

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6097845号
(P6097845)

(45) 発行日 平成29年3月15日(2017.3.15)

(24) 登録日 平成29年2月24日(2017.2.24)

(51) Int. Cl.	F I
G06F 12/0868 (2016.01)	G06F 12/08 501F
G06F 12/08 (2016.01)	G06F 12/08 541C
G06F 3/06 (2006.01)	G06F 12/08 557
G06F 12/16 (2006.01)	G06F 3/06 302A
	G06F 3/06 304Z
請求項の数 13 (全 31 頁) 最終頁に続く	

(21) 出願番号	特願2015-554332 (P2015-554332)	(73) 特許権者	000005108
(86) (22) 出願日	平成25年12月24日(2013.12.24)		株式会社日立製作所
(86) 国際出願番号	PCT/JP2013/084502		東京都千代田区丸の内一丁目6番6号
(87) 国際公開番号	W02015/097751	(74) 代理人	110000279
(87) 国際公開日	平成27年7月2日(2015.7.2)		特許業務法人ウィルフォート国際特許事務所
審査請求日	平成27年12月1日(2015.12.1)		所
		(72) 発明者	岡田 尚也
			東京都千代田区丸の内一丁目6番6号 株
			株式会社日立製作所内
		(72) 発明者	野中 裕介
			東京都千代田区丸の内一丁目6番6号 株
			株式会社日立製作所内
		(72) 発明者	荒木 亮彦
			東京都千代田区丸の内一丁目6番6号 株
			株式会社日立製作所内
			最終頁に続く

(54) 【発明の名称】 不揮発性キャッシュメモリにデータをバックアップするストレージシステム

(57) 【特許請求の範囲】

【請求項1】

記憶デバイスと、
 ホストコンピュータからデータを受信し、受信したデータを前記記憶デバイスに格納するストレージコントローラと
 を備えるストレージシステムであって、

前記ストレージコントローラは、プロセッサと、前記プロセッサに接続され前記データを一時的に格納する揮発性の第1のキャッシュメモリと、前記マイクロプロセッサに接続され前記データを一時的に格納する不揮発性の第2のキャッシュメモリと、電断が起こった場合に少なくとも前記プロセッサおよび前記第1のキャッシュメモリに電力を供給するバッテリーと、を備え、

前記第2のキャッシュメモリは、前記記憶デバイスに格納されていないデータであるパーティデータを格納するパーティデータ領域と、前記パーティデータ領域以外の残りの領域とを含み、

前記プロセッサは、電断が起こった場合、前記第1のキャッシュメモリに格納されている全部または一部のデータを対象データとして、前記第2のキャッシュメモリの前記残りの領域に格納する
 ストレージシステム。

【請求項2】

前記第2のキャッシュメモリの前記残りの領域は、前記記憶デバイスに格納されている

データであるクリーンデータを格納するクリーンデータ領域と、データが格納されていない空き領域とを含み、

前記プロセッサは、電断が起こった場合、前記第1のキャッシュメモリの容量と、前記第2のキャッシュメモリの前記クリーンデータ領域および前記空き領域の合計容量とを比較し、

前記第1のキャッシュメモリの容量が前記合計容量以下の場合、前記対象データは、前記第1のキャッシュメモリに格納されている全てのデータであり、

前記第1のキャッシュメモリの容量が前記合計容量より大きい場合は、前記対象データは、前記第1のキャッシュメモリに格納されているデータのうちのパーティデータのみである、

10

請求項1に記載のストレージシステム。

【請求項3】

前記プロセッサは、電断が起こった場合、

前記対象データが前記第1のキャッシュメモリのどこに格納されているのかを示す対象データテーブルと、前記対象データを前記第2のキャッシュメモリのどこに格納するのかを示す退避先データテーブルとを前記第1のキャッシュメモリ上に作成し、

前記対象データテーブルおよび前記退避先データテーブルを前記対象データとともに前記第2のキャッシュメモリの前記残りの領域に格納する

請求項1に記載のストレージシステム。

【請求項4】

20

前記第2のキャッシュメモリの前記残りの領域は、前記記憶デバイスに格納されているデータであるクリーンデータを格納するクリーンデータ領域と、データが格納されていない空き領域とを含み、

前記プロセッサは、前記クリーンデータ領域および前記空き領域のうち、前記空き領域に優先的に前記対象データの格納先にする

請求項1に記載のストレージシステム。

【請求項5】

前記クリーンデータ領域は、複数の領域を有し、

前記プロセッサは、前記複数の領域のうち大きいサイズ領域から順に前記対象データの格納先にする

30

請求項4に記載のストレージシステム。

【請求項6】

前記ストレージコントローラは、

前記第1のキャッシュメモリの容量と第2のキャッシュメモリの容量とを比較し、

前記第1のキャッシュメモリの容量が前記第2のキャッシュメモリの容量よりも小さい場合、前記第1のキャッシュメモリの容量と第2のキャッシュメモリの容量との差分を、前記第2のキャッシュメモリで保持できるパーティデータの最大容量に設定し、

前記第1のキャッシュメモリの容量が前記第2のキャッシュメモリの容量以上である場合、前記第1のキャッシュメモリで保持できるパーティデータの最大容量と前記第2のキャッシュメモリの容量とを比較し、

40

前記第2のキャッシュメモリの容量が前記第1のキャッシュメモリで保持できるパーティデータの最大容量より小さい場合、前記第1のキャッシュメモリで保持できるパーティデータの最大容量を前記第2のキャッシュメモリの容量と等しくし、前記第2のキャッシュメモリが保持できるパーティデータの最大容量を0にし、

前記第2のキャッシュメモリの容量が前記第1のキャッシュメモリで保持できるパーティデータの最大容量以上である場合、前記第2のキャッシュメモリの容量と前記第1のキャッシュメモリで保持できるパーティデータの最大容量との差分を、前記第2のキャッシュメモリで保持できるパーティデータの最大容量に設定する

請求項1乃至請求項5のいずれか一項に記載のストレージシステム。

【請求項7】

50

前記ストレージコントローラは、前記プロセッサに接続される不揮発性媒体を更に備え

、
前記プロセッサは、電断が起こった場合、前記対象データを、前記第2のキャッシュメモリの前記残りの領域および前記不揮発性媒体に分けて格納する請求項1に記載のストレージシステム。

【請求項8】

前記プロセッサは、前記第1のキャッシュメモリから前記第2のキャッシュメモリへの単位時間当たりのデータ転送量である第1の転送スループット、および、前記第1のキャッシュメモリから前記不揮発性媒体への単位時間当たりのデータ転送量である第2の転送スループットを算出し、

10

前記プロセッサは、前記第1の転送スループットおよび前記第2の転送スループットに基づき、前記第2のキャッシュメモリの前記残りの領域に格納する第1のデータ量および前記不揮発性媒体に格納する第2のデータ量を決定する

請求項7に記載のストレージシステム。

【請求項9】

前記対象データは、前記第2のキャッシュメモリにおいて、複数のセグメントに分けて格納され、

前記プロセッサは、電断発生後、電断が解消された場合に、ダーティデータを含む前記格納領域に対してパリティが生成されているか否かを前記セグメント毎に判断し、

前記プロセッサは、前記パリティが生成されていない前記セグメントに格納された前記対象データに対するダーティデータの割合に応じて、前記パリティが生成されていない前記セグメントに格納された前記対象データを、前記第1のキャッシュメモリに転送し、又は、前記記憶デバイスに記憶する

20

請求項1乃至請求項8のいずれか一項に記載のストレージシステム。

【請求項10】

前記プロセッサは、前記パリティが生成されている前記セグメントに格納された前記対象データを、前記セグメントから前記記憶デバイスに格納する

請求項9に記載のストレージシステム。

【請求項11】

前記プロセッサは、前記パリティが生成されていない前記セグメントに格納された前記対象データを、前記記憶デバイスに記憶する時に、記憶する前記対象データに対しパリティを生成する

30

請求項9または請求項10に記載のストレージシステム。

【請求項12】

前記第2のキャッシュメモリの前記残りの領域は、前記記憶デバイスに既に格納されたクリーンデータを格納する複数のクリーンデータ領域を含み、

前記プロセッサは、各クリーンデータ領域に格納されたデータへのアクセスの頻度に応じて、各クリーンデータ領域を順位付けし、アクセスの頻度が低い前記クリーンデータ領域から順に前記対象データを格納する

請求項1に記載のストレージシステム。

40

【請求項13】

電断が起こった場合にバッテリーからの電力で動作するストレージシステムの制御方法であって、

ホストコンピュータからデータを受信する都度に、受信したデータを、一時的に格納する揮発性の第1のキャッシュメモリと、データを一時的に格納する不揮発性の第2のキャッシュメモリとのいずれかに格納し、

前記第1又は第2のキャッシュメモリに格納されているデータを記憶デバイスに格納し、

電断が起こった場合、前記第1のキャッシュメモリに格納されている全部または一部の前記データを対象データとして、前記第2のキャッシュメモリのダーティデータ領域と残

50

りの領域のうち前記残りの領域に格納し、

前記パーティデータ領域は、前記記憶デバイスに格納されていないデータであるパーティデータを格納する領域であり、

前記残りの領域は、前記第2のキャッシュメモリのうちの前記パーティデータ領域以外の領域である、

ストレージシステムの制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、概して、ストレージシステムのキャッシュ制御に関する。

10

【背景技術】

【0002】

電源遮断時の主記憶デバイス上のデータの消失を防ぐために、主記憶メモリ上のデータを退避させるための不揮発性メモリをストレージコントローラに搭載することがある。特許文献1は、退避すべき主記憶メモリ上のデータのみを選択し、不揮発性メモリに退避する技術を開示している。例えば、上位装置からのアクセスコマンドに従うデータの一時格納先は、揮発性キャッシュメモリとしている。一次電源から揮発性キャッシュメモリへの給電が無くなった場合に、バッテリーからの給電により、揮発性キャッシュメモリに記憶されているデータを不揮発性メモリにコピーしている。

【先行技術文献】

20

【特許文献】

【0003】

【特許文献1】特開2008-108026号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかし、特許文献1に開示された技術では、不揮発性メモリと揮発性キャッシュメモリの容量が考慮されていないので、揮発性キャッシュメモリが大容量である場合、揮発性キャッシュメモリの全てのデータをバックアップできない場合がある。また、主記憶キャッシュメモリの容量増加により電源遮断時に必要なバッテリー容量とバックアップ媒体容量が増大する。

30

【0005】

よって、ストレージシステムにおいて、電断時のデータ転送時間を短縮させることにより、バッテリー容量を削減し、コストを削減することが課題である。

【課題を解決するための手段】

【0006】

上記課題を解決するために、ストレージシステムは、記憶デバイスと、ホストコンピュータからデータを受信し、受信したデータを記憶デバイスに格納するストレージコントローラとを備える。ストレージコントローラは、プロセッサと、プロセッサに接続されデータを一時的に格納する揮発性の第1のキャッシュメモリと、マイクロプロセッサに接続されデータを一時的に格納する不揮発性の第2のキャッシュメモリと、電断が起こった場合に少なくともプロセッサおよび第1のキャッシュメモリに電力を供給するバッテリーとを備える。第2のキャッシュメモリは、記憶デバイスに格納されていないデータであるパーティデータを格納するパーティデータ領域と、パーティデータ領域以外の残りの領域とを含む。プロセッサは、電断が起こった場合、第1のキャッシュメモリに格納されている全部または一部のデータを対象データとして、第2のキャッシュメモリの残りの領域に格納する。

