセス先が決定すると、チャンク管理部３は、そのアクセス先を示す情報を応答の該当エントリに設定する（ステップＳ４１１）。例えば、チャンク管理部３は、対象チャンクについての応答のエントリに、アクセス先として決定した記憶装置の記憶装置ＩＤと、必要に応じてその記憶装置内でのアクセス情報（チャンクアクセスＩＤ）とを設定する。
【０１３６】
次に、チャンク管理部３は、ステップＳ４１２において、対象チャンクについての応答のエントリに、そのチャンク内での読み出し開始位置（チャンク内オフセットアドレス）

と読み出しサイズとを設定する。なお、この動作は、対象チャンクが未割り当てであった場合にも行う。

【０１３７】

チャンク内オフセットアドレスおよび読み出しサイズは、例えば、既に説明した変換対象とするチャンクの範囲の把握動作において求めたチャンクサイズｃ（またはｃ$_j$）や先頭チャンクの読み出し開始位置ｇと末尾チャンクの読み出し終了位置ｈとによって求められる。

【０１３８】

具体的には、対象チャンクが先頭チャンクである場合、チャンク内オフセットアドレスはｇ，読み出しサイズはｃ－ｇとなる。ここで、ｃは対象チャンクのチャンクサイズを示す。また、対象チャンクが末尾チャンクである場合、チャンク内オフセットアドレスは０，読み出しサイズはｈとなる。また、対象チャンクが先頭チャンクおよび末尾チャンク以外のチャンクである場合、チャンク内オフセットアドレスは０，読み出しサイズはｃとなる。なお、先頭チャンク＝末尾チャンク（ｅ＝ｆ）の場合、チャンク内オフセットアドレスはｇ，読み出しサイズはｈ－ｇとなる。

【０１３９】

なお、複数のチャンクを連結して１つのチャンクとする場合であって、連結前のチャンク毎に１つのエントリとする場合には、各複製チャンクのチャンク内オフセットアドレスには、連結チャンクにおける対象チャンクの位置を加味する必要がある。

【０１４０】

対象チャンクについてのアドレス変換（エントリの設定）が終了すると、チャンク管理部３は、変換対象とするチャンクの範囲内でアドレス変換が終わっていないチャンクがなければ処理を終了し、あればステップＳ４１４に移行する（ステップＳ４１３）。ステップＳ４１４では、対象チャンクを次のチャンクに設定し、ステップＳ４０６に移行する。

【０１４１】

なお、ステップＳ４０１において、チャンクサイズの取得は、チャンクサイズがシステムに対し予め定められている場合には省略してもよい。また、コンテンツ毎に可変である場合には、例えば、そのコンテンツの分割管理情報に含まれているチャンクサイズを参照すればよい。また、各チャンク毎に可変である場合には、例えば、各チャンクのチャンク参照情報に含まれているチャンクサイズを参照すればよい。

【０１４２】

また、本例では、ステップＳ４０３において、全対象データが未割り当てであった場合には、全体を１つのチャンクとして未割り当てを示すエントリの応答として返す例を示している。

【０１４３】

以上のように、同一内容のチャンクを複数の記憶装置に記憶した上で、アクセス先を分散させることによって、複数のホストから同一のチャンクが読み出されるような場合であっても、読み出し負荷を分散させることができる。

【０１４４】

なお、本例では、チャンク管理部３がアクセス先を決定する例を示したが、チャンク管理部３では、１つのエントリに各複製チャンクの格納先を示す情報群を格納してホスト１に返信するようにし、ホスト１側でアクセス先とする記憶装置を決定してもよい。なお、ホスト側におけるアクセス先の決定方法としては、例えば、乱数を用いたり、ホスト１に予め定められた値を用いて該当チャンクを記憶した複数の記憶装置から１つの記憶装置を選べばよい。

【０１４５】

例えば、チャンク管理部３は、ステップＳ４１０においてアクセス先を決定せずに、対象チャンクの複製管理情報で示されるその対象チャンクの全配置先を、その対象チャンクについての応答のエントリに設定する。図１３および図１４は、ホスト１側でアクセス先とする記憶装置を決定する場合の応答フォーマットの例を示す説明図である。チャンク管

理部３は、例えば、図１３に示すように、１つの対象チャンクにつき１つのエントリとして、そのチャンク内での読み出し開始位置（チャンク内オフセットアドレス）と読み出しサイズと複製数と、全配置先情報として複製数分の記憶装置ＩＤおよびチャンクアクセスＩＤとを格納すればよい。

【０１４６】

なお、図１４は、記憶装置内のチャンクアクセスＩＤが記憶装置間で共通である場合の応答フォーマットの例である。図１４に示す例では、１つの対象チャンクにつき１つのエントリとして、そのチャンク内での読み出し開始位置（チャンク内オフセットアドレス）と読み出しサイズと複製数とそのチャンクの全複製チャンクに対し割り当てられた共通のチャンクアクセスＩＤと、複製数分の記憶装置ＩＤとを格納する例を示している。また、同じ内容のチャンクであっても、記憶装置内においてそのチャンク内での読み出し開始位置が異なる場合（例えば、記憶装置内においてチャンクを連結する場合）には、複製チャンク毎にその記憶装置内におけるチャンク内での読み出し開始位置（チャンク内オフセットアドレス）を格納すればよい。

【０１４７】

そして、ホスト１は、ステップＳ１０４において記憶装置２に読み出しを要求する前に、読み出し対象エントリ中に全配置先情報として示されている複数の記憶装置ＩＤから１つの記憶装置ＩＤを選び出すことによって、アクセス先とする記憶装置を決定する。この際、ホスト１は、乱数あるいは、ホストに規定された数字を用いてもよい。

【０１４８】

また、図１５は、書き込みアドレス変換処理の処理フローの一例を示すフローチャートである。図１５に示すように、チャンク管理部３は、まず、アドレス変換の対象として指定されたコンテンツＩＤと対応づけられた分割管理情報を読み込む（ステップＳ５０１）。チャンク管理部３は、例えば、指定されたコンテンツＩＤと対応づけられた分割管理情報をチャンク管理ＤＢ７から検索して読み込めばよい。ここで、チャンク管理部３は、指定されたコンテンツＩＤと対応づけられた分割管理情報がチャンク管理ＤＢ７中に存在していればステップＳ５１１に移行し、存在していなければステップＳ５０３に移行する（ステップＳ５０２）。なお、未割り当てを示すコンテンツＩＤが指定された場合、チャンク管理部３は、新たにコンテンツＩＤを割り当てて、ステップＳ５０３に移行すればよい。

【０１４９】

ステップＳ５０３では、新規コンテンツ用に、チャンク管理ＤＢ７中に、そのコンテンツＩＤと対応づけた分割管理情報を記憶するための領域を確保し、その分割管理情報を初期化する。チャンク管理部３は、例えば、チャンクサイズを決定し、決定したチャンクサイズによって定まるチャンクの分割数に応じて、各チャンクのチャンク参照情報を生成し、分割管理情報として登録することによって初期化すればよい。チャンクサイズは、システムとして規定値を持たせてもよいし、コンテンツ毎に設定できるようにしてもよいし、コンテンツ内のチャンク毎に設定できるようにしてもよい。また、１回の書き込み毎にチャンクを分割してその都度可変とする方法もある。なお、各チャンクのチャンク参照情報には、チャンクサイズに応じて算出されるチャンクのオフセットアドレス等を登録してもよい。なお、各チャンクのチャンクコンテンツＩＤは、そのチャンクの複製管理情報が作成されるまでに登録されていればよく、この時点では未割り当てを示す情報を登録してもよいし、ここで割り当ててその値を登録してもよい。

【０１５０】

次いで、チャンク管理部３は、指定された書き込み開始位置と書き込みサイズ、およびステップＳ５０３の初期化の際に決定されたチャンクサイズならびにチャンク参照情報から、アドレス変換対象領域をチャンクに分割する（ステップＳ５０４）。具体的には、チャンク管理部３は、指定された書き込み開始位置と書き込みサイズ、および各チャンクのコンテンツ内におけるオフセットアドレスとチャンクサイズとから、書き込み対象として指定されたデータがそのコンテンツ内においてどのチャンクに該当するかを算出すること

によって、アドレス変換の対象とするチャンクの範囲をチャンクコンテンツＩＤにより特定する。

【０１５１】

アドレス変換の対象とするチャンクの範囲を特定すると、チャンク管理部３は、先頭のチャンクを対象チャンクとして設定するとともに、その対象チャンクについてのエントリを応答の先頭のエントリに設定する（ステップＳ５０５）。そして、対象チャンクの複製チャンク（対象チャンクと同一内容の記憶装置におけるチャンク）を記憶装置に割り当てるチャンク作成処理を行う（ステップＳ５０６）。なお、チャンク作成処理の詳細については後述する。

【０１５２】

チャンク作成処理が完了すると、対象チャンクの配置先情報を、応答の該当エントリに設定する（ステップＳ５０７）。チャンク管理部３は、例えば、チャンク作成処理によってチャンク管理ＤＢ７中に作成された、対象チャンクのチャンクコンテンツＩＤと対応づけられた複製管理情報を読み込み、その複製管理情報で配置先として示される複数の記憶装置ＩＤとアクセス情報（チャンクアクセスＩＤ）とを応答のエントリに設定する。また、チャンク管理部３は、対象チャンクについての応答のエントリに、そのチャンク内での書き込み開始位置（チャンク内オフセットアドレス）と書き込みサイズとを設定する（ステップＳ５０８）。なお、チャンク内オフセットアドレスおよび書き込みサイズの算出方法は、読み出しアドレス変換処理におけるチャンク内オフセットアドレスおよび読み出しサイズの算出方法と同様である。

【０１５３】

対象チャンクについてのアドレス変換（エントリの設定）が終了すると、チャンク管理部３は、変換対象とするチャンクの範囲内でアドレス変換が終わっていないチャンクがなければ処理を終了し、あればステップＳ５１０に移行する（ステップＳ５０９）。ステップＳ５１０では、対象チャンクを次のチャンクに設定し、ステップＳ５０６に移行する。

【０１５４】

また、チャンク管理部３は、指定されたコンテンツＩＤと対応づけられた分割管理情報が存在していた場合には（ステップＳ５０２のＹｅｓ）、指定された書き込み開始位置と書き込みサイズ、および読み込んだ分割管理情報によって特定される各チャンクのコンテンツ内におけるオフセットアドレスとチャンクサイズとから、アドレス変換対象領域をチャンクに分割する（ステップＳ５１１）。具体的には、チャンク管理部３は、ステップＳ５０４と同様に、書き込み対象として指定されたデータがそのコンテンツ内においてどのチャンクに該当するかを算出することによって、アドレス変換の対象とするチャンクの範囲をチャンクコンテンツＩＤにより特定する。

【０１５５】

アドレス変換の対象とするチャンクの範囲を特定すると、チャンク管理部３は、先頭のチャンクを対象チャンクとして設定するとともに、その対象チャンクについてのエントリを応答の先頭のエントリに設定する（ステップＳ５１２）。

【０１５６】

対象チャンクが設定されると、チャンク管理部３は、まず、対象チャンクの複製管理情報を読み込む（ステップＳ５１３）。ここで、チャンク管理部３は、分割管理情報において、対象チャンクにチャンクコンテンツＩＤが割り当てられており、さらに複製管理情報において、そのチャンクに対し配置先である記憶装置が割り当てられていればステップＳ５１６に移行し、そうでなければステップＳ５１５に移行する（ステップＳ５１４）。

【０１５７】

ステップＳ５１５では、対象チャンクに対し、配置先として複製数分の記憶装置を割り当てるチャンク作成処理を行う。チャンク作成処理は、ステップＳ５０６と同様である。なお、チャンク管理部３は、チャンク作成処理が完了すると、チャンク作成処理によって作成された対象チャンクの複製管理情報を読み込む。

【０１５８】

また、ステップＳ５１６では、チャンク管理ＤＢ７に登録されている対象チャンクの複製管理情報を読み込む。チャンク管理部３は、例えば、対象チャンクのチャンクコンテンツＩＤと対応づけられた複製管理情報をチャンク管理ＤＢ７中から検索して読み込めばよい。

【０１５９】

そして、対象チャンクの複製管理情報に含まれるチャンクの配置先情報を参照して（ステップＳ５１７）、配置先として示されている複数の記憶装置ＩＤとアクセス情報（チャンクアクセスＩＤ）とを、応答のエントリに設定する（ステップＳ５１８）。また、チャンク管理部３は、対象チャンクについての応答のエントリに、そのチャンク内での書き込み開始位置（チャンク内オフセットアドレス）と書き込みサイズとを設定する（ステップＳ５１９）。ここで、複製管理情報において既に対象チャンクの配置先として記憶装置が割り当てられていた場合には、再書き込みを行わせないようにするために、応答のエントリに配置先情報を設定しないようにすることも可能である。なお、この場合においても、ホスト１に次のチャンクのオフセットアドレスを知らせるために書き込みサイズは設定しておく。

【０１６０】

対象チャンクについてのアドレス変換（エントリの設定）が終了すると、チャンク管理部３は、変換対象とするチャンクの範囲内でアドレス変換が終わっていないチャンクがなければ処理を終了し、あればステップＳ５２０に移行する（ステップＳ５１９）。ステップＳ５３０では、対象チャンクを次のチャンクに設定し、ステップＳ５１３に移行する。

【０１６１】

ここで、ステップＳ５０４，Ｓ５１１における変換対象とするチャンクの範囲の特定方法は、読み出しアドレス変換処理における方法と同様である。また、ステップＳ５０８，Ｓ５１８で設定するチャンク内オフセットアドレスおよび書き込みサイズの算出方法についても、読み出しアドレス変換処理における算出方法と同様である。

【０１６２】

次に、チャンク作成処理について説明する。図１６は、チャンク管理部３におけるチャンク作成処理の処理フローの一例を示すフローチャートである。図１６に示すように、まず、チャンク管理部３は、対象チャンクの配置先を管理するために、チャンク管理ＤＢ７中に、その対象チャンクについての複製管理情報を記憶するための領域を確保する（ステップＳ６０１）。チャンク管理部３は、対象チャンクにチャンクコンテンツＩＤが割り当てられていなければ、チャンクコンテンツＩＤを割り当て、そのチャンクコンテンツＩＤと対応づけた複製管理情報を記憶するための領域を確保し、その複製管理情報を初期化する。チャンク管理部３は、例えば、複製数を決定し、決定した複製数を複製管理情報に登録することによって初期化すればよい。チャンクの複製数はシステムの規定の値としてもよいし、複製数計画部６によって決定させてもよい。なお、チャンクの配置先情報には、未割り当てを示す情報を登録すればよい。

【０１６３】

次に、対象チャンクの配置先とする記憶装置を決定する（ステップＳ６０２）。なお、配置先の決定方法については後述する。そして、決定した配置先を示す情報を、その対象チャンクのチャンクコンテンツＩＤと対応づけた複製管理情報に記録する（ステップＳ６０３）。

【０１６４】

複製数分の配置先を決定していればステップＳ６０５に移行し、まだ決定されていなければステップＳ６０２に移行する（ステップＳ６０４）。最後に、対象チャンクのチャンクコンテンツＩＤをステップＳ６０１で割り当てていれば、その割り当てたチャンクコンテンツＩＤを、その対象チャンクを含むコンテンツの分割管理情報の該当チャンクのチャンク参照情報に記憶し、終了する（ステップＳ６０５）。本例では、実際に記憶装置へのチャンクの書き込みは、配置先を通知されたホスト１側で行われるが、例えば、後述の複製チャンク数追加処理で示すように、チャンク管理部３の指示に応じて複製処理部５が行

うようにしてもよい。

【０１６５】

次に、記憶装置を追加した際の処理について説明する。記憶装置を追加した旨が通知されると、チャンク管理部３は、少なくともチャンク管理ＤＢ７中の記憶装置情報にその記憶装置の情報を登録する。チャンク管理部３は、記憶装置情報に登録することによって、以降のチャンク作成処理で、追加された記憶装置をチャンクの割り当て先として認識させる。なお、チャンク管理部３は、記憶装置が追加されたことを契機にして、チャンクの再配置を行ってもよい。

【０１６６】

次に、システムに障害が発生した際の処理について説明する。例えば、障害検出部４は、障害が発生し使用不能になった記憶装置を検出すると、その記憶装置の記憶装置ＩＤをチャンク管理部３に通知する。障害検出部４は、例えば、記憶装置が正常動作をしていることを外部に知らせるために送出している信号を監視し、一定時間以上その送信が確認されないときに使用不能を検出してもよいし、例えば、システム管理者の操作に応じて、障害対応動作が必要な記憶装置ＩＤを入力することによって、使用不能な記憶装置を検出してもよい。使用不能の旨が通知されると、チャンク管理部３は、次に示すような障害対応処理を行う。

【０１６７】

図１７は、チャンク管理部３における障害対応処理の処理フローの一例を示すフローチャートである。図１７に示すように、チャンク管理部３は、まず、通知された記憶装置ＩＤで示される記憶装置をチャンクの割り当て対象から外す（ステップＳ７０１）。チャンク管理部３は、例えば、チャンク管理ＤＢ７中に記憶装置情報として登録されている情報のうち、その記憶装置ＩＤと対応づけられた稼働状況を停止中とすればよい。

【０１６８】

次いで、本システムが管理しているチャンクコンテンツＩＤ（チャンク管理ＤＢ７に記録されているチャンクコンテンツＩＤ）のうち、未検査のチャンクを選び、そのチャンクの複製管理情報を読み込む（ステップＳ７０２）。そして、その複製管理情報において、割り当て対象から外した記憶装置ＩＤが、チャンクの配置先として登録されているか否かを確認し、その記憶装置ＩＤが登録されていた場合ステップＳ７０４に、そうでなければステップＳ７０５に移行する（ステップＳ７０３）。

【０１６９】

ステップＳ７０４では、複製管理情報において、チャンクの配置先として登録されているその装置ＩＤを削除する。例えば、チャンク管理部３は、読み込んだ複製管理情報においてチャンクの配置先情報として示されている、割り当て対象から外した装置ＩＤを、未割り当てを示す情報に変更し、変更した複製管理情報をチャンク管理ＤＢ７に記録する。また、ステップＳ７０５では、チャンク管理ＤＢ７に記録されている全チャンクコンテンツＩＤについて検査が完了していれば処理を終了し、完了していなければステップＳ７０２に移行する。

【０１７０】

このように、チャンク管理部３により、障害等により使用不能となった記憶装置はチャンク管理ＤＢ７中の複製管理情報から削除されるので、その後の読み出しアドレス変換処理においてアクセス先として決定されることはなく、ホスト１が使用不能となった記憶装置にアクセスすることはなくなる。また、チャンク管理ＤＢ７中の記憶装置情報にも停止中である旨が登録されることにより、チャンク作成処理において配置先とする記憶装置の対象からも外されるので、使用できない記憶装置がチャンクの配置先として割り当てられることもない。従って、データの保全性および可用性が向上する。

【０１７１】

また、チャンク管理部３は、障害対応処理の一環として、複製数を維持するための処理を行ってもよい。なお、複製数を維持するための処理は、例えば、障害対応動作を行った後、システムで規定した時間が経過する、または、負荷がシステムで規定した値を下回っ