40

【発明の効果】

【0007】

電断時のデータ転送時間を短縮させることにより、バッテリー容量を削減し、コストを削

50

減することができる。

【図面の簡単な説明】

【0008】

【図1】第1実施例に係るストレージシステムを示す。

【図2】主記憶メモリと、NVRAMとに格納するデータの構成を示す。

【図3】ユーザーデータの構成を示す。

【図4】主記憶メモリ管理テーブルの構成を示す。

【図5】ダーティデータテーブルの構成を示す。

【図6】NVRAM管理テーブルの構成を示す。

【図7】退避先データテーブルの構成を示す。

【図8】第1実施例において、ストレージ装置が電断した時のストレージコントローラによるバックアップ処理の動作を示す。

【図9】ダーティ最大容量決定処理のフローチャートを示す。

【図10】バックアップ処理のフローチャートを示す。

【図11】第1実施例に係るバックアップデータ選択処理のフローチャートを示す。

【図12】第1実施例に係る退避先選択処理のフローチャートを示す。

【図13】第1実施例に係るデータ転送処理のフローチャートを示す。

【図14】第1実施例に係るバックアップデータの復旧処理のフローチャートを示す。

【図15】第2実施例に係るストレージシステムを示す。

【図16】第2実施例のストレージ装置が電断した時のストレージコントローラによるバックアップ処理の動作を示す。

【図17】第2実施例に係るバックアップデータの退避先判定処理のフローチャートを示す。

【図18】第2実施例に係るバックアップデータの復旧処理のフローチャートを示す。

【図19】第3実施例に係るバックアップデータの復旧処理の動作を示す。

【図20】第3実施例に係るバックアップデータの復旧処理のフローチャートを示す。

【図21】復旧対象データ選択処理のフローチャートを示す。

【図22】デステージデータテーブルの構成を示す。

【図23】デステージ処理のフローチャートを示す。

【図24】第3実施例に係るNVRAM管理テーブルの構成を示す。

【図25】第3実施例に係るバックアップデータ選択処理のフローチャートを示す。

【図26】第3実施例に係る退避先選択処理のフローチャートを示す。

【図27】クリーン領域選択処理のフローチャートを示す。

【図28】第3実施例に係るデータ転送処理のフローチャートを示す。

【図29】第3実施例に係る、バックアップデータの復旧処理のフローチャートを示す。

【発明を実施するための形態】

【0009】

以下、図面を参照しながら本発明の実施の形態を説明する。

図1は、本発明の第1実施例におけるストレージシステムの構成を示す。

【0010】

ストレージシステムは、ホストコンピュータ2と、ホストコンピュータ2に接続されたストレージ装置1とを有する。ストレージ装置1は、データ処理、演算を行うホストコンピュータ2に接続されている。ストレージ装置1は、ホストコンピュータ2とFibre Channel (FC)、およびiSCSI等によるデータ転送路で接続されている。なお、ホストコンピュータ2の数は一つでも複数でも良い。

【0011】

ストレージ装置1は、ストレージコントローラ3と、ドライブ筐体4と、電源5とを有する。ドライブ筐体5は、複数の記憶ドライブ6（本実施例では3つドライブ）を収容する。電源5は、ストレージコントローラ3とドライブ筐体4とに電力を供給する。また、ストレージ装置1は、商用電源7に接続されている。

10

20

30

40

50

【0012】

ストレージ装置1が含むストレージコントローラ3、ドライブ筐体4、電源5のそれぞれの数は設計に依存する。記憶ドライブの種別は問わない。例えば、ハードディスクドライブ(HDD)でも良いし、フラッシュメモリなどの不揮発性能メモリを搭載したSolid State Drive(SSD)であっても良い。

【0013】

記憶ドライブ6は、例えば不揮発性の記憶媒体であり、ホストコンピュータ2から転送されたユーザーデータを格納する。例えば、複数の記憶ドライブ6がRAID(Redundant Arrays of Inexpensive Disks)を構成しデータの冗長化を行っても良い。これにより、記憶ドライブ6に障害が発生しデータを消失したとしても、記憶ドライブ6のデータおよびパリティに基づいて消失したデータを復元することができる。なお、記憶ドライブ6は、記憶デバイスの一例である。

10

【0014】

ストレージコントローラ3は複数のクラスタ10を有し、本例において、2つのクラスタ10を有する。ストレージコントローラ3は、可用性を重視し、サービスを継続して提供するために2つのクラスタ2を備えている。なお、2つのクラスタ10は、同一の構成であるので、同一の符号を付し、一方のクラスタ10についてのみ説明する。

【0015】

クラスタ10は、ホストインターフェース(I/F)LSI11と、ドライブI/F LSI12と、マイクロプロセッサ(以下、単にMPとする)13と、主記憶メモリ14と、NVRAM15と、バッテリー16とを含む。

20

【0016】

ホストI/F LSI11は、ホストコンピュータ2との通信に用いられるデバイスであり、FC、Fibre Channel Over Ethernet(登録商標)(FCoE)、iSCSI等の通信プロトコルをストレージ装置1内部で用いられるPCIeプロトコルに変換する機能を有する。なお、ホストコンピュータ2からストレージ装置1に対し、512B単位でデータ転送が行われるが、ホストI/F LSI11はそのデータに対し8BのCRCを付加するように構成されている。

【0017】

ドライブI/F LSI12は、複数のドライブ6を収容するドライブ筐体4と、ストレージコントローラ3とを接続するためのデバイスである。例えば、ドライブI/F LSI12は、記憶ドライブ6とストレージコントローラ3との通信に用いられるFC、Serial Attached SCSI(SAS)、Serial Advanced Technology Attachment(SATA)、Non-Volatile Memory Express(NVMe)等の通信プロトコルをストレージコントローラ3内部で用いられるプロトコル、例えば、PCIeのプロトコルに変換する機能を有する。

30

【0018】

MP13は、クラスタ10内における内部処理を実行する。MP13は、演算やデータの読み書きを実行してストレージコントローラの制御を行うコア13A、主記憶メモリ14とのデータ転送を行うメモリI/F13B、ホストI/F LSI11とMP13とを接続するPCIeインターフェース13C、ドライブI/F LSI12とMP13とを接続するPCIe I/F13D、NVRAM15とMP13とを接続するPCIe I/F13E、主記憶メモリ14と他方のクラスタ2(リモートクラスタとも呼ぶ。)とのデータ転送を行うDMA13F、リモートクラスタ2との通信I/FとなるNon Transparent Bridge(NTB)13Gとを有する。MP13の内部構成要素は内部ネットワークで相互に接続されている。メモリI/F13Bは、例えばDDRインターフェースである。MP13のコア13Aは、ユーザーデータを主記憶メモリ14から記憶ドライブ6にコピーする際に後述のセグメント32毎にパリティを生成し、記憶ドライブ6に格納(デステージ)する。

40

50

【0019】

主記憶メモリ14は、MP13に接続されており、Double Data Rate (DDR) Synchronous DRAM等の半導体メモリで構成される揮発性の記憶デバイスである。主記憶メモリ14の容量は、例えば100~500GBである。主記憶メモリ14は、DDRインターフェースであるメモリI/F13Bを介してMP13と接続している。主記憶メモリ14は、ホストコンピュータ2から読み書きされるデータを一時的に記憶し(ユーザーデータのキャッシング)、また、MP13が扱うデータやストレージコントローラ10を制御するためのプログラムを格納する。主記憶メモリ14が格納するデータの詳細は後述する。なお、主記憶メモリ14は、第1のキャッシュメモリの一例である。

10

【0020】

NVRAM15は、典型的にはPCIe I/F13Eを介してマイクロプロセッサ13に接続されており、NAND型フラッシュメモリ、Phase Change RAM (PCM)、Resistance RAM (ReRAM)等の半導体メモリで構成される不揮発性の記憶デバイスである。NVRAM15の容量は、例えば500GB~1TBである。NVRAM15は、主記憶メモリ14に記憶されているデータを、NVRAM15に転送するためのDirect Memory Access (DMA)15Aを搭載する。NVRAM15はホストコンピュータ2から読み書きされるデータを一時的に格納する。一般に、NVRAM15は、DRAMと比較して面積あたりのデータ記録密度は大きい、データ読み込み速度、データ書き込み速度が小さく、また、データ書き換え可能回数は少ない。NVRAM15が格納するデータの詳細は後述する。なお、NVRAM15は、第2のキャッシュメモリの一例である。

20

【0021】

バッテリー16は、例えば、再充電可能な二次電池であり、電源5の故障等によりストレージコントローラ3に電力が供給出来なくなった場合に電源の代わりに電力を供給する。なお、バッテリー16を構成する電池の種別は問わない。例えば鉛蓄電池、リチウムイオン電池等であっても構わない。バッテリー16は、クラスタ10内の少なくともMP13、主記憶メモリ14、NVRAM15の3つを一定時間動作させられるだけの電荷量を保持する。バッテリー16は、電力をクラスタ10に供給する回路の電圧降下(電断とも呼ぶ)を契機にクラスタ10に電力の供給を開始するように構成されている。

30

【0022】

図2は、主記憶メモリ14と、NVRAM15とに格納するデータの構成例を示す。

【0023】

主記憶メモリ14には、ストレージコントローラ3を制御するプログラムである制御プログラムデータ20と、ホストコンピュータ2から読み書きされるユーザーデータ30と、ユーザーデータ30を管理する情報を格納する主記憶メモリ管理テーブル40と、NVRAM15に記憶されたユーザーデータを管理する情報を格納するNVRAM管理テーブル50と、主記憶メモリ14上のバックアップ対象となるデータを管理するバックアップデータテーブル60と、ユーザーデータ30が退避可能なNVRAM15上の領域を管理する退避先データテーブル70とが格納されている。また、NVRAM15には、ユーザーデータ80が格納されている。

40

【0024】

ユーザーデータ30、80、管理テーブル40、50、バックアップデータテーブル60、退避先データテーブル70の構造についての詳細は後述する。NVRAM15上にはユーザーデータ80のみが格納される。NVRAM15は、基本的にはキャッシュメモリとして用いられるが、ライト性能がDRAMである主記憶メモリ14よりも低いため、アクセス頻度が高い管理テーブル40、50、および制御プログラムデータ20は、NVRAM15に格納しない。

【0025】

ホストコンピュータ2からストレージ装置1に転送されたユーザーデータのうち、キャ

50

ッシュメモリに格納されており、記憶ドライブ6に既にコピーされているユーザーデータをクリーンデータと呼ぶ。すなわち、記憶ドライブ6にデステージされたユーザーデータがクリーンデータである。一方、ホストコンピュータ2からストレージ装置1に転送されたユーザーデータのうち、キャッシュメモリに格納されているが、記憶ドライブ6に未だコピーされていないユーザーデータをダーティデータと呼ぶ。すなわち、記憶ドライブ6にまだデステージされていないユーザーデータがダーティデータである。

【0026】

図3は、ユーザーデータ30の構造を示している。

【0027】

ユーザーデータ30は、ホストコンピュータ2から転送されたデータの集合体を表している。図3に示すように、一つのスロット31は、複数のセグメント32から構成されており、各セグメント32はそれぞれ複数のブロック33から構成される。なお、図示はしないが、NVRAM15に格納されたユーザーデータ80も、ユーザーデータ30と同じ構成である。

10

【0028】

一つのスロット31は、例えば256KBであり、スロット31の数は、設計に依存する。各セグメント32は主記憶メモリ14上の連続領域を表す。また、セグメント32の大きさは設計に依存し、スロット31が持てるセグメント32の数は設計に依存する。各ブロック33は、ユーザーデータ30を管理する最小の単位であって、例えば512Bであり、主記憶メモリ14上の連続領域を表す。

20

【0029】

そして、各ブロック33に、クリーンデータ33Aおよびダーティデータ33Bが格納される。また、MP13が、新たなダーティデータを格納するためにクリーンデータが格納されたブロック33またはセグメント32を初期化して解放する事により、一時的に空き領域のブロック33Cが生じる。

【0030】

また、ストレージ装置1においては、主記憶メモリ14に保持可能なダーティデータの最大容量を、ユーザーにより設定可能に構成されている。

【0031】

図4は、主記憶メモリ管理テーブル40の構成を示す。

30

【0032】

主記憶メモリ管理テーブル40は、主記憶メモリ14に格納されたユーザーデータ30を管理するテーブルである。主記憶メモリ管理テーブル40は、スロット管理ポインタテーブル41と、スロット管理テーブル42と、セグメント管理テーブル43とを有する。

【0033】

スロット管理ポインタテーブル41は、I/O先情報を格納するテーブルである。すなわち、スロット管理ポインタテーブル41は、ホストコンピュータ2から受けたI/O要求に含まれる情報で、I/O先を表す情報を格納する。典型的には、I/O先情報は、LUN(Logical Unit Number)とLBA(Logical Block Addressing)である。LUNは、論理ユニット(論理ボリューム)の番号であり、ホストコンピュータ2が認識している番号である。LBAは、記憶ドライブ6の論理ユニットにおける領域(論理ブロック)を表すアドレスです。

40

【0034】

スロット管理ポインタテーブル41は、所定数のブランクのエントリ(レコード)を有する。I/O要求を受けたときに、ブランクのエントリに、I/O先情報であるスロット管理ポインタが書かれる。そして、スロット管理テーブル42から、空きスロットに対応したレコードが特定されて、そのレコードに、スロット管理ポインタが紐付けられる。これにより、どのI/O先に書かれる(あるいはどのI/O先から読み出された)データが、キャッシュメモリである主記憶メモリ14のどこに入っているのかが分かる。

【0035】

50

なお、スロット管理ポインタテーブル41の1行は、最初からは、スロット管理テーブル42の1行に1対1に対応していない。なぜなら、そうしてしまうと、予め、論理ボリュームの容量分のポインタを用意しなければならず、テーブルのサイズもキャッシュのサイズも膨大に必要になってしまうからである。

【0036】

スロット管理テーブル42は、スロットと特定するための情報であるスロットID42Aと、Most Recent Used (MRU) 情報42Bと、Least Recent Used (LRU) 情報42Cと、スロットステータス42Dと、セグメント管理ポインタ42Eとを格納する。

【0037】

MRU情報42Bは、対応スロットに対するアクセスが高頻度であるか否かを示す情報である。LRU情報42Cは、対応スロットに対するアクセスが低頻度であるか否かを示す情報である。

【0038】

スロットステータス42Dは、クリーンまたはダーティのキャッシュ制御情報、およびスロット31の各セグメント32に対しパリティを生成したことを示す情報が格納する。なお、本実施例では、主記憶メモリ14およびNVRAM15に格納されているダーティデータの割合が、所定の割合を超えた場合に、ダーティデータに関するパリティがセグメント32毎に生成され、ダーティデータのデステージ処理が実行される。

【0039】

セグメント管理ポインタ42Eは、スロットが内包するセグメントに対応するセグメント管理テーブルのポインタを格納する。

【0040】

セグメント管理テーブル43は、各スロットに対応して複数設けられている。セグメント管理テーブル43は、セグメントを特定するための情報であるセグメントID43Aと、データ格納アドレス43Bと、クリーンデータビットマップ43Cと、ダーティデータビットマップ43Dとを格納する。データ格納アドレス43Bは、セグメントが配置されている位置を示す。クリーンデータビットマップ43Cは、対応セグメント内でクリーン状態になっているデータ格納位置を示す。すなわち、クリーンデータビットマップ43Cは、対応セグメント内でクリーンデータを格納している位置を示す。ダーティデータビットマップ43Dは、対応セグメント内でダーティ状態になっているデータ格納位置を示す。ダーティデータビットマップ43Dは、対応セグメント内でダーティデータを格納している位置を示す。

【0041】

図5は、バックアップデータテーブル60の構成を示す。

【0042】

バックアップデータテーブル60は、電断時にMP13のコア13Aにより、主記憶メモリ管理テーブル40に基づき作成される。バックアップデータテーブル60は、ユーザーデータ30中のバックアップ対象データに関するビットマップテーブルである。すなわち、バックアップデータテーブル60は、バックアップ対象となるデータがどこに格納されているのかを指し示すテーブルである。また、バックアップデータテーブル60は、スロット管理ポインタ61と、スロットID62と、セグメントID63と、データサイズ64と、データ先頭アドレス65とを格納する。なお、主記憶メモリ14からNVRAM15へ退避されるデータは、全ユーザーデータである場合と、ダーティデータのみである場合があるが、いずれのデータをバックアップ対象とした時も図5に示したバックアップデータテーブル60をバックアップデータテーブルとして説明する。なお、バックアップデータテーブル60は対象データテーブルに相当し、バックアップデータは対象データに相当する。また、電断時に主記憶メモリ14からNVRAM15へバックアップする対象であるためバックアップデータと表現しているが、電源が復旧したときにはNVRAM15から主記憶メモリ14へ復旧の対象となるデータでもある。

10

20

30

40

50

【 0 0 4 3 】

スロット管理ポインタ61と、スロットID62と、セグメントID63とは、それぞれ図4におけるスロット管理ポインタテーブル41と、スロットID42Aと、セグメントID43Aとに対応する。データサイズ64は、対応するセグメントに格納されているバックアップデータのサイズである。データ先頭アドレス65は、対応するセグメントに格納されているバックアップデータの先頭のアドレスを示す。

【 0 0 4 4 】

上記のように、ホストコンピュータ2から512B単位で転送されてくるデータに対し8BのCRCが付加される。そうすると、8Bずつアドレスがずれることとなる。このため、同じブロックでもダーティデータとクリーンデータが混在することとなる。したがって、バックアップ対象をデータサイズとアドレスで管理することにより、ユーザーデータ30からバックアップデータのみを抽出して退避させることができ、バックアップデータ量を減少させることができる。

10

【 0 0 4 5 】

図6は、NVRAM管理テーブル50の構成を示す。

【 0 0 4 6 】

NVRAM管理テーブル50は、NVRAM15に格納されたユーザーデータ80を管理するテーブルである。NVRAM管理テーブル50は、スロット管理ポインタテーブル51と、スロット管理テーブル52と、セグメント管理テーブル53とを格納する。スロット管理ポインタテーブル51と、スロット管理テーブル52と、セグメント管理テーブル53とは、それぞれ主記憶メモリ管理テーブル40のスロット管理ポインタテーブル41と、スロット管理テーブル42と、セグメント管理テーブル43と同様のテーブルであるので、詳細な説明は省略する。

20

【 0 0 4 7 】

なお、スロット管理ポインタテーブル51のスロット管理ポインタは、スロット管理ポインタテーブル41のスロット管理ポインタと同様に、I/O先情報を示すが、主記憶メモリ14に入ったデータのI/O先情報とは別のNVRAM15に入ってデータのI/O先情報を示している。また、スロット管理テーブル52は、スロットID52Aと、MRU情報52Bと、LRU情報52Cと、スロットステータス52Dと、セグメント管理ポインタ52Eとを有する。セグメント管理テーブル53は、各スロットに対応して複数設けられ、セグメントID53Aと、データ格納アドレス53Bと、クリーンデータビットマップ53Cと、ダーティデータビットマップ53Dとを有する。

30

【 0 0 4 8 】

図7は、退避先データテーブル70の構成を示す。

【 0 0 4 9 】

退避先データテーブル70は、電断時にMP13のコア13Aにより作成される。退避先データテーブル70は、電断時に、主記憶メモリ14からNVRAM15へユーザーデータ30中のダーティデータを退避させるための退避先(バックアップ先)を示すビットマップテーブルである。また、退避先データテーブル70は、スロット管理ポインタ71と、スロットID72と、セグメントID73と、データサイズ74と、データ先頭アドレス75とを有する。なお、主記憶メモリ14からNVRAM15へ退避されるデータは、全ユーザーデータである場合と、ダーティデータのみである場合があるが、いずれのデータをバックアップ対象とした時も図7に示した退避先データテーブル70を退避先データテーブルとして説明する。

40

【 0 0 5 0 】

スロット管理ポインタ71は、図5のスロット管理ポインタ61に対応する。すなわち、図5において管理しているバックアップデータのI/O先情報(スロット管理ポインタ)を、図7の退避先データテーブルにおいても管理している。これにより、主記憶メモリ14上のバックアップ対象となるデータと、NVRAM15上に退避されるデータとが対応付けられる。そして、主記憶メモリ14にバックアップデータを復旧したときに、図7

50

のテーブルのスロット管理ポインタ（I/O先情報）が表すI/O先に、バックアップデータを書込むことができる。

【0051】

また、スロットID72と、セグメントID73とは、それぞれ図5におけるスロットID62と、セグメントID63とに対応する。また、退避先データサイズ74は、対応するセグメントに格納されているダーティデータのサイズである。退避先データ先頭アドレス75は、対応するセグメントに格納されているダーティデータ先頭のアドレスを示す。

【0052】

図8は、第1実施例において、ストレージ装置1が電断した時のストレージコントローラ3によるバックアップ処理の動作を示す。なお、図8では、一方のクラスタ10におけるバックアップ処理の動作のみを示す。

10

【0053】

電源5の故障等により電圧が低下し、電源5からストレージコントローラ3へ電力を供給する回路8が開放状態に遷移した事を契機に、クラスタ10はバックアップ処理を開始する。バックアップ処理とは、主記憶メモリ14に格納されている、ユーザーデータ30、主記憶メモリ管理テーブル40、NVRAM管理テーブル50、バックアップデータテーブル60、および退避先テーブル70を、NVRAM15のクリーンデータ領域15Bおよび空き領域15Cにデータ転送F1を実行する処理を表す。すなわち、バックアップ処理では、バックアップデータが、ダーティデータ領域以外の残りの領域（クリーンデータ領域15Bおよび空き領域15C）に格納される。バックアップ処理の具体的な処理動作は後述する。なお、バックアップ処理時には、クラスタ10はバッテリー16により稼働する。

20

【0054】

図9は、ダーティ最大容量決定処理のフローチャートを示す。この処理は、ストレージ装置1の起動時、又は、主記憶メモリ14に保持可能なダーティデータの最大容量をユーザーが変更した時に実行される。

【0055】

MP13は、主記憶メモリ14の容量情報とNVRAM15の容量情報を取得し、NVRAM容量が主記憶メモリ容量より大きいかなかを判定する（S10）。NVRAM容量が主記憶メモリ容量より大きいと判定された場合（S10：YES）、MP13は、NVRAM容量と主記憶メモリ容量との差分を、NVRAM15で保持できるダーティデータの最大容量に設定する（S11）。主記憶メモリ14の方がライト性能は高いため、主記憶メモリ14にダーティデータを保持させるようにする事で、ストレージ装置1の性能を維持する事が可能となる。

30

【0056】

NVRAM容量が主記憶メモリ容量以下であると判定された場合（S10：NO）、MP13は、NVRAM容量が、主記憶メモリ14が保持できるダーティデータの最大容量よりも小さいかなかを判定する（S12）。NVRAM容量が、主記憶メモリ14が保持できるダーティデータの最大容量より小さいと判定された場合（S12：YES）、MP13は、主記憶メモリ14が保持できるダーティデータの最大容量をNVRAMの容量と同等になるようにストレージ装置1の設定を変更する（S13）。さらに、NVRAM15が保持できるダーティデータの容量を0にする（S13）。この場合、NVRAM15には、クリーンデータのみが格納されることとなる。これにより、主記憶メモリ14の方がライト性能は高いため、主記憶メモリ14にダーティデータを保持させるようにする事で、ストレージ装置1の性能を維持する事が可能となる。

40

【0057】

また、ステップS13の処理において、NVRAM15にダーティデータが格納されている場合には、そのダーティデータを全て記憶ドライブ6に書込んで（デステージして）、NVRAM15を全てクリーン状態にしても良い。また、ステップS13の処理におい