たことを契機に自動的に行うようにしてもよい。また、複製数を維持するための処理は、障害対応動作の一環として行うに限らず、記憶装置２が追加される等のシステムで規定した契機や、ユーザからの指示を契機に行うことも可能である。

【０１７２】

図１８は、複製数を維持するための処理の処理フローを示すフローチャートである。図１８に示すように、チャンク管理部３は、まず、チャンク管理ＤＢ７に記録されている各チャンクの複製管理情報から、そのチャンクの複製数と、実際に記憶装置に割り当てられている複製チャンクの数とが一致しないものを検索する（ステップＳ１１０１）。ここで、チャンクの複製数は、システムで規定した値やその複製管理情報に登録されている値によって認識すればよい。また、実際に記憶装置に割り当てられている複製チャンクの数は、その複製管理情報に、チャンクの配置先として登録されている記憶装置ＩＤの数によって認識すればよい。ステップＳ１１０１において、一致しないものが見つかった場合にはステップＳ１１０３に移行し、一致しないものが見つからなかった場合には処理を終了する（ステップＳ１１０２）。

【０１７３】

ステップＳ１１０３では、検索された複製管理情報に基づき、当該チャンクの複製数と実際に記憶装置に割り当てられている複製チャンクの数とを比較し、その差が、複製数のほうが多い場合にはステップＳ１１０５に移行し、逆に複製数のほうが少ない場合にはステップＳ１１０４に移行する。

【０１７４】

ステップＳ１１０４では、当該チャンクの複製数と実際に記憶装置に割り当てられている複製チャンクの数との差をチャンクの追加数として、当該チャンクについて複製チャンク追加処理を行う。なお、複製チャンク追加処理については後述するが、チャンク管理部３はコピー元およびコピー先とする記憶装置を決定するにとどめ、実際の記憶装置間のチャンクのコピー処理は複製処理部５に行わせることが好ましい。このようにすることによって、チャンク管理部３において、障害回復や再配置のための処理による通常のチャンクアクセスにかかる処理への負担が増大しないようにする。

【０１７５】

また、ステップＳ１１０５では、当該チャンクの複製数と実際に記憶装置に割り当てられている複製チャンクの数との差をチャンクの削除数として、当該チャンクについて複製チャンク削除処理を行う。なお、複製チャンク削除処理についても同様に、チャンク管理部３は削除対象とする記憶装置を決定するにとどめ、実際の記憶装置からのチャンクの削除は複製処理部５に行わせることが好ましい。

【０１７６】

当該チャンクについての複製チャンク追加処理または複製チャンク削除処理が完了すると、次の対象チャンクを検索するため、ステップＳ１１０１に移行する。

【０１７７】

なお、複製チャンク追加処理または複製チャンク削除処理は、複数のチャンクについて並行に実行してもよい。さらに、同時にコピー元およびコピー先として処理を行う記憶装置２や削除対象として処理を行う記憶装置２が異なれば処理時間が短縮される。処理時間が短縮されれば、複製数が一致していない期間が短縮される。特に、障害により使用不能になった記憶装置が割当先として削除された場合には、複製チャンクの数の減少によって冗長度が減少していることになるので、その期間が短縮されれば、信頼度が向上する。

【０１７８】

また、本例では、チャンク数追加処理またはチャンク数追加処理だけを複製処理部５に行わせる例を示したが、この複製数を維持するための処理全体を複製処理部５に行わせ、チャンク管理部３は、複製処理部３からの依頼に応じてコピー元およびコピー先チャンクや削除対象チャンクを決定するだけであってもよい。

【０１７９】

また、この複製数を維持するための処理は、障害回復のために行われるだけでなく、複

製数計画部６によって複製数が変更された場合にも行われる。例えば、複製数計画部６は、各コンテンツのアクセス状況に応じて複製数を変更してもよい。複製数計画部６は、例えば、各コンテンツまたはコンテンツ内の各チャンクに求められる複製数の算出を行い、その時点での複製数と算出した複製数とが異なる場合に、チャンク管理ＤＢ７中の該当チャンクの複製管理情報における複製数を更新して、そのチャンクについて、チャンク管理部３に複製数を維持するための処理を行わせてもよい。または、複製数計画部６が、複製数の変更対象をチャンクとして意識しない場合（コンテンツを単位にする場合）や、コンテンツ内における複数のチャンクを対象とする場合には、算出した複製数とコンテンツＩＤと、必要に応じてそのコンテンツ内における対象チャンクの範囲とを指定して、チャンク管理部３に後述の複製数変更処理を行わせてもよい。対象チャンクの範囲の指定は、例えば、コンテンツ内におけるオフセットアドレスとサイズで指定すればよい。

【０１８０】

複製数の算出を行うタイミングとしては、コンテンツへのその時点でのアクセス数がシステムで規定した閾値を超えた場合あるいは閾値を下回った場合、コンテンツのアクセス予想数がシステムで規定した閾値を超えた場合あるいは閾値を下回った場合などがある。また、記憶装置２－１～ｎ中の未使用記憶容量がシステムで規定した閾値を超えた場合あるいは閾値を下回った場合でもよい。

【０１８１】

図１９は、チャンク管理部３における複製数変更処理の処理フローを示すフローチャートである。複製数変更処理は、複製数計画部６から複製数の変更が指示された際に行われる。図１９に示すように、チャンク管理部３は、複製数計画部６から指定されたコンテンツＩＤに対応する分割管理情報をチャンク管理ＤＢ７から読み込む（ステップＳ８０１）。ここで、対象チャンクの範囲としてオフセットアドレスとサイズとが指定されている場合にはステップＳ８０４に移行し、そうでなければステップＳ８０３に移行する（ステップＳ８０２）。

【０１８２】

ステップＳ８０３では、そのコンテンツ内の全チャンクを複製数の変更対象として設定する。一方、ステップＳ８０４では、指定されたオフセットアドレスとサイズからチャンクの範囲を計算によって把握し、把握したチャンクを複製数の変更対象として設定する。チャンクの範囲の特定方法は、読み出しアドレス変換処理のステップＳ４０４において説明した方法と同様である。

【０１８３】

変更対象とするチャンクの範囲を特定すると、変更対象中のチャンクからまだ未処理のチャンクを選ぶ（ステップＳ８０５）。ここで、未処理のチャンクがない場合には処理を終了し、あった場合にはステップＳ８０７に移行する（ステップＳ８０６）。

【０１８４】

ステップＳ８０７では、選択したチャンクに対応した複製管理情報をチャンク管理ＤＢ７から読み込む。次いで、指定された複製数と複製管理情報中の複製数とを比較し、一致した場合にはステップＳ８０５へ、または、指定された複製数のほうが小さかった場合にはステップＳ８１０へ、逆に指定された複製数のほうが大きかった場合にはステップＳ８０９へ移行する（ステップＳ８０８）。

【０１８５】

ステップＳ８０９では、指定された複製数と複製管理情報中の複製数との差をチャンクの追加数として、選択したチャンクについて複製チャンク追加処理を行う。また、ステップＳ８１０では、指定された複製数と複製管理情報中の複製数との差をチャンクの削除数として、選択したチャンクについて複製チャンク削除処理を行う。選択したチャンクについての複製チャンク追加処理または複製チャンク削除処理が完了すると、次のチャンクを選択するため、ステップＳ８０５に移行する。

【０１８６】

なお、ステップＳ８０５以降の処理は並行に実行してもよい。このようにすることで処

理時間は短縮される。さらに、同時にコピー元およびコピー先として処理を行う記憶装置２や削除の処理を行う記憶装置２が異なれば処理時間が短縮される。

【０１８７】

次に、複製チャンク削除処理について説明する。図２０は、チャンク管理部３における複製チャンク削除処理の処理フローの一例を示すフローチャートである。なお、本例では、複製チャンク削除処理は、削除対象となるチャンクのチャンクコンテンツＩＤと現時点での複製管理情報と削除数とが指定されて行われる。図２０に示すように、チャンク管理部３は、まず、対象チャンク（コンテンツ内におけるチャンク）の複製管理情報によって示される、対象チャンクの複製チャンクとして現時点で実際に記憶装置が割り当てられているチャンクの中から、削除対象とする複製チャンクを決定する（ステップＳ９０１）。各複製チャンクは異なる記憶装置に格納されているため、ここでは、削除対象とする記憶装置を決定することとなる。

【０１８８】

削除対象の記憶装置が決定すると、チャンク管理部３は、対象チャンクの複製管理情報から、削除対象とする記憶装置の記憶装置ＩＤを削除する（ステップＳ９０２）。すなわち、対象チャンクの複製管理情報において、その対象チャンクの配置先情報として示されている記憶装置ＩＤのうち削除対象とした記憶装置の記憶装置ＩＤを、未割り当てを示す情報に変更する。そして、変更した複製管理情報をチャンク管理ＤＢ７に反映させる（ステップＳ９０３）。

【０１８９】

次いで、複製処理部５に対し、削除対象の記憶装置内における対象チャンク（削除対象として決定した複製チャンク）の削除処理を指示する（ステップＳ９０４）。例えば、削除対象とする記憶装置の記憶装置ＩＤと、削除対象とする複製チャンクのチャンクアクセスＩＤとを指定して、対象記憶装置内におけるチャンクの削除処理を指示すればよい。そして、複製処理部５からの削除完了の通知を待つ（ステップＳ９０５）。

【０１９０】

指定された削除数分、削除処理が完了した場合には処理を終了し、指定された削除数に満たない場合はステップＳ９０１へ移行する（ステップＳ９０６）。

【０１９１】

なお、本例では、記憶装置内におけるチャンクの削除処理をシーケンシャルに行う例を示しているが、削除対象とする複製チャンクを削除数分決定した上で、複数の記憶装置に対し並行に削除処理を行わせるようにしてもよい。このようにすることで、処理時間が短縮される。なお、記憶装置２でのチャンク削除処理は、それぞれ異なる記憶装置２で行われるため並列処理の処理効率がよい。