50

て、主記憶メモリ14に格納されているパーティデータの量が、主記憶メモリ14が保持できるパーティデータの最大容量を超えていた場合には、超えている分のパーティデータを記憶ドライブ6に書込んで良い。

【0058】

NVRAM15の容量が、主記憶メモリ14が保持できるパーティデータの最大容量以上であると判定された場合(S12:NO)、MP13は、NVRAM15と主記憶メモリ14が保持できるパーティデータの最大容量との差分を、NVRAM15で保持できるパーティデータの最大容量に設定する(S14)。これにより、主記憶メモリ14が保持できるパーティデータの最大容量分だけ、NVRAM15はクリーンデータを保持することになる。

10

【0059】

図10は、電断時におけるストレージ装置1のバックアップ処理のフローチャートを示す。

【0060】

まず、MP13は、電源5の故障等により電圧が低下し、電源5からストレージコントローラ3へ電力を供給する回路8が開放状態に遷移したか否かを判定する(S20)。すなわち、MP13は、電断が発生したか否かを判定する。電断が発生していなければ(S20:NO)、バックアップ処理は終了する。一方、電断が発生していれば(S20:YES)、MP13は、主記憶メモリ14上のバックアップ対象であるデータを選択するバックアップデータ選択処理を実行する(S21)。バックアップデータ選択処理の詳細については後述する。バックアップデータ選択処理が終了すると、MP13は、バックアップデータの退避先をNVRAM15から選択する退避先選択処理を実行する(S22)。退避先選択処理の詳細については後述する。退避先選択処理の終了後、MP13は、ステップS21で選択したバックアップデータをステップS22で選択した退避先に転送するデータ転送処理を実行する(S23)。データ転送処理の詳細については後述する。データ転送処理の完了をもってストレージ装置1のバックアップ処理を終了する。

20

【0061】

図11は、主記憶メモリ14上のバックアップデータを選択するバックアップデータ選択処理のフローチャートを示す。

【0062】

MP13は、バックアップデータテーブル60を作成する(S30)。この時点で、バックアップデータテーブル60は空となっている。次に、MP13は、主記憶メモリ14の容量がNVRAM15上のクリーンデータの容量と空き容量との合計量以下であるか否かを判定する(S31)。主記憶メモリ14の容量が合計量以下であると判定された場合(S31:YES)、MP13は、主記憶メモリ14上の全データをバックアップ対象として選択する(S32)。これにより、一律に全てのデータが選択されるので、逐一バックアップ対象となるデータを判定するための処理時間が削減される。よって、全体のバックアップ処理時間の削減が可能となる。

30

【0063】

MP13は、ステップS32で選択した全ユーザーデータの情報をバックアップデータテーブルに反映させる(S33)。具体的には、主記憶メモリ管理テーブル40に基づき、スロットポインタ、スロットID、セグメントID、データサイズ、データ先頭アドレスをバックアップデータテーブル60に反映させる。これにより、図5に示した全ユーザーデータに関するバックアップデータテーブル60が作成される。

40

【0064】

一方、主記憶メモリ容量が合計量より大きいと判定された場合(S31:NO)は、主記憶メモリ上のデータの全てを選択することができない。そのため、MP13は、主記憶メモリ管理テーブル40に基づき、パーティデータを格納するスロットとセグメントを特定し、パーティデータをバックアップ対象として選択する(S34)。この時、特定したパーティデータのサイズをパーティデータビットマップ43Dから計算する。次に、MP

50

13は、特定したパーティデータの情報に基づき、スロットポインタ、スロットID、セグメントID、パーティデータのデータサイズ、パーティデータのデータ先頭アドレスをバックアップデータテーブル60に反映させる(S35)。これにより、バックアップデータとして、図5に示したバックアップデータテーブル60が作成される。

【0065】

次に、MP13は、全てのユーザーデータについてパーティデータの判定処理が完了したか否かを判定する(S36)。MP13は、この判定処理が完了していなければステップS34に戻り(S36:NO)、パーティデータの選択処理を繰り返す。全てのユーザーデータについてパーティデータの判定処理が完了していると判定した場合(S36:YES)、MP13は、ステップS33およびステップS35で作成したバックアップデータテーブル60に基づき、バックアップ対象となるバックアップデータの合計サイズを計算する(S37)。ステップS37でバックアップデータの合計サイズを計算した後、MP13は、バックアップデータ選択処理を終了する。

10

【0066】

図12は、バックアップデータの退避先をNVRAM15から選択する退避先選択処理のフローチャートを示す。

【0067】

MP13は、NVRAM管理テーブル50に基づき、退避先データテーブルを作成する(S40)。なお、この時点において、退避先データテーブルには、データは何も反映されていない。MP13は、NVRAM管理テーブル50、および図11のバックアップデータ選択処理で作成したバックアップデータテーブル60を読み込む(S41)。

20

【0068】

ここで、NVRAM15の容量効率を高めるため、MP13はNVRAM15上にバックアップ用データ退避領域を設けずにバックアップデータ格納先としてクリーンデータが格納されているユーザーデータ領域を選択する。そのため、MP13はNVRAM15のユーザーデータの管理テーブルを読み込み、バックアップデータ格納先となるクリーンデータの位置を特定する必要がある。この時バックアップ格納先を選択する優先順位は以下の順となる。(1)ユーザーデータが格納されていない空き領域、(2)最も大きいデータ管理単位であるスロットの内、クリーン状態のスロット、(3)スロットの次に大きいデータ管理単位であるセグメントの内、クリーン状態のセグメント、(4)最も小さいデータ管理単位であるブロックの内、クリーン状態のブロック。このように、大きい領域から順に選択することにより、主記憶メモリ14からNVRAM15への書込み時間を短縮することができる。よって、バッテリー16によるクラスタ10への給電時間を短くことができ、バッテリー16の容量を削減することができる。ここで、クリーン状態とは、その対応する領域に格納されているデータが全てクリーンデータであり、パーティデータを格納していない状態を示す。

30

【0069】

MP13は、NVRAM管理テーブル50に基づき、NVRAM15に空き領域があるか否かを判定する(S42)。NVRAM15に空き領域がある場合(S42:YES)、MP13は、空き領域を選択する(S43)。一方、NVRAM15に空き領域がない場合(S42:NO)、MP13は、NVRAM15にクリーン状態のスロットがあるか否かを判定する(S44)。NVRAM15にクリーン状態のスロットがある場合(S44:YES)、MP13は、クリーン状態のスロットを選択する(S45)。NVRAM15にクリーン状態のスロットがない場合(S44:NO)、MP13は、NVRAM15にクリーン状態のセグメントがあるか否かを判定する(S46)。NVRAM15にクリーン状態のセグメントがある場合(S46:YES)、MP13は、クリーン状態のセグメントを選択する(S47)。NVRAM15にクリーン状態のセグメントがない場合(S46:NO)、MP13は、クリーン状態のブロックを選択する(S48)。このようにして、MP13は、退避領域を選択するので、NVRAM15中のクリーンデータなるべく残しておくことができる。

40

50

【 0 0 7 0 】

これらの退避先選択処理の後、MP 1 3 は、選択した領域の書込回数に 1 を足した回数が、最大書込可能回数を超えるか否かを判定する (S 4 9)。選択した領域の書込回数に 1 を足した回数が最大書込可能回数以上である場合 (S 4 9 : Y E S)、S 4 2 に戻り、退避先選択処理を繰り返す。一方、選択した領域の書込回数に 1 を足した回数が最大書込可能回数未満である場合 (S 4 9 : N O)、MP 1 3 は、N V R A M 管理テーブル 5 0 における選択した退避領域のキャッシュステートをダーティに遷移させる (S 5 0)。ダーティに遷移させることにより、リカバリ時に誤ってバックアップデータを破棄することを防止することができる。MP 1 3 は、選択した領域に対し、退避させるユーザーデータを割り当て、退避先データテーブル 7 0 に、割り当てたユーザーデータに関するスロットポ
10
インタ、スロット I D、セグメント I D、データサイズ、データ先頭アドレスを反映させる (S 5 1)。

【 0 0 7 1 】

次に、MP 1 3 は、図 1 1 の S 3 7 で算出したバックアップデータの合計サイズ分だけ退避領域を選択したか否かを判定する (S 5 2)。バックアップデータの合計サイズ分だけ退避領域を選択したと判定した場合 (S 5 2 : Y E S)、MP 1 3 は、退避先選択処理を終了する。一方、バックアップデータの合計サイズ分だけ退避領域を選択していないと判定した場合 (S 5 2 : N O)、MP 1 3 は、ステップ S 4 2 に戻る。

【 0 0 7 2 】

図 1 3 は、主記憶メモリ 1 4 から N V R A M 1 5 へバックアップデータを転送するデー
20
タ転送処理のフローチャートを示す。

【 0 0 7 3 】

MP 1 3 は、バックアップデータテーブル 6 0 と、退避先データテーブル 7 0 とを読み込む (S 6 0)。次に、MP 1 3 は、読み込んだ 2 つのテーブルに基づいて、N V R A M 1 5 の D M A 1 5 A でデータ転送するためのデータ転送命令を主記憶メモリ 1 4 に作成する (S 6 1)。この時、主記憶メモリ管理テーブル 4 0、N V R A M 管理テーブル 5 0、バックアップデータテーブル 6 0、および退避先データテーブル 7 0 も転送対象に含める。次に、MP 1 3 は、D M A 1 5 A に作成したデータ転送命令を読み込ませ、D M A 1 5 A にデータ転送を実行させる (S 6 2)。

【 0 0 7 4 】

MP 1 3 は、D M A 1 5 A にデータ転送を実行させた後、一定時間待機し (S 6 3)、D M A 1 5 A によるデータ転送が完了したか否かを判定する (S 6 4)。データ転送が完了していなければ (S 6 4 : N O)、MP 1 3 は、ステップ S 6 3 に戻り待機する。データ転送が完了していれば (S 6 4 : Y E S)、MP 1 3 は、全バックアップデータについて転送を完了したか否かを判定する (S 6 5)。全バックアップデータについて転送を完了していなければ (S 6 5 : N O)、MP 1 3 はステップ S 6 2 に戻り、転送処理を繰り返す。全バックアップデータについて転送を完了してしれば (S 6 5 : Y E S)、MP 1 3 はデータ転送処理を終了する。
30

【 0 0 7 5 】

図 1 4 は、電源が復旧した時の、MP 1 3 によるバックアップデータの復旧処理のフロー
40
チャートを示す。

【 0 0 7 6 】

MP 1 3 は、N V R A M 1 5 に復旧データが存在するか否かを判定する (S 7 0)。例えば、N V R A M 1 5 にバックアップデータテーブル 6 0 および退避先データテーブル 7 0 が格納されているか否かを判定する。N V R A M 1 5 に復旧データが存在しないと判定された場合 (S 7 0 : N O)、MP 1 3 は復旧処理を終了する。一方、N V R A M 1 5 に復旧データが存在すると判定された場合 (S 7 0 : Y E S)、MP 1 3 は、N V R A M 1 5 から退避先データテーブル 7 0 を読み込み、N V R A M 1 5 上のバックアップデータを特定する (S 7 1)。

【 0 0 7 7 】

次に、MP 13は、ステップS 72で特定したバックアップデータを主記憶メモリ14に再展開するために、NVRAM 15から主記憶メモリ14にステップS 72で特定したバックアップデータを転送する(S 72)。また、MP 13は、主記憶メモリ管理テーブル40およびNVRAM管理テーブル50を転送する。MP 13は、バックアップデータテーブル60を参照して、主記憶メモリ管理テーブル40を再構成し(S 73)、復旧処理を終了する。また、退避先データテーブル70を参照して、NVRAM管理テーブル50を再構成しても良い。バックアップデータの転送時には、バックアップデータに付されたCRCコードをチェックし、CRCエラーを検出した場合には、そのデータを破棄する。