【０１９２】

また、ステップＳ９０５でチャンクの削除処理の完了待ちを行ったが、完了を待たずに処理を進めてもよい。このようにすることで、さらに処理時間が短縮される。本例では、チャンク管理ＤＢ７の更新後に記憶装置２に対するチャンクの削除を行っているので、ホスト１がチャンク削除処理中のチャンクへアクセスする可能性を低くでき、ホストからのアクセスの安定性が向上する。

【０１９３】

なお、削除処理を並列に行うと、処理を行う記憶装置２の負荷となる場合がある。そのような場合には、記憶装置２毎に同時に発行する削除処理数の上限を設け、それを越える場合には削除処理の発行を待つようにしてもよい。このようにすることで、記憶装置２の他の処理に使用できる処理能力を保障することができる。すなわちアクセス性能が安定する。

【０１９４】

次に、複製処理部５での記憶装置内におけるチャンクの削除処理を説明する。複製処理部５には、記憶装置ＩＤとチャンクアクセスＩＤとが指定され、対象記憶装置内における対象チャンクを削除する旨が指示される。複製処理部５は、指定された記憶装置ＩＤによ

り特定される記憶装置２－１～ｎのいずれかに対し、指定されたチャンクアクセスＩＤで示されるチャンクの削除要求を発行する。チャンクの削除要求は、具体的には、ブロックデバイスに対するデータ削除命令であったり、ファイルシステムに対するファイルの削除命令であったり、ＯＳＤに対するオブジェクトの削除命令である。

【０１９５】

次に、複製チャンク追加処理（冗長度増加処理）について説明する。図２１は、チャンク管理部３における複製チャンク追加処理の処理フローの一例を示すフローチャートである。なお、複製チャンク追加処理は、追加対象となるチャンクのチャンクコンテンツＩＤと現時点での複製管理情報と追加数とが指定されて行われることとする。図２１に示すように、チャンク管理部３は、まず、追加対象のチャンク（コンテンツ内におけるチャンク）の複製管理情報に基づき、新たな配置先を決定する（ステップＳ１００１）。ステップＳ１００１では、後述する配置先決定方法に従って、新たな配置先とする記憶装置をコピー先記憶装置として決定する。次いで、ステップＳ１００１で割り当てた記憶装置２にデータをコピーする際のコピー元記憶装置を決定する（ステップＳ１００２）。なお、ステップＳ１００１，ステップＳ１００２においては、具体的には、コピー先の記憶装置ＩＤとその記憶装置内でのチャンクアクセスＩＤとを決定したり、コピー元の記憶装置ＩＤとその記憶装置内でのチャンクアクセスＩＤとを決定する。

【０１９６】

次いで、複製処理部５に対し、記憶装置間での対象チャンク（コピー元およびコピー先として決定した複製チャンク）のコピー処理を指示する（ステップＳ１００３）。例えば、コピー先とする記憶装置ＩＤおよびチャンクアクセスＩＤと、コピー元とする記憶装置ＩＩＤおよびチャンクアクセスＩＤとを指定して、対象記憶装置間でのチャンクのコピー処理を指示すればよい。なお、コピー先記憶装置でのチャンクアクセスＩＤが、コピー処理の結果割り当てられる場合には、この時点では指定しなくてもよい。そして、複製処理部５からのコピー完了の通知を待って（ステップＳ１００４）、複製管理情報の対象チャンクの配置先情報に、追加した複製チャンクの配置先を示す情報（ここでは、記憶装置ＩＤおよびチャンクアクセスＩＤ）を追加する（ステップＳ１００５）。そして、変更した複製管理情報をチャンク管理ＤＢ７に反映させる（ステップＳ１００６）。

【０１９７】

指定された増加数分、コピー処理が完了した場合には処理を終了し、指定された増加数に満たない場合はＳ１００１へ移行する（ステップＳ１００７）。

【０１９８】

なお、本例では、記憶装置間でのチャンクのコピー処理をシーケンシャルに処理を行っているが、コピー元およびコピー先対象とする複製チャンクを追加数分決定した上で、複数の記憶装置に対し並行にコピー処理を行わせるようにしてもよい。このようにすることで、処理時間が短縮される。なお、記憶装置２でのチャンクのコピー処理は、コピー先が異なるだけでなく、コピー元もそれぞれ異なる記憶装置２とすることができるので、並列処理の処理効率はよい。

【０１９９】

また、本例では、ステップＳ１００４でチャンクのコピー処理の完了を待った上で、チャンク管理ＤＢ７の更新を行っているので、ホスト１がチャンクコピー処理中のチャンクへアクセスする可能性を低くでき、ホストからのアクセスの安定性が向上する。

【０２００】

次に、複製処理部５での記憶装置間のチャンクのコピー処理を説明する。複製処理部５には、コピー元の記憶装置ＩＤおよびその記憶装置内でのチャンクアクセスＩＤと、コピー先の記憶装置ＩＤおよびその記憶装置内でのチャンクアクセスＩＤとが指定され、対象記憶装置間で対象チャンクをコピーする旨が指示される。複製処理部５は、指定されたコピー元記憶装置ＩＤにより特定される記憶装置に対し、指定されたコピー元チャンクアクセスＩＤで示されるチャンクの読み出し要求を発行する。そして、対象チャンクのデータが読み出されると、指定されたコピー元記憶装置ＩＤにより特定される記憶装置に対し、

指定されたコピー先チャンクアクセスＩＤで示されるチャンクへの読み出したデータの書き込み要求を発行する。なお、コピー先記憶装置に該当チャンクがない場合、チャンクの作成を行う。チャンクの作成はファイルシステムのファイルの作成、あるいはＯＳＤでのオブジェクトの作成と同様である。

【０２０１】

また、本例では、データを複製処理部５がいったん読み出してからコピーしたが、ストレージのサードパーティコピー機能と同様に、複製処理部５を介さずに、コピー元記憶装置とコピー先記憶装置間で直接チャンクのコピーを行ってもよい。

【０２０２】

次に、チャンクの配置先決定方法について説明する。なお、チャンクの配置先決定方法は、図１６に示すチャンク作成処理のステップＳ６０２や、図２１に示す複製チャンク追加処理のステップＳ１００１において、チャンク割り当て処理として実施される。

【０２０３】

まず、同一チャンクの重複配置を防止するための配置先決定方法について説明する。図２２は、チャンク管理部３におけるこの方法でのチャンク割り当て処理の処理フローの一例を示すフローチャートである。なお、本例では、チャンク割り当て処理は、割り当て対象となるチャンクのチャンクコンテンツＩＤと現時点での複製管理情報とが指定されて行われる。なお、チャンク割り当て処理における割当数は１である。図２２に示すように、まず、チャンク管理部３は、割り当て可能な全記憶装置をリストアップする（ステップＳ１２０１）。割り当て可能な記憶装置は、例えば、チャンク管理ＤＢ７中に登録されている記憶装置情報の稼働状況に基づいて、現時点で稼働中の記憶装置の記憶装置ＩＤを集合させた割り当て候補リストを作成すればよい。

【０２０４】

次いで、対象チャンクの複製管理情報で示される、対象チャンクの複製チャンクに既に割り当てられている記憶装置を、割り当て候補リストから削除する（ステップＳ１２０２）。ここで、割り当て候補リストには、現在稼働中であって対象チャンクの複製チャンクに対しまだ割り当てられていない記憶装置がリストアップされていることとなる。最後に、割り当て対象リストから、例えば乱数を用いて１の記憶装置を選び、選んだ記憶装置を配置先として決定する（ステップＳ１２０３）。

【０２０５】

このようにすることで、同じ内容のチャンクが異なる記憶装置に配置される。従って、一つの記憶装置に障害が発生しても他の記憶装置に記憶したチャンクにアクセスすることができるので、データの保全性・可用性が向上する。また、コンテンツ内におけるチャンクと同じ内容のチャンク（同じチャンクコンテンツＩＤのチャンク）が複数あることで、チャンクへのアクセスのスケーラビリティが向上する。

【０２０６】

なお、ステップＳ１２０３では乱数により配置先を決定したが、例えば、各記憶装置が記憶しているチャンク数や未使用データ領域をテーブルで保持しておき、記憶しているチャンクの数が最も少ない記憶装置や、未使用データ領域が最も多い記憶装置に割り当てる方法がある。また、各記憶装置の負荷状態がわかる場合には、最も負荷が低い記憶装置に割り当てる方法もある。

【０２０７】

なお、チャンク管理部３は、ステップＳ１２０において割り当て候補リストから全ての記憶装置が削除される場合を考慮するならば、各々の記憶装置に均等化されるように配置先を決定すればよい。具体的には、ステップＳ１２０２において、割り当て候補リストから記憶装置を削除する前に、その割り当て候補リストに対象チャンクの割当数を登録するようにし、その後、割当数が最も少ない記憶装置以外の記憶装置を割り当て候補から削除すればよい。

【０２０８】

次に、コンテンツ内において連続するチャンクの重複配置を防止するための配置先決定

方法について説明する。図２３は、チャンク管理部３におけるこの方法でのチャンク割り当て処理の処理フローの一例を示すフローチャートである。本例でも、チャンク割り当て処理は、割り当て対象となるチャンクのチャンクコンテンツＩＤと現時点での複製管理情報とが指定されて行われる。なお、チャンク割り当て処理における割当数は１である。なお、図２３に示すステップＳ１３０１，Ｓ１３０２については、図２２におけるステップＳ１２０１，Ｓ１２０２と同様であるため、説明省略している。