【0078】

上記のように、電断時に主記憶メモリ14に格納されているデータをNVRAM 15の空き領域および/またはクリーンデータ領域に退避するようにしたので、NVRAM 15の容量効率を高めることができる。また、バックアップデータの退避用としての不揮発性メモリを設ける必要がなくなるので、コストの増大を防止することができる。また、キャッシュメモリとして使用するNVRAM 15にバックアップデータを退避させるので、データ転送時間を短縮することができ、バッテリー16の容量を削減することができ、コストを減少させることができる。

【0079】

次に、本発明の第2実施例におけるストレージ装置101について説明する。

【0080】

図15は、第2実施例におけるストレージシステムを構成するストレージ装置101を示している。なお、第1実施例と同一の構成については同一の番号を付して説明を省略し、異なる部分についてのみ説明する。

【0081】

図15に示すように、本実施例の各クラスタ10は、バックアップ専用の不揮発性媒体であるSSD 17と、SSD I/F LSI 18とを有している。SSD 17は、NAND型フラッシュメモリ、Phase Change RAM (PCM)、Resistance RAM (ReRAM)等の半導体メモリで構成される記憶デバイスであり、その容量は1~2TBである。SSD 17は、通常はSATAインターフェースでMP 13と接続している。また、SSD 17は、DRAMと比較して面積あたりのデータ記録密度は大きい、データ読み込み速度、データ書き込み速度が小さく、また、データ書き換え可能回数は少ない。

本実施例では、SSD 17は、電断時に主記憶メモリ14に格納されたデータを退避させるためだけに用いられ、バックアップ処理時以外でホストコンピュータ2から読み書きされるデータを格納することは無い。SSD I/F LSI 18は、SSD 17がMP 13と通信を行うためのコントローラである。

【0082】

図16は、本実施例のストレージ装置101が電断した時のストレージコントローラ3によるバックアップ処理の動作を示す。なお、図16では、一方のクラスタ10におけるバックアップ処理の動作のみを示す。

【0083】

電源5の故障等により電圧が低下し、電源5からストレージコントローラ3へ電力を供給する回路8が開放状態に遷移した事を契機に、クラスタ10はバックアップ処理を開始する。本実施例のバックアップ処理では、主記憶メモリ14に格納されている、ユーザーデータ30、主記憶メモリ管理テーブル40、NVRAM管理テーブル50、バックアップデータテーブル60、および退避先データテーブル70を、NVRAM 15のクリーンデータ領域および空き領域にデータ転送F1、SSD 17にデータ転送F2を実行する。このように、本実施例のバックアップ処理では、第1実施例のバックアップ処理と異なり、主記憶メモリ14からNVRAM 15へのバックアップデータ転送と、主記憶メモリ14からSSD 17へのバックアップデータ転送を並列して実行する。バックアップ処理の

10

20

30

40

50

具体的な処理動作は後述する。

【 0 0 8 4 】

本実施例においても、第 1 実施例の図 1 0 と同様のバックアップ処理が実行される。しかし、本実施例では、ステップ S 2 2 に代えて、図 1 7 のバックアップデータの退避先判定処理が実行される。

【 0 0 8 5 】

図 1 7 は、本実施例におけるバックアップデータの退避先判定処理のフローチャートを示す。

【 0 0 8 6 】

MP 1 3 は、クラスタ 1 0 に S S D 1 7 が存在するか否かを判定する (S 8 0)。S S D 1 7 が存在しないと判定すれば (S 8 0 : N O)、MP 1 3 は、図 1 0 のステップ S 2 2 と同じ処理を実行する。すなわち、バックアップデータの退避先が N V R A M 1 5 しかないので、図 1 2 の処理のみを実行する。

【 0 0 8 7 】

一方、S S D 1 7 が存在すると判定した場合 (S 8 0 : Y E S)、バックアップデータの格納先となる不揮発性媒体が二つあるため、MP 1 3 は、各媒体に格納するデータ量を設定する処理を実行する。まず、MP 1 3 は、N V R A M 1 5 および S S D 1 7 のデータ転送スループットを測定する (S 8 1)。例えば、MP 1 3 は、ダミーデータを作成し、主記憶メモリ 1 4 から N V R A M 1 5 と S S D 1 7 とに対しダミー転送を実行し、データ転送量とデータ転送時間とから 1 秒あたりのデータ転送量を計算する。これにより、N V R A M 1 5 のデータ転送スループット ($V_{N V R A M}$ [G B / S]) および S S D 1 7 のデータ転送スループット ($V_{S S D}$ [G B / S]) が測定される。

【 0 0 8 8 】

次に、MP 1 3 は、下記の式 (1)、(2) を用いて、N V R A M 1 5 と S S D 1 7 にバックアップするデータ量を算出する (S 8 2)。

【 数 1 】

$$B_{SSD} = C_{MaxDirty} \times \frac{V_{SSD}}{V_{SSD} + V_{NVRAM}} \dots \text{式 1}$$

10

20

30

【 数 2 】

$$B_{NVRAM} = C_{MaxDirty} \times \frac{V_{NVRAM}}{V_{SSD} + V_{NVRAM}} \dots \text{式 2}$$

ここで、各変数は

$V_{S S D}$: ステップ S 8 1 で測定した S S D のデータ転送スループット [G B / S]

40

$V_{N V R A M}$: ステップ S 8 1 で測定した N V R A M のデータ転送スループット [G B / S]

$C_{M a x D i r t y}$: 主記憶メモリに格納されているダーティデータ量 [G B]

$B_{S S D}$: S S D にバックアップするデータ量 [G B]

$B_{N V R A M}$: N V R A M にバックアップするデータ量 [G B]

である。

【 0 0 8 9 】

次に、MP 1 3 は、ステップ S 8 2 で算出した $B_{S S D}$ [G B] および $B_{N V R A M}$ [G B] に基づき、S S D 1 7 および N V R A M 1 5 のそれぞれにバックアップするデータを決定する。MP 1 3 は、ステップ S 8 3 で決定した N V R A M 1 5 に退避するバックア

50

ップデータについて、図10のステップS22と同じ処理を実行する。

【0090】

次に、MP13は、ステップS83で決定したSSD17に退避するバックアップデータについて、図示せぬ退避テーブルを作成する(S84)。退避テーブルでは、少なくともI/O先情報と、バックアップデータとが対応付けられる。これにより、主記憶メモリ14にバックアップデータを復旧した時に、I/O先情報が示すI/O先にバックアップデータを書込むことができる。MP13は、退避テーブル作成処理の後、退避先判定処理を終了する。このように、NVRAM15およびSSD17のデータ転送スループットに基づき、バックアップするデータ量を決定するので、NVRAM15およびSSD17のそれぞれに対するデータ転送を同時に終了させることができる。よって、バックアップデータを退避させる時間を最短にすることができ、バッテリー16の容量を削減することができる。

10

【0091】

図18は、本実施例における電源が復旧した時の、MP13によるバックアップデータの復旧処理のフローチャートを示す。なお、図14に示した復旧処理と同一の処理については同一の番号を付して説明を省略し、異なる処理についてのみ説明する。

【0092】

本実施例の復旧処理では、ステップS72とステップS73との間に、ステップ90とステップ91とが追加されている。MP13は、ステップ90において、SSD17にバックアップデータが格納されているか否かが判定する。SSD17にバックアップデータが格納されていなければ、MP13は、ステップS73に進む。一方、SSD17にバックアップデータが格納されていれば、SSD17に退避されたバックアップデータを主記憶メモリ14に転送する。

20

【0093】

次に、本発明の第3実施例におけるストレージ装置について説明する。本実施例では、第1実施例のストレージ装置と比較して、バックアップデータの復旧処理のみが異なる。よって、バックアップデータの復旧処理についてのみ説明を行う。この復旧処理では、NVRAM15上のバックアップデータのうち、主記憶メモリ14へ転送するデータと主記憶メモリ14へ転送しないデータとをそれぞれ選択し、主記憶メモリ14へ転送するデータのみを主記憶メモリ14へ転送する。これにより、バックアップデータの復旧処理の時間を削減している。なお、第1実施例と同一の構成については同一の番号を付して説明を省略し、異なる部分についてのみ説明する。

30

【0094】

図19は、本実施例における電源が復旧した時の、MP13によるバックアップデータの復旧処理の動作を示す。

【0095】

図19に示すように、本実施例では、NVRAM15に退避されたバックアップデータ19をセグメント毎に、主記憶メモリ14に戻すのか、記憶ドライブ6にデステージするのかが判定される。例えば、セグメントを、パリティが未生成であるパリティ未生成セグメント19Aと、パリティを生成済みであるパリティ生成済みセグメント19Bに分類する。さらに、パリティ未生成セグメント19Aに含まれるダークデータの割合が所定の閾値未満である第1のパリティ未生成セグメント19A1と、その割合が所定の閾値以上である第2のパリティ未生成セグメント19A2とに分類する。

40

【0096】

そして、バックアップデータの復旧処理において、第1のパリティ未生成セグメント19A1に含まれるダークデータを主記憶メモリ14に転送するデータ転送F3、パリティ生成済みセグメント19Bおよび第2のパリティ未生成セグメント19A2に含まれるダークデータを記憶ドライブ6に転送するデータ転送F4を実行する。これにより、主記憶メモリ14に転送する転送データ量の削減を図ることができ、復旧処理を短時間で終了させることができる。

50

【 0 0 9 7 】

図 2 0 は、本実施例における電源が復旧した時の、MP 1 3 によるバックアップデータの復旧処理のフローチャートを示す。なお、図 1 4 に示した復旧処理と同一の処理については同一の番号を付して説明を省略し、異なる処理についてのみ説明する。

【 0 0 9 8 】

S 7 0、S 7 1 の処理を実行後、MP 1 3 は、N V R A M 1 5 から主記憶メモリ管理テーブル 4 0 および N V R A M 管理テーブル 5 0 を読み込み、主記憶メモリ 1 4 上に構成する (S 1 0 0)。テーブルを構成後、MP 1 3 は、復旧対象データ選択処理を実行する (S 1 0 1)。復旧対象データ選択処理の詳細については後述する。復旧対象データ選択処理が終了すると、MP 1 3 は、復旧対象データ選択処理において主記憶メモリ 1 4 に復旧させるデータとして選択したバックアップデータ (ダーティデータ) を主記憶メモリ 1 4 に転送する (S 1 0 2)。バックアップデータの転送時には、バックアップデータに付された C R C コードをチェックし、C R C エラーを検出した場合には、そのデータを破棄する。その後、MP 1 3 は、デステージ処理を実行し (S 1 0 3)、バックアップデータの復旧処理を終了する。デステージ処理の詳細については後述する。

10

【 0 0 9 9 】

図 2 1 は、本実施例における復旧対象データ選択処理のフローチャートを示す。

【 0 1 0 0 】

まず、MP 1 3 は、図 2 2 に示すデステージデータテーブル 8 0 を主記憶メモリ 1 3 上に作成する (S 1 1 0)。なお、図 2 2 では、デステージデータテーブル 8 0 にデータが入力されているが、作成時点ではデステージデータテーブル 8 0 は空となっている。