【０２０９】

本例では、さらに、対象チャンクのチャンクコンテンツＩＤに対応づけられた分割管理情報から、対象チャンクの前後Ｎ個以内のチャンクを調べ、それらチャンクの複製管理情報で示される、対象チャンクの前後Ｎ個以内のチャンクの複製チャンクに既に割り当てられている記憶装置を、割り当て候補リストから削除する（ステップＳ１３０３）。ここで、割り当て候補リストには、現在稼働中であって対象チャンクの複製チャンクおよびコンテンツ内においてその対象チャンクの前後Ｎ個以内のチャンクの複製チャンクに対しまだ割り当てられていない記憶装置がリストアップされていることとなる。最後に、割り当て対象リストから１の記憶装置を選び、選んだ記憶装置を配置先として決定する（ステップＳ１３０４）。なお、割り当て対象リストからの選択動作については、ステップＳ１２０３と同様である。

【０２１０】

このようにすることで、同じ内容のチャンクが異なる記憶装置に配置されるだけでなく、コンテンツ中の前後Ｎ個以内のチャンクと同じ内容のチャンクとも異なる記憶装置に配置される。従って、例えば、チャンクアクセスがシーケンシャルに行われる場合にも、各アクセスが異なる記憶装置に分散される。すなわち、シーケンシャルアクセス時のスループットが向上する。

【０２１１】

なお、前後Ｎ個以内のＮは、１つであってもよいが、例えば、バッファを使用して先読みを行うようなコンテンツを扱う場合には、先読みで同時にアクセスする量に相当するチャンクが含まれるように規定することが好ましい。そうすることで、先読みアクセスにおいてもスループットの向上とレスポンスタイムの短縮が期待される。

【０２１２】

次に、記憶装置間の相関を平均化するための配置先決定方法について説明する。、例えば、記憶装置内のチャンクの配置内容が、他の記憶装置内のチャンクの配置内容とを一致させないように配置決定する方法である。チャンク管理部３は、２つの記憶装置間で共有する同じ内容の複製チャンクの数を共有数とした場合に、各々の記憶装置の組み合わせにおける共有数が均等化されるように、複製チャンクの配置先を決定すればよい。このようにすることで、記憶装置内のチャンクの配置パターンが均等化され、ある記憶装置が障害等により使用できなくなった際に他の稼動できる記憶装置への代替による負荷が均等化する。

【０２１３】

実現方法の１つを説明する。チャンク管理ＤＢ７に記憶装置間の共通度を格納するテーブルを用意する。

【０２１４】

このテーブルの操作について説明する。このテーブルは、記憶装置数ｎ×ｎの二次元配列テーブルであって、２記憶装置間で共有するチャンク数（以下、単に共有数という。）をその２記憶装置の記憶装置ＩＤを添字とする配列要素に示すテーブルである。以下、このテーブルを配列Ａと呼ぶ。配列Ａの各配列要素の初期値は全て０とする。なお、ここでは、記憶装置ＩＤには０からの整数が割り当てられているものとする。チャンク管理部３は、同一内容の複数のチャンクを一度に配置決定した場合、同一内容のチャンクの配置先となった記憶装置ＩＤの組み合わせを添字とする配列要素で示される共有数を＋１する。例えば、記憶装置２－ａ，ｃ，ｋに同一内容のチャンクを割り当てた場合には、配列Ａ［ａ］［ｃ］，Ａ［ｃ］［ａ］，Ａ［ａ］［ｋ］，Ａ［ｋ］［ａ］，Ａ［ｃ］［ｋ］，Ａ［

ｋ］［ｃ］に、１を加算する。

【０２１５】

また、チャンク管理部３は、あるチャンクを記憶装置から削除した場合には、そのチャンクの配置先であった記憶装置ＩＤと、現時点でそのチャンクと同一内容のチャンクが割り当てられている他の記憶装置ＩＤとの組み合わせを添字とする配列要素で示される共有数を－１する。例えば、記憶装置２－ａ，ｃ，ｋに割り当てられているチャンクのうち、記憶装置２－ｋに割り当てられているチャンクを削除した場合には、配列Ａ［ａ］［ｋ］，Ａ［ｋ］［ａ］，Ａ［ｃ］［ｋ］，Ａ［ｋ］［ｃ］から１を減算する。

【０２１６】

また、チャンク管理部３は、あるチャンクを記憶装置に追加した場合には、そのチャンクの配置先となった記憶装置ＩＤと、現時点でそのチャンクと同一内容のチャンクが割り当てられている他の記憶装置ＩＤとの組み合わせを添字とする配列要素で示される共有数を＋１する。例えば、記憶装置２－ａ，ｃ，ｋに配置されているチャンクと同じ内容のチャンクを新たに記憶装置２－ｈに割り当てた場合、配列Ａ［ａ］［ｈ］，Ａ［ｃ］［ｈ］，Ａ［ｋ］［ｈ］，Ａ［ｈ］［ａ］，Ａ［ｈ］［ｃ］，Ａ［ｈ］［ｋ］に１を加算する。

【０２１７】

チャンク管理部３が新たなチャンクの配置先を決定する場合には、このように扱われるテーブルを用いて、そのチャンクと同一内容のチャンクが既に割り当てられている記憶装置と最も相関の低い記憶装置を配置先として決定する。具体的には、既に記憶装置に割り当てられているチャンクと同一のチャンクの配置を決定する場合には、そのチャンクが割り当てられている記憶装置ＩＤを添字の１つとして固定して配列Ａの各配列要素を見た場合に、共有数が最も小さくなる添字を示す記憶装置ＩＤを配置先として決定する。なお、既に記憶装置に割り当てられているチャンクが複数ある場合には、各記憶装置ＩＤを添字の１つにそれぞれ固定した複数の配列Ａの和として見ればよい。例えば、既に記憶装置２－ａ，ｃに配置されているチャンクと同一内容のチャンクの配置先を決定する場合には、配列Ａ［ａ］［ｉ］＋配列Ａ［ｃ］［ｉ］が最小となるｉを示す記憶装置２－ｉを配置先として決定すればよい。

【０２１８】

なお、チャンクの割り当て処理としては、例えば、図２２に示すステップＳ１２０３や図２３に示すステップＳ１３０４において、割り当て候補リストにある記憶装置の中から、上記ｉを満たす記憶装置を選べばよい。なお、複数の記憶装置でｉを満たす場合には、乱数等により配置先を決定すればよい。

【０２１９】

次に、チャンク管理部３における読み出し元記憶装置の決定方法について説明する。なお、読み出し元記憶装置決定方法は、図２１に示す冗長度増加処理のステップＳ１００２におけるコピー元複製チャンクの決定動作として実施される。チャンク管理部３は、各記憶装置で行っているチャンクのコピー処理の数をテーブルにより保持しているものとする。図２１に示す例では、ステップＳ１００３でコピー元記憶装置のコピー処理数を＋１し、ステップＳ１００４が完了するとその記憶装置のコピー処理数を－１すればよい。

【０２２０】

そして、チャンク管理部３は、読み出し元を決定する際に、上記テーブルを参照して、追加対象チャンクが割り当てられている記憶装置の中で最もコピー数が少ない記憶装置をコピー元記憶装置に決定すればよい。なお、選択候補の記憶装置のコピー数がシステムで規定した数よりも大きかった場合に、コピー数が規定数未満になるまでコピー処理を待つ方法もある。このようにすることで、コピーに使用する帯域を制限でき、コピー処理中の記憶装置で他のアクセスに使用する帯域を確保することができる。よって、アクセスの安定性が保証される。

【０２２１】

また、上記例では、コピー元記憶装置についてコピー数を管理しているが、コピー先の記憶装置についてもコピー数を管理し、コピー先記憶装置のコピー数がシステムで規定し

た数よりも大きかった場合に、コピー処理を待つ方法もある。このようにすることで、コピー先の記憶装置で他の処理に使用できる処理能力を保障することができる。よって、アクセス性能が安定する。

【０２２２】

また、障害等により使用できなくなった記憶装置の代替として新規に記憶装置を追加した場合など、割り当て先を代替用に追加した記憶装置に固定する方法もある。更に代替記憶装置にコピーを行う期間中は、読み出し対象としないなど代替記憶装置をコピー以外のアクセスの対象外とするとともに、他の記憶装置からのコピー処理を並行に行う方法がある。このようにすると、代替記憶装置へのコピー処理については、代替記憶装置のバンド幅全てが使用できるため、コピー処理の時間を短縮できる。すなわち、冗長度が減少している時間が短縮され、可用性が向上する。また、代替記憶装置が短い時間で復帰すれば、使用不能となった記憶装置に割り当てられたチャンクアクセスの代替による負荷がかかる記憶装置へのアクセス性能も安定する。

【０２２３】

次に、チャンク管理部３における削除対象記憶装置の決定方法について説明する。なお、削除対象記憶装置の決定方法は、図２０に示す複製チャンク数削除処理のステップＳ９０１における削除対象の複製チャンクの決定動作として実施される。

【０２２４】

チャンク管理部３は、例えば、対象チャンクの複製管理情報を参照し、対象チャンクの複製チャンクが割り当てられている記憶装置の中から乱数により１の記憶装置を選択してもい。また、例えば、各記憶装置が記憶しているチャンク数や未使用データ領域をテーブルで保持しておき、記憶しているチャンクの数が最も多い記憶装置や、未使用データ領域が最も少ない記憶装置を削除対象とする方法もある。また、各記憶装置の負荷状態がわかる場合には、最も負荷が高い記憶装置を削除対象とする方法もある。

【０２２５】

次に、複製数計画部６におけるチャンクの複製数の決定方法について説明する。複製数は、可用性・保全性から決定される可用性面での複製数と、スループット・レスポンスタイム等から決定されるアクセス面での複製数とがあるが、ここでは、両者の大きい方の値を複製数として決定する。

【０２２６】

可用性面での複製数は、記憶装置が障害等によりアクセスできなくなった際に、障害で使えなくなった記憶装置に格納されていたチャンクと同じチャンクが他のいずれかの記憶装置でアクセス可能であることを保障するための数である。

【０２２７】

さらに、障害発生後にコピーを作成して複製数を維持する処理を行う場合、そのコピーが完了するまでの間に他の記憶装置にも障害が発生する可能性を考慮して決定される。コピー完了までの時間が長くなればなるほど、可用性面の複製数は多くなる。すなわち、記憶装置の容量とコピーに使用できる帯域と記憶装置の信頼度から算出される値である。可用性面での複製数は記憶装置が１つ使用できなくなっても冗長度が０とならないよう、３以上と定める。

【０２２８】

アクセス面での複製数は、コンテンツの要求予測量と、各記憶装置の処理能力と、システムで規定した安全係数とに基づいて決定される。例えば、そのコンテンツの必要スループットを各記憶装置のスループットで割ったものにシステムで規定した安全係数をかけたものである。ここで、安全係数とは、予想以上に要求が増えた場合に備えた安全係数の意と、複数のコンテンツで記憶装置を共有することによるそのコンテンツでの記憶装置の占有率の意を含めた係数である。安全係数は、１以下の値を設定する。コンテンツの必要スループットは、コンテンツのエンコードのビットレートと要求ユーザ数との積によって求めてもよい。なお、記憶装置の平均スループットや最低スループットを用いてもよい。または、配置決定した記憶装置のスループットの和をとって、必要スループットに足りなか

ったら複製数を追加するという方法もある。

【０２２９】

また、コンテンツの先頭やコンテンツにチャプタが付いている場合のチャプタの先頭に位置するチャンクの複製数を多くする方法もある。ビデオのようなメディアコンテンツのアクセスではバッファを用いてアクセスの安定性を確保するが、コンテンツの先頭やチャプターの直後はバッファによる安定化ができないため、例えば、コンテンツの先頭やチャプターの直後に該当するデータを含むチャンクについては、他のチャンクよりも複製数を多くすることで再生の安定化が図れる。

【０２３０】

次に、複製チャンクの再配置について説明する。複製チャンクの再配置を行う契機としては、例えば、ユーザが指示した場合や、記憶装置が追加された場合、記憶装置が削除された場合、ある記憶装置の記憶容量の使用率またはｂｕｓｙ率が閾値（８０％，９０％等）を超えた場合、ある記憶装置の使用率（またはｂｕｓｙ率）とシステム全体の平均使用率（またはｂｕｓｙ率）との差が閾値を超えた場合などが考えられる。

【０２３１】

例えば、記憶装置が追加された場合、まず、チャンク管理部３が、チャンク管理ＤＢ７に記憶されている各チャンクの複製管理情報において、追加された記憶装置以外の記憶装置が配置先として割り当てられているチャンク（複製チャンク）をランダムに抽出する。そして、抽出されたチャンクを追加された記憶装置に格納できるか否かを判断する。ここで、格納できないチャンクの例としては、追加された記憶装置を既に格納先としたチャンクと同一内容のチャンクや、追加された記憶装置を既に格納先としてチャンクの前後Ｎ個以内のチャンクや、追加された記憶装置に記憶することである記憶装置間の相関が高くなるようなチャンクである。

【０２３２】

抽出したチャンクが追加された記憶装置に格納できる場合には、そのチャンクを対象チャンクとして新たな記憶装置に移動させる。格納できない場合には、他のチャンクを再度抽出する。チャンクの移動方法としては、例えば、複製処理部５に対象チャンクのコピー処理を行わせたのち、その対象チャンクの配置先情報を、元の記憶装置から新たな記憶装置に変更してチャンク管理ＤＢ７に反映させる。そして、複製処理部５に元の記憶装置からの対象チャンクの削除処理を行わせればよい。以上の動作を、例えば、システムで規定されているチャンク数Ｍ分だけ繰り返し行うようにすればよい。なお、チャンク数Ｍは、システムに予め設定された値を用いてもよいし、再配置を行う際にユーザが指定した値を用いてもよい。また、追加された記憶装置の容量に応じて決定する方法もある。

【０２３３】

このようにすることで、追加された記憶装置に即座にチャンクが割り当てられ、追加された記憶装置が使用されるようになる。また、追加された記憶装置に対してもチャンクの配置先決定ルールが適用されるので、可用性面での複製数が低くなることも、アクセス面での複製数が低くなることも、記憶装置間の相関が高くなることもない。なお、上記で示したチャンクの移動方法によれば、移動中のチャンクにアクセスされる可能性も低い。

【０２３４】

また、記憶装置が削除された場合の再配置動作は、使用不能になった記憶装置に対する障害対処理における複製数を維持するための処理と同様である。

【０２３５】

また、それ以外の場合の再配置動作としては、例えば、チャンク管理部３が、任意の記憶装置または使用率やｂｕｓｙ率において高負荷が検出された記憶装置から、移動対象とするチャンクをランダムに抽出する。次に、移動対象チャンクを抽出した記憶装置以外の記憶装置から、移動先とする記憶装置をランダムに選択する。そして、移動対象チャンクを移動先として選択した記憶装置に格納できるか否かを判断する。

【０２３６】

抽出した移動対象チャンクが移動先として選択した記憶装置に格納できる場合には、そ

のチャンクを選択した記憶装置に移動させる。格納できない場合には、移動先とする記憶装置を再度選択する。再度選択した記憶装置にも格納できない場合や全記憶装置に格納できない場合には、移動対象チャンクを再度抽出する。なお、チャンクの移動方法は、既に説明した方法と同様である。

【０２３７】

このようにすることで、システム全体としての記憶装置間の相関を更に低くすることができる。さらに、再配置に伴って可用性面での複製数が低くなることも、アクセス面での複製数が低くなることもない。

【０２３８】

以上のように、本実施の形態によれば、コンテンツを複数のチャンクに分割し、さらにそのチャンクを複製してできる複製チャンクを、所定の配置先決定ルールに従って複数の記憶装置に分散させて格納することによって、データ分散格納における特性・性能の向上を図ることができる。

【０２３９】

例えば、チャンク管理部３が、同一内容のチャンクがそれぞれ異なる記憶装置に格納されるように各チャンクの配置を決定することによって、記憶装置が障害により使用できない場合であってもその記憶装置に格納されているチャンクと同一内容のチャンクを格納している他の記憶装置でそのチャンクアクセスが可能とあり、可用性を向上させることができる。また、チャンク管理部３が複数チャンクのアクセス先を決定すれば、ホストからの読み出しを確実に分散させることができ、アクセス性を向上させることができる。

【０２４０】

また、例えば、チャンク管理部３が、コンテンツ内において連続するチャンクと同一内容のチャンクがそれぞれ異なる記憶装置に格納されるように各チャンクの配置を決定することによって、さらにシーケンシャルアクセスでのアクセス性を向上させることができる。

【０２４１】

また、例えば、チャンク管理部３が、上記条件に加えて、記憶装置間の相関を低くするように配置決定することによって、障害等により使用できなくなった記憶装置に格納されていたチャンクへのアクセスを、他の記憶装置全体に平均的に分散することができる。従って、障害発生時のアクセス性低下の度合いを低減させることができる。また、障害発生後に複製数を維持させる場合であっても、そのためのコピー処理におけるコピー元が他の記憶装置全体に分散されるので、並列処理を可能とし、回復に要する時間を短縮することができる。すなわち、障害発生後の可用性低下の度合いを低減させることができる。なお、連続するチャンクと同一内容の複製チャンクとの記憶装置の共有を防ぎつつ、記憶装置間の相関を低くするように配置決定することも可能である。

【０２４２】

また、チャンクの記憶装置群への配置管理をチャンク管理部３で行うことにより、コンテンツが形成する論理データ空間と記憶装置群が形成する物理データ空間とのマッピングが柔軟になり、記憶装置の追加に対して柔軟に対応できる。

【０２４３】

また、チャンク管理部３が障害対応動作として使用できなくなった記憶装置をアクセス対象外からはずすことによって、ホストが障害によりアクセスできなくなった記憶装置にアクセスすることがなくなる。すなわち障害発生時のアクセス失敗の可能性を低減することができる。

【０２４４】

また、複製数計画部６が、コンテンツへのアクセス要求の変化に即して、コンテンツの複製数を変更することによって、変化するコンテンツへのデマンドに対処できかつ、記憶容量の無駄を抑えることが可能となる。すなわち、変化するコンテンツのデマンドに適したアクセス性を提供しながら容量効率を向上できる。なお、複製数画部６がコンテンツ内におけるチャンクの位置に応じて複製数を決定することによって、コンテンツの位置によ

り異なるデマンドに対応できる。すなわち、変化するコンテンツのデマンドに適したアクセス性を提供しながら容量効率を向上できる。

【産業上の利用可能性】

【０２４５】

本発明の活用例として、ストリーミング配信サーバがある。特に、不特定多数のユーザに向けて多量のコンテンツをサービス対象とするストリーミング配信サーバのように、多くの記憶装置を制御対象とするデータ分散格納システムに好適に適用可能である。