20

【 0 1 0 1 】

次に、MP 1 3 は、主記憶メモリ管理テーブル 4 0 を読み込む (S 1 1 1)。この処理により、電断前の主記憶メモリ 1 4 のキャッシュ管理の状態を確認することができる。MP 1 3 は、セグメント管理テーブル 4 3 からダーティデータを含む 1 つのセグメントを選択する (S 1 1 2)。選択したセグメントについて、パリティ生成済みであるか否かを判定する (S 1 1 3)。パリティ生成済みであるか否かは、スロット管理テーブル 4 2 のスロットステータス 4 2 D を参照して判定する。選択したセグメントがパリティ生成済みであった場合 (S 1 1 3 : Y E S)、MP 1 3 は、選択したセグメントに含まれるダーティデータを主記憶メモリ 1 4 に転送しないデータとして選択する (S 1 1 4)。すなわち、N V R A M 1 5 に格納されたバックアップデータのうち、選択したセグメントに含まれるダーティデータに対応するダーティデータを主記憶メモリ 1 4 に転送しないデータとして選択する。次に、MP 1 3 は、S 1 1 4 において選択したダーティデータを図 2 2 のデステージデータテーブルに追加する (S 1 1 5)。

30

【 0 1 0 2 】

一方、選択したセグメントがパリティ未生成である場合 (S 1 1 3 : N O)、MP 1 3 は、選択したセグメントに含まれるデータに対するダーティデータの割合が所定の閾値以上であるか否かを判定する (S 1 1 6)。選択したセグメントに含まれるデータに対するダーティデータの割合は、セグメント管理テーブル 4 3 (図 4) を参照して算出される。また、所定の閾値は、ユーザーが自由に設定可能である。なお、バックアップ処理において、ダーティデータのみを退避させた場合には、MP 1 3 は、選択したセグメントにおける空き領域を含む全体のサイズに対するダーティデータの割合が所定の閾値以上であるか否かを判定する。ダーティデータの割合が所定の閾値以上である場合 (S 1 1 6 : Y E S)、MP 1 3 は、S 1 1 4 へ進む。

40

【 0 1 0 3 】

一方、ダーティデータの割合が所定の閾値未満である場合 (S 1 1 6 : N O)、MP 1 3 は、選択したセグメントに含まれるダーティデータを主記憶メモリ 1 4 に転送するデータとして選択し、データ転送命令を作成する (S 1 1 7)。ダーティデータの割合が所定の閾値より小さいセグメントに含まれるダーティデータの量は比較的少ないと考えられ、データ復旧時の主記憶メモリ 1 3 へのデータ転送量を少なくすることができる。よって、

50

復旧処理を短時間で終了させることができる。なお、S 1 1 7ではダーティデータのみを選択したが、クリーンデータも選択して復旧させても良い。

【 0 1 0 4 】

次に、MP 1 3は、全てのダーティデータを含むセグメントについて処理を完了したか否かを判定し、完了していればS 1 1 9に進み（S 1 1 8：YES）、完了していなければS 1 1 2に戻る（S 1 1 8：NO）。MP 1 3は、S 1 1 7およびS 1 1 4で選択したデータに基づき、主記憶メモリ管理テーブル40とNVRAM管理テーブル50とを更新する。すなわち、主記憶メモリ管理テーブル40およびNVRAM管理テーブル50から、記憶ドライブ6にデステージするダーティデータに関する情報を削除し、新たなダーティデータを格納することを可能にする。これにより、主記憶メモリ14のキャッシュ利用効率を高めることができる。また、主記憶メモリ管理テーブル40およびNVRAM管理テーブル50を更新することにより、バックアップデータの復旧処理後の主記憶メモリ14およびNVRAM15のユーザーデータの状態と、管理テーブル40、50との整合性をとることができる。

10

【 0 1 0 5 】

図22は、デステージ処理の対象であるダーティデータを管理するデステージデータテーブル80の構成を示す。

【 0 1 0 6 】

デステージデータテーブル80は、スロット管理ポインタ81と、スロットID82と、セグメントID83と、ダーティデータサイズ84と、ダーティデータ先頭アドレス85とを格納する。スロットID82、セグメントID83、ダーティデータサイズ84、およびダーティデータ先頭アドレス85は、NVRAM15における情報である。

20

【 0 1 0 7 】

図23は、本実施例におけるデステージ処理のフローチャートを示す。

【 0 1 0 8 】

まず、MP 1 3は、主記憶メモリ13からデステージデータテーブル80を読み込む（S 1 2 0）。次に、MP 1 3は、デステージ対象であるダーティデータのうち、パリティ未生成のダーティデータに対しパリティを生成し（S 1 2 1）、デステージ対象であるダーティデータとパリティを記憶ドライブ6にデステージ処理する命令を作成する（S 1 2 2）。なお、パリティの生成は、NVRAM15のDMA15Aが行っても良い。

30

【 0 1 0 9 】

MP 1 3は、ドライブI/F LSI12に作成した命令を読み込ませ、デステージ対象であるダーティデータとパリティを記憶ドライブ6に転送する（S 1 2 3）。転送後、MP 1 3は、デステージデータテーブル80の全デステージ対象データについて、デステージ処理を完了したか否かを判定し（S 1 2 4）、完了していない場合はS 1 2 0に戻り、完了していればデステージ処理を終了する。

【 0 1 1 0 】

次に、本発明の第4実施例におけるストレージ装置について説明する。本実施例では、NVRAM15上の退避先を選択する際に、利用頻度の低いクリーンスロットを優先的に選択し、選択したクリーンスロットにバックアップデータを退避させるようにしている。これにより、利用頻度の高いクリーンスロットを残し、電源復旧後の装置のキャッシュヒット率を向上させることができる。なお、第1実施例と同一の構成については同一の番号を付して説明を省略し、異なる部分についてのみ説明する

40

【 0 1 1 1 】

図24は、本実施例におけるNVRAM管理テーブル150の構成を示す。NVRAM管理テーブル150は、MRU情報テーブル54およびLRU情報テーブル55を格納している。なお、図示していないが、NVRAM管理テーブル150は、図6で示したスロット管理ポインタテーブル51、スロット管理テーブル52、およびセグメント管理テーブル53も格納している。

【 0 1 1 2 】

50

M R U 情報テーブル 5 4 は、どのスロットにより直近にアクセスがあったかを示すテーブルである。すなわち、M R U 情報テーブル 5 4 では、直近にアクセスのあったスロットがより上位に配置されるように構成されている。換言すれば、直近にアクセスのないスロットほど下位に配置される。M R U 情報テーブル 5 4 は、アクセス履歴情報 5 4 A と、スロット位置ポインタ 5 4 B とを有する。アクセス履歴情報 5 4 A は、その数値が小さいほど、より直近にスロットに対しアクセスがあったこと示す情報である。スロット位置ポインタ 5 4 B は、アクセスのあったスロットの位置を示す情報である。したがって、M P 1 3 は、あるスロットにアクセスがあった場合、そのスロットを M R U 情報テーブル 5 4 の最も上位に配置する。

【 0 1 1 3 】

10

L R U 情報テーブル 5 5 は、どのスロットにより直近にアクセスがなかったかを示すテーブルである。すなわち、L R U 情報テーブル 5 5 では、直近にアクセスがなかったスロットがより上位に配置されるように構成されている。換言すれば、直近にアクセスのあったスロットほど下位に配置される。L R U 情報テーブル 5 5 は、アクセス履歴情報 5 5 A と、スロット位置ポインタ 5 5 B とを有する。アクセス履歴情報 5 5 A は、その数値が小さいほど、より直近にスロットに対しアクセスがなかったこと示す情報である。スロット位置ポインタ 5 5 B は、アクセスのなかったスロットの位置を示す情報である。また、N V R A M 1 5 の全スロットは、M R U 情報テーブル 5 4 および L R U 情報テーブル 5 5 のいずれかに格納される。

【 0 1 1 4 】

20

なお、M R U 情報テーブル 5 4 は、直近にアクセスのあったスロットがより上位に配置されるように構成したが、所定の期間内におけるスロットへのアクセスの回数を記録しておき、回数が多いスロットほど上位に配置するようにしても良い。同様に、L R U 情報テーブル 5 5、直近にアクセスのなかったスロットほど、より上位に配置されるように構成したが、所定の期間内におけるスロットへのアクセスの回数を記録しておき、回数が少ないスロットほど上位に配置するようにしても良い。

【 0 1 1 5 】

図 2 5 は、本実施例における主記憶メモリ 1 4 上のバックアップデータを選択するバックアップデータ選択処理のフローチャートを示す。なお、図 1 1 に示したバックアップデータ選択処理と同一の処理については同一の番号を付して説明を省略し、異なる処理についてのみ説明する。

30

【 0 1 1 6 】

バックアップデータテーブル 6 0 を作成後、M P 1 3 は、主記憶管理テーブル 4 0 と N V R A M 管理テーブル 1 5 0 とを読み込む (S 1 3 0)。これにより、M P 1 3 は、ストレージ装置 1 が有するクリーンデータの量とダーティデータの量を把握することができる。

【 0 1 1 7 】

主記憶メモリ容量が合計量より大きいと判定された場合 (S 3 1 : N O)、M P 1 3 は、主記憶メモリ 1 4 のダーティデータを含むセグメント数が、N V R A M 1 5 のクリーン状態のセグメント数以下であるか否かを判定する (S 1 3 1)。主記憶メモリ 1 4 のダーティデータを含むセグメント数が、N V R A M 1 5 のクリーン状態のセグメント数以下であった場合 (S 1 3 1 : Y E S)、M P 1 3 は、ダーティデータを含むセグメントをバックアップデータとして選択する (S 1 3 2)。これにより、M P 1 3、セグメント中のダーティデータのみを選択・抽出して、N V R A M 1 5 に転送しなくても、セグメントをコピーする処理のみでダーティデータを全て退避させることができる。M P 1 3 は、N V R A M 1 5 上のクリーン状態のセグメントを開放し、主記憶メモリ 1 4 のダーティデータを含むセグメントをコピーする処理を行うことにより、ダーティデータをバックアップすることができる。M P 1 3 は、S 3 0 で作成したバックアップデータテーブル 6 0 を破棄する (S 1 3 3)。すなわち、セグメント単位でデータをバックアップするので、バックアップデータの先頭アドレス等の情報が必要なくなるため、バックアップデータテーブル 6

40

50

0を破棄する。

【0118】

図26は、本実施例におけるバックアップデータの退避先をNVRAM15から選択する退避先選択処理のフローチャートを示す。なお、図12に示した退避先選択処理と同一の処理については同一の番号を付して説明を省略し、異なる処理についてのみ説明する。

【0119】

まず、MP13は、主記憶メモリ14上にバックアップデータテーブル60が存在するか否かを判定する(S140)。バックアップデータテーブル60が存在しない場合(S140:NO)、すなわち、図26のS133でバックアップデータテーブル60を破棄している場合、MP13は、退避先リストを作成する(S141)。その後、MP13は、NVRAM管理テーブル150を読み込み(S142)、退避先を選択する処理を行う(S143)。S143のクリーン領域選択処理の詳細については後述する。

10

【0120】

一方、バックアップデータテーブル60が存在する場合(S140:YES)、MP13は、図12で説明したS40~S42の処理を実行し、退避先を選択する処理を行う(S43、S141)。

【0121】

MP13は、選択した領域の書込回数に1を足した回数が最大書込可能回数以上である場合(S49:YES)において、S49の処理に至るルートが、S40~S42の処理を実行するルートであった場合はS42の処理に戻り、S140~S141の処理を実行するルートであった場合はS143の処理に戻る。

20

【0122】

また、MP13は、S51の処理において、S51の処理に至るルートが、S40~S42の処理を実行するルートであった場合は、退避先データテーブル70に、割り当てたユーザーデータに関する情報を反映させ、S140~S141の処理を実行するルートであった場合は、退避先リストにS143で選択したスロットおよびセグメントに関する情報を反映させる。

【0123】

図27は、クリーン領域選択処理のフローチャートを示す。

【0124】

まず、MP13は、S41の処理又はS141の処理で読み込んだNVRAM管理テーブル150のLRU情報テーブル55に、退避先として未選択のクリーンスロットがあるか否かを判定する(S150)。LRU情報テーブル55に未選択のクリーンスロットがある場合(S150:YES)、最上位のクリーンスロットを選択し(S151)、クリーン領域選択処理を終了する。

30

【0125】

一方、LRU情報テーブル55に未選択のクリーンスロットがない場合(S150:NO)、MP13は、MRU情報テーブル54に、退避先として未選択のクリーンスロットがあるか否かを判定する(S152)。MRU情報テーブル54に未選択のクリーンスロットがある場合(S150:YES)、最下位のクリーンスロットを選択し(S151)、クリーン領域選択処理を終了する。このように、クリーンスロットを選択することにより、利用頻度の高いクリーンスロットを残すことができるので、電源復旧後におけるキャッシュヒット率を向上させることができる。

40

【0126】

MRU情報テーブル54に未選択のクリーンスロットがない場合(S152:NO)、MP13は、S154の処理へ進む。S154~S156の処理は、図12のS46~S48と同じであるので説明を省略する。

【0127】

図28は、本実施例における主記憶メモリ14からNVRAM15へバックアップデータを転送するデータ転送処理のフローチャートを示す。なお、図13に示した退避先選択

50

処理と同一の処理については同一の番号を付して説明を省略し、異なる処理についてのみ説明する。

【0128】

まず、MP13は、主記憶メモリ14にバックアップデータテーブル60が存在するかどうかを判定する(S160)。主記憶メモリ14にバックアップデータテーブル60が存在しない場合(S160:NO)、すなわち、図25のS133の処理でバックアップデータテーブル60を破棄している場合、MP13は、主記憶メモリ管理テーブル40と退避先リストとを読み込む。そして、S61の処理において、退避先リストのセグメントに主記憶メモリ管理テーブル40におけるパーティータを含むセグメントをコピーする命令を作成する。なお、主記憶メモリ管理テーブル40、NVRAM管理テーブル50、退避先リストを転送対象に含める。

10

【0129】

一方、主記憶メモリ14にバックアップデータテーブル60が存在する場合(S160:YES)、MP13は、S60の処理を行う。

【0130】

図29は、本実施例における、電源が復旧した時の、MP13によるバックアップデータの復旧処理のフローチャートを示す。なお、図13に示した退避先選択処理と同一の処理については同一の番号を付して説明を省略し、異なる処理についてのみ説明する。

【0131】

MP13は、S70において、NVRAM15に、バックアップデータテーブル60、退避先データテーブル70、又は退避先リストが格納されているかどうかを判定する。NVRAM15に復旧データが存在すると判定された場合(S70:YES)、MP13は、NVRAM15に退避先テーブルが存在するかどうかを判定する(S170)。NVRAM15に退避先テーブルが存在する場合、MP13は、S71~S73の処理を実行する。すなわち、NVRAM15に退避先テーブルが存在する場合は、全ユーザーデータまたはパーティータを退避させた場合である。

20

【0132】

一方、NVRAM15に退避先テーブルが存在しない場合、MP13は、主記憶メモリ管理テーブル40およびNVRAM管理テーブル50を主記憶メモリ14に転送し、それらの管理テーブル49、50を主記憶メモリ14上に再構成する(S171)。すなわち、データ転送処理においてNVRAM15に退避させたデータを、復旧処理において主記憶メモリ14に戻さない。これにより、復旧処理にかかる時間を短縮することができ、ホストコンピュータ2は、電源復旧後直ちにNVRAM15のユーザーデータにアクセスすることができる。

30

【0133】

なお、上記実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに限定されるものではない。ある実施例の構成の一部を、他の実施例の構成に置き換えることができる。ある実施例の構成に他の実施例の構成を加えることもできる。各実施例の構成の一部について、他の構成の追加、削除、置換が可能である。例えば、上記の各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等により、ハードウェアで実現してもよい。あるいは、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行することによりソフトウェアで実現してもよい。

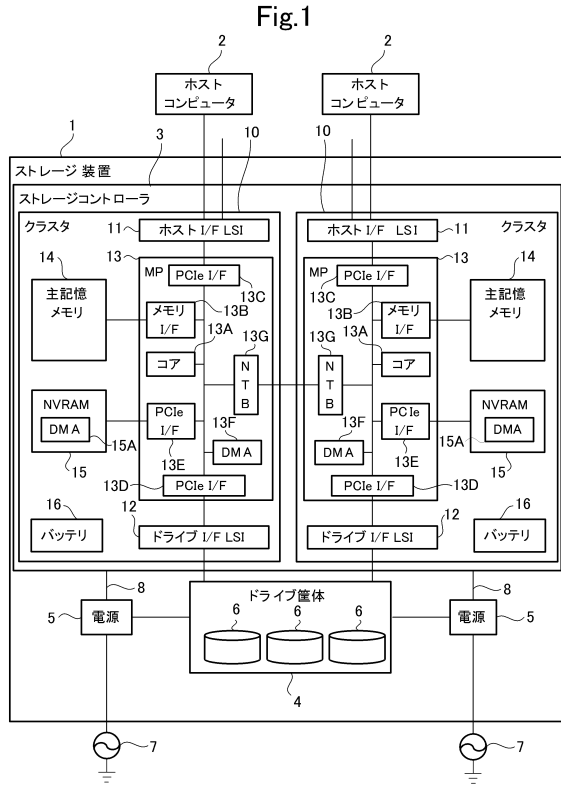
40

【符号の説明】

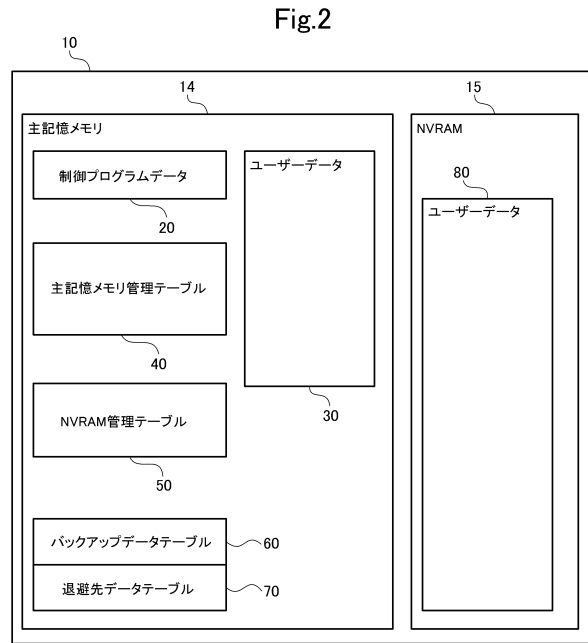
【0134】

1：ストレージ装置、2：ホストコンピュータ、6：記憶ドライブ、4：ドライブ筐体、3：ストレージコントローラ、14：主記憶メモリ、15：NVRAM、16：バッテリー、60：バックアップデータテーブル、70：退避先データテーブル

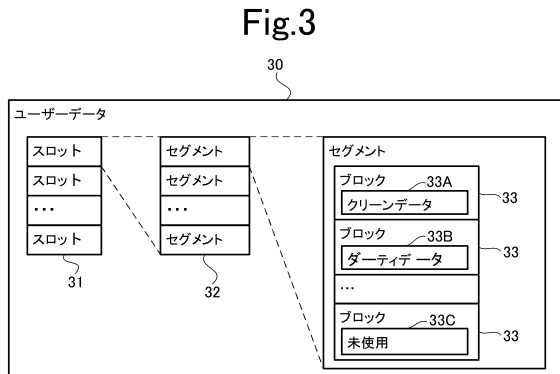
【図1】



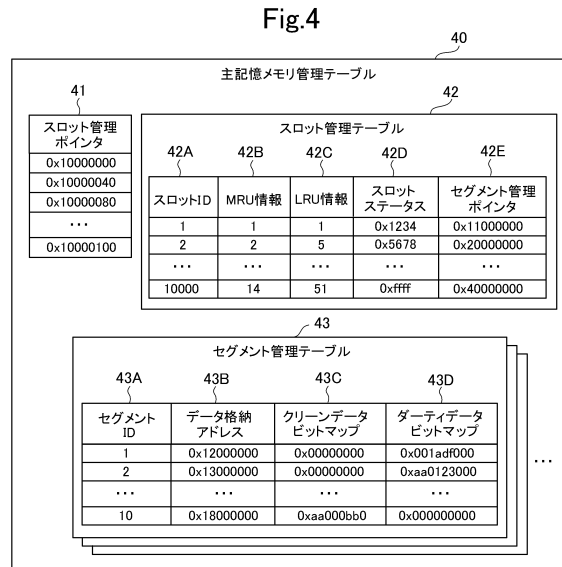
【図2】



【図3】



【図4】



【図5】

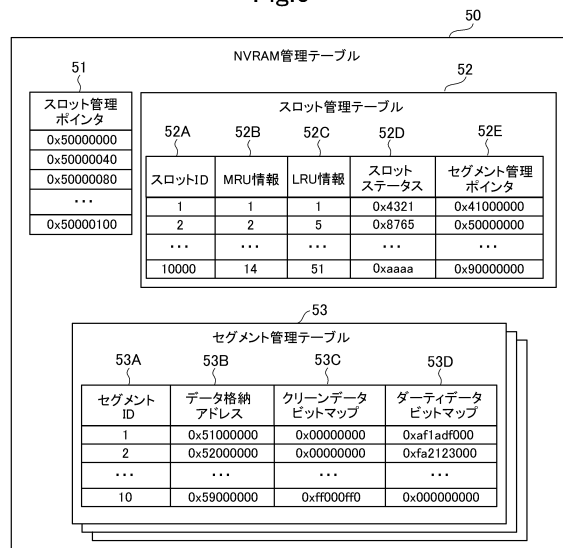
Fig.5

バックアップデータテーブル

61	62	63	64	65
スロット管理 ポインタ	スロット ID	セグメント ID	データサイズ	データ 先頭アドレス
0x10000040	2	2	0xffff	0x10a0110aa
0x10000000	1	4	0x1000	0x10a000000
...
0x10020000	14	6	0x80000	0x20a0110aa

【図6】

Fig.6



【図7】

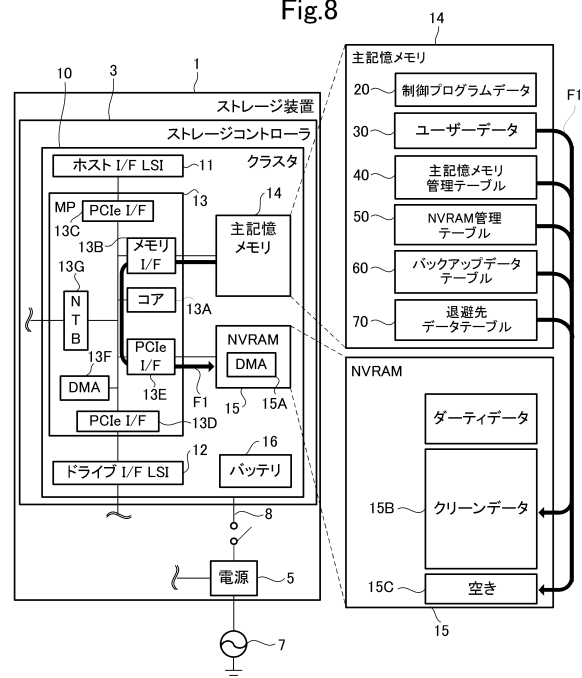
Fig.7

退避先データテーブル

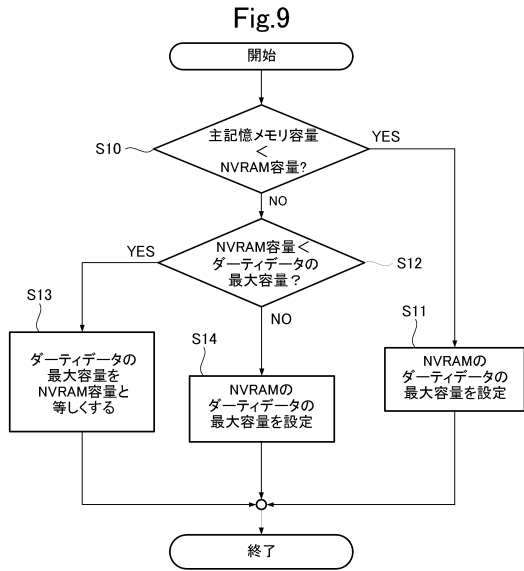
71	72	73	74	75
スロット管理 ポインタ	スロット ID	セグメント ID	データサイズ	データ 先頭アドレス
0x10000040	3	1	0xffff	0x80a0110aa
0x10000000	2	3	0x1000	0x80a000000
...
0x10020000	16	8	0x80000	0x90a0110aa