【図面の簡単な説明】

【０２４６】

【図１】本発明によるデータ分散格納システムの構成例を示すブロック図である。

【図２】本実施の形態によるデータ分散格納システムの構成例を示すブロック図である。

【図３】分割管理情報の例を示す説明図である。

【図４】複製管理情報の一例を示す説明図である。

【図５】チャンクの連結結果の一例を示す説明図である。

【図６】チャンクの連結結果の一例を示す説明図である。

【図７】コンテンツ読み出し動作の一例を示すフローチャートである。

【図８】読み出し用アドレス変換要求に対する応答フォーマットの一例を示す説明図である。

【図９】コンテンツ書き込み動作の一例を示すフローチャートである。

【図１０】書き込み用アドレス変換要求に対する応答フォーマットの一例を示す説明図である。

【図１１】チャンク管理部３におけるアドレス変換処理の処理フローの一例を示すフローチャートである。

【図１２】読み出しアドレス変換処理の処理フローの一例を示すフローチャートである。

【図１３】読み出し用アドレス変換要求に対する応答フォーマットの他の例を示す説明図である。

【図１４】読み出し用アドレス変換要求に対する応答フォーマットの他の例を示す説明図である。

【図１５】書き込みアドレス変換処理の処理フローの一例を示すフローチャートである。

【図１６】チャンク管理部３におけるチャンク作成処理の処理フローの一例を示すフローチャートである。

【図１７】チャンク管理部３における障害対応処理の処理フローの一例を示すフローチャートである。

【図１８】複製数を維持するための処理の処理フローを示すフローチャートである。

【図１９】チャンク管理部３における複製数変更処理の処理フローを示すフローチャートである。

【図２０】チャンク管理部３における複製チャンク削除処理の処理フローの一例を示すフローチャートである。

【図２１】チャンク管理部３における複製チャンク追加処理の処理フローの一例を示すフローチャートである。

【図２２】チャンク管理部３におけるチャンク割り当て処理の処理フローの一例を示すフローチャートである。

【図２３】チャンク管理部３におけるチャンク割り当て処理の処理フローの一例を示すフローチャートである。

【符号の説明】

【０２４７】

１　ホスト

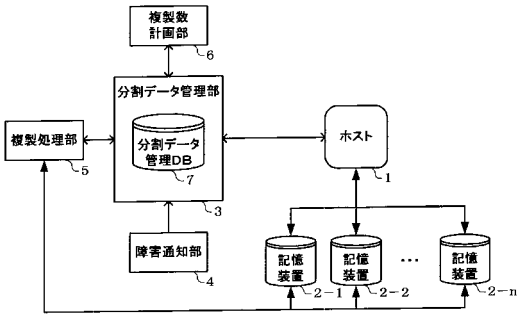
２－１～ｎ　記憶装置

３　分割データ管理部、チャンク管理部

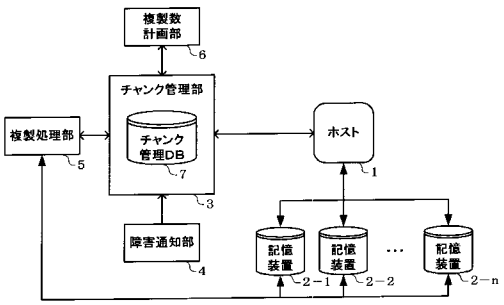
４　障害通知部

5　複製処理部
6　複製数計画部
7　分割データ管理ＤＢ、チャンク管理ＤＢ

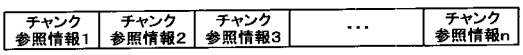
【図１】



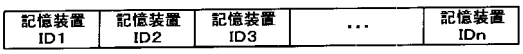
【図２】



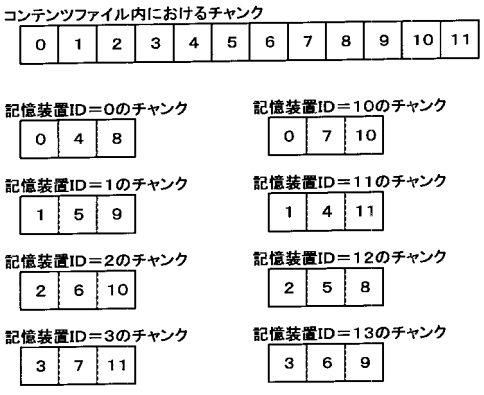
【図３】



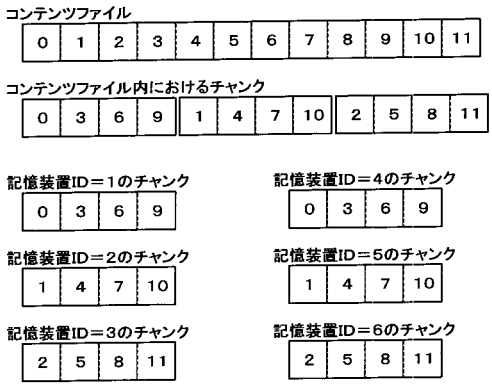
【図４】



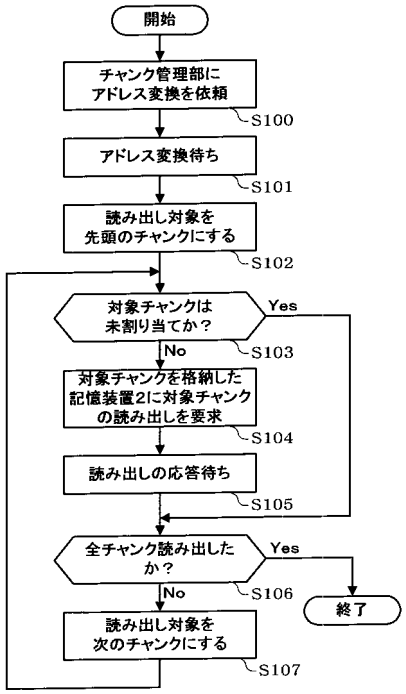
【図５】



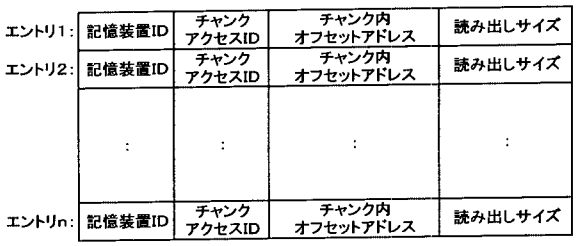
【図６】


コンテンツファイル
0 1 2 3 4 5 6 7 8 9 10 11
コンテンツファイル内におけるチャンク
0 3 6 9 1 4 7 10 2 5 8 11
記憶装置ID＝1のチャンク
0 3 6 9
記憶装置ID＝2のチャンク
1 4 7 10
記憶装置ID＝3のチャンク
2 5 8 11
記憶装置ID＝4のチャンク
0 3 6 9
記憶装置ID＝5のチャンク
1 4 7 10
記憶装置ID＝6のチャンク
2 5 8 11

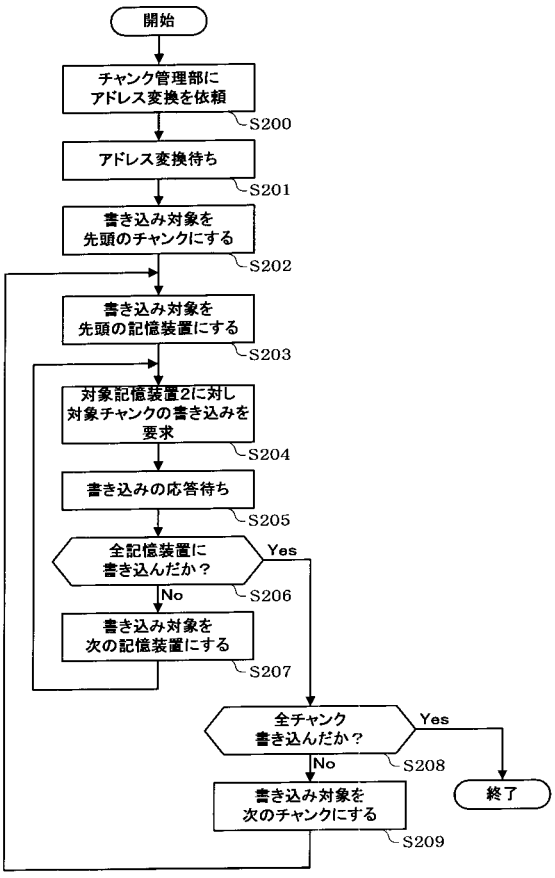
【図７】


開始
チャンク管理部にアドレス変換を依頼
S100
アドレス変換待ち
S101
読み出し対象を先頭のチャンクにする
S102
対象チャンクは未割り当てか？
Yes
No
S103
対象チャンクを格納した記憶装置2に対象チャンクの読み出しを要求
S104
読み出しの応答待ち
S105
全チャンク読み出したか？
Yes
No
S106
終了
読み出し対象を次のチャンクにする
S107

【図８】


エントリ1：
記憶装置ID
チャンクアクセスID
チャンク内オフセットアドレス
読み出しサイズ
エントリ2：
記憶装置ID
チャンクアクセスID
チャンク内オフセットアドレス
読み出しサイズ
：
：
：
：
エントリn：
記憶装置ID
チャンクアクセスID
チャンク内オフセットアドレス
読み出しサイズ

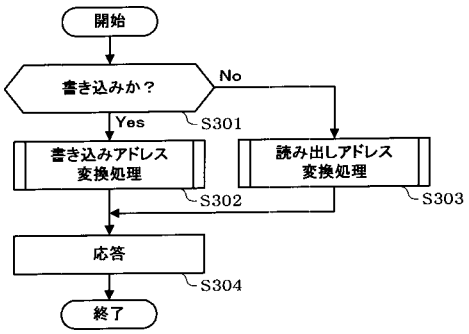
【図９】


開始
チャンク管理部にアドレス変換を依頼
S200
アドレス変換待ち
S201
書き込み対象を先頭のチャンクにする
S202
書き込み対象を先頭の記憶装置にする
S203
対象記憶装置2に対し対象チャンクの書き込みを要求
S204
書き込みの応答待ち
S205
全記憶装置に書き込んだか？
Yes
No
S206
書き込み対象を次の記憶装置にする
S207
全チャンク書き込んだか？
Yes
No
S208
終了
書き込み対象を次のチャンクにする
S209