【図8】

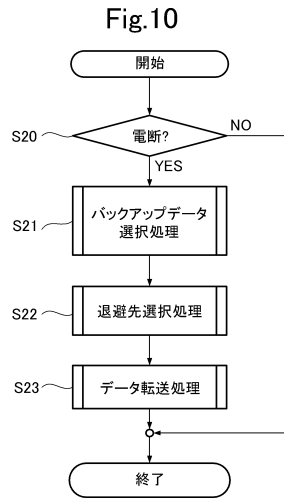
Fig.8



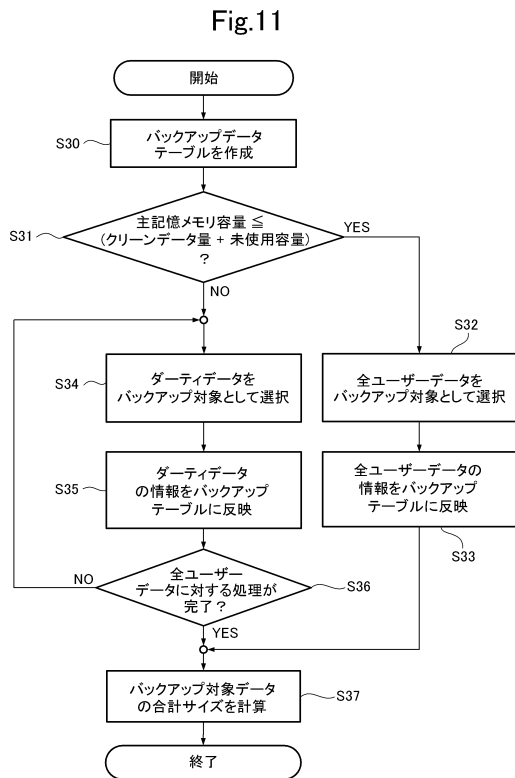
【図9】



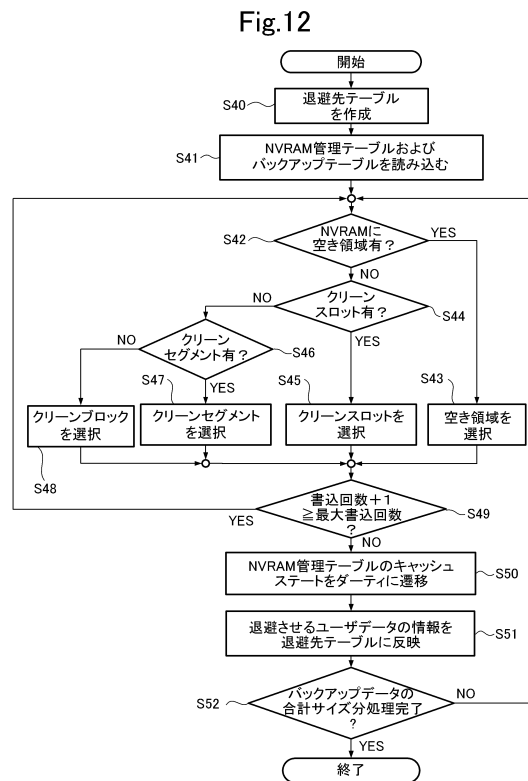
【図10】



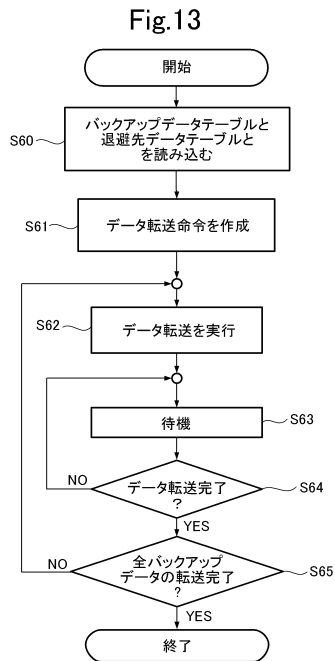
【図11】



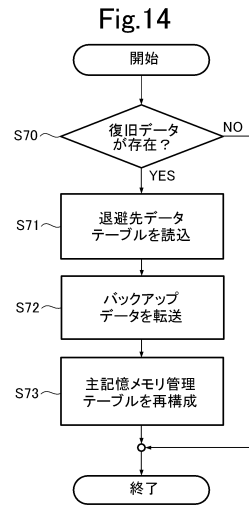
【図12】



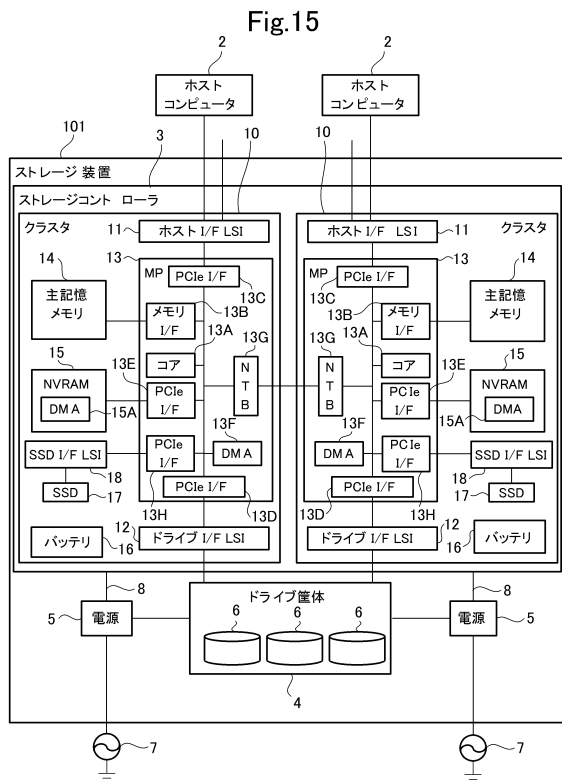
【図13】



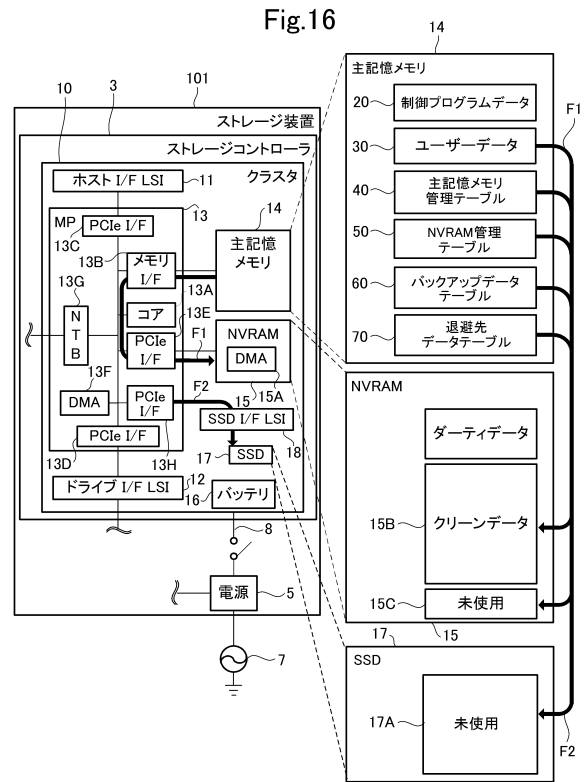
【図14】



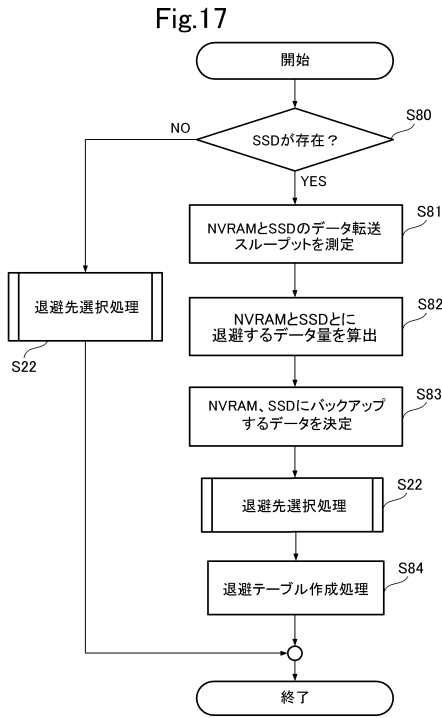
【図15】



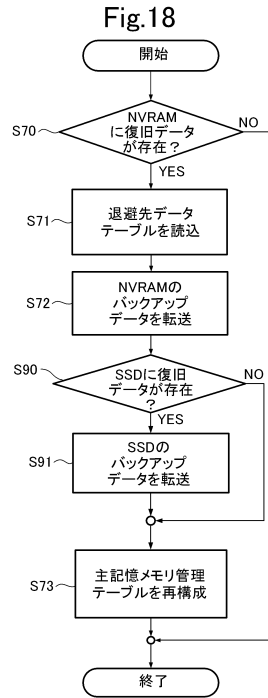
【図16】



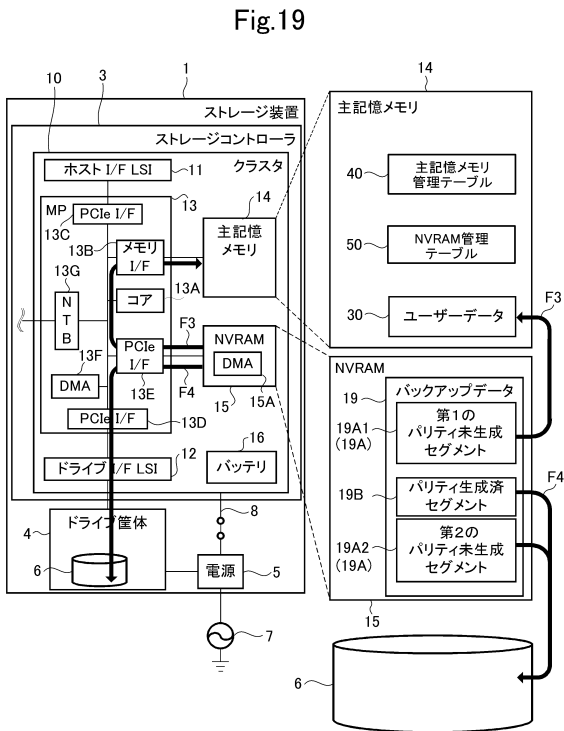
【図17】



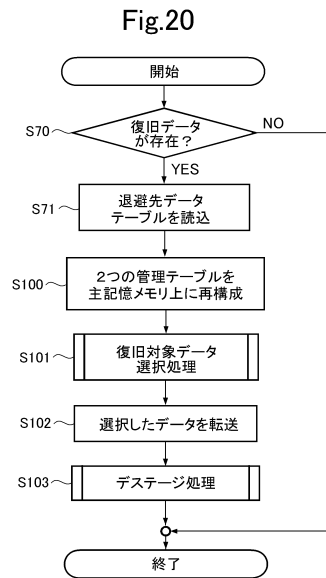
【図18】



【図19】