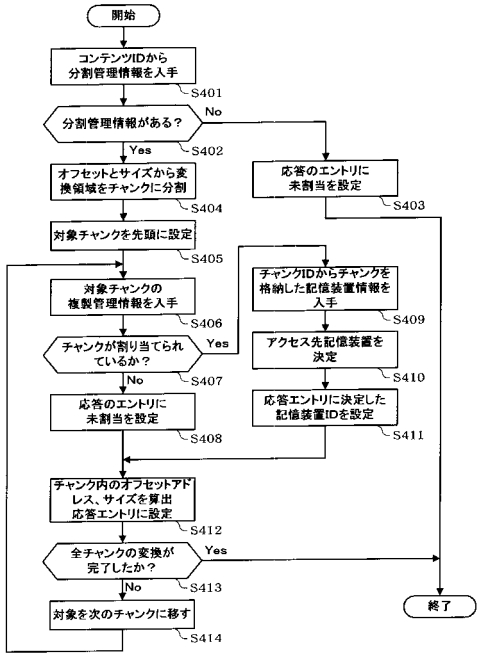
【図１０】

エントリ1:	書き込みサイズ	チャンク内オフセットアドレス	複製数	記憶装置ID	チャンクアクセスID	…	記憶装置ID	チャンクアクセスID
エントリ2:	書き込みサイズ	チャンク内オフセットアドレス	複製数	記憶装置ID	チャンクアクセスID	…	記憶装置ID	チャンクアクセスID
	:	:	:	:				
エントリn:	書き込みサイズ	チャンク内オフセットアドレス	複製数	記憶装置ID	チャンクアクセスID	…	記憶装置ID	チャンクアクセスID

【図１１】



【図１２】



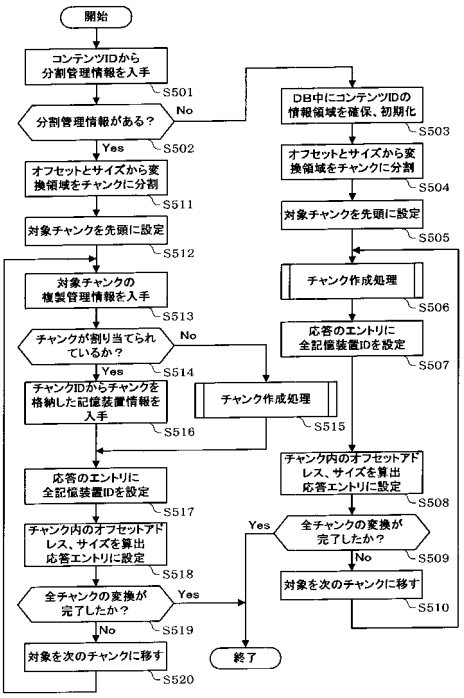
【図１３】

エントリ1:	チャンク内オフセットアドレス	読み出しサイズ	複製数	記憶装置ID	チャンクアクセスID	…	記憶装置ID	チャンクアクセスID
エントリ2:	チャンク内オフセットアドレス	読み出しサイズ	複製数	記憶装置ID	チャンクアクセスID	…	記憶装置ID	チャンクアクセスID
	:	:	:	:				
エントリn:	チャンク内オフセットアドレス	読み出しサイズ	複製数	記憶装置ID	チャンクアクセスID	…	記憶装置ID	チャンクアクセスID

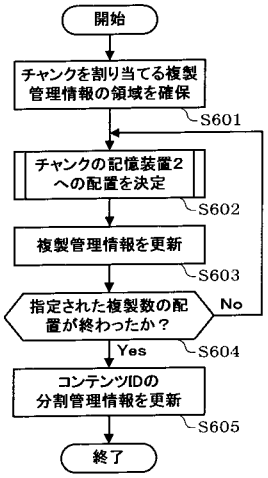
【図１４】

エントリ1:	チャンク内オフセットアドレス	読み出しサイズ	複製数	チャンクアクセスID	記憶装置ID	…	記憶装置ID
エントリ2:	チャンク内オフセットアドレス	読み出しサイズ	複製数	チャンクアクセスID	記憶装置ID	…	記憶装置ID
	:	:	:	:			
エントリn:	チャンク内オフセットアドレス	読み出しサイズ	複製数	チャンクアクセスID	記憶装置ID	…	記憶装置ID

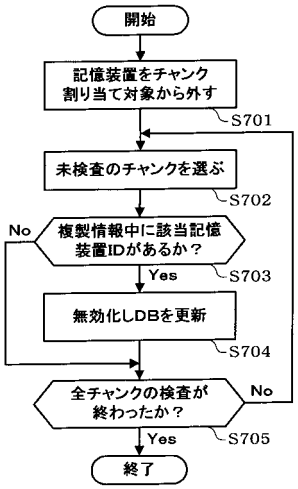
【図１５】



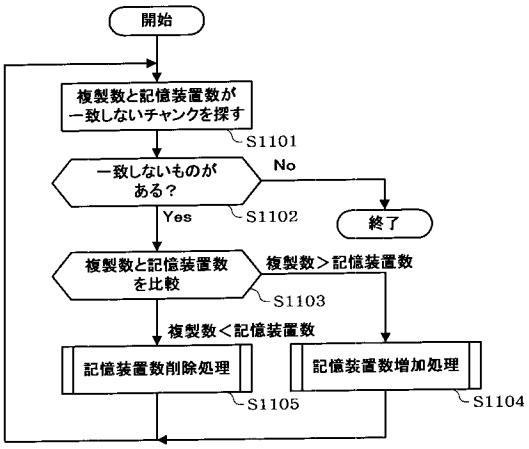
【図１６】



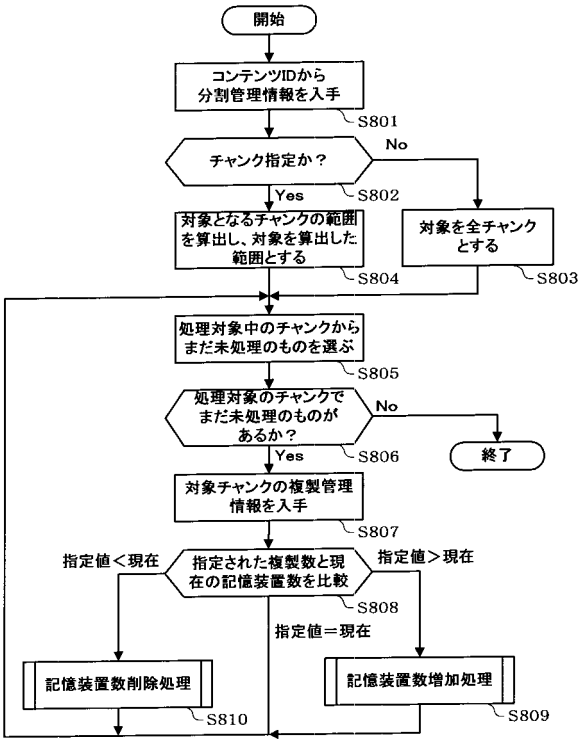
【図１７】



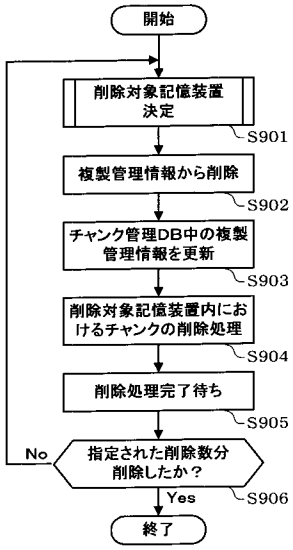
【図１８】



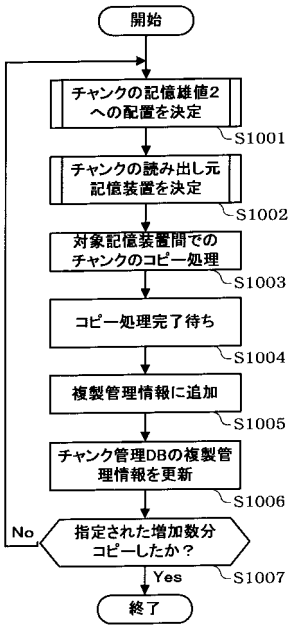
【図１９】



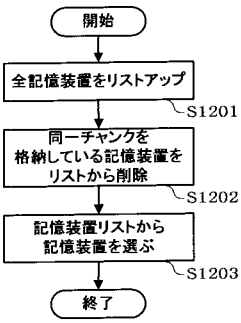
【図２０】



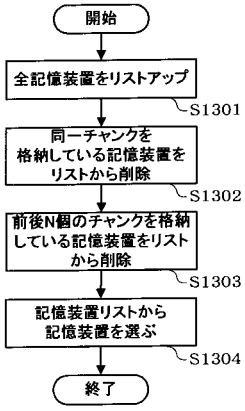
【図２１】



【図２２】



【図２３】



フロントページの続き

(56)参考文献 特開平０７－１４１２３２（ＪＰ，Ａ）
特開２００６－３５０４７０（ＪＰ，Ａ）
特開平０９－０９１１８６（ＪＰ，Ａ）
特開２００４－０４６３５２（ＪＰ，Ａ）
特開２００５－０７１２４７（ＪＰ，Ａ）
特開２００１－２４３０９９（ＪＰ，Ａ）
藤田　憲治，ディスク・アレイ導入の勧め，日経バイト，日本，日経ＢＰ社，１９９９年　４月２２日，第１９０号，pp.104～117
ＩＢＭ　ＳｅｒｖｅＲＡＩＤ　ユーザーズ・リファレンス　第４版，日本，日本アイ・ビー・エム株式会社，２００７年　１月３１日，第4版，pp.24～26

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１２／００
Ｇ０６Ｆ　３／０６
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩＩ）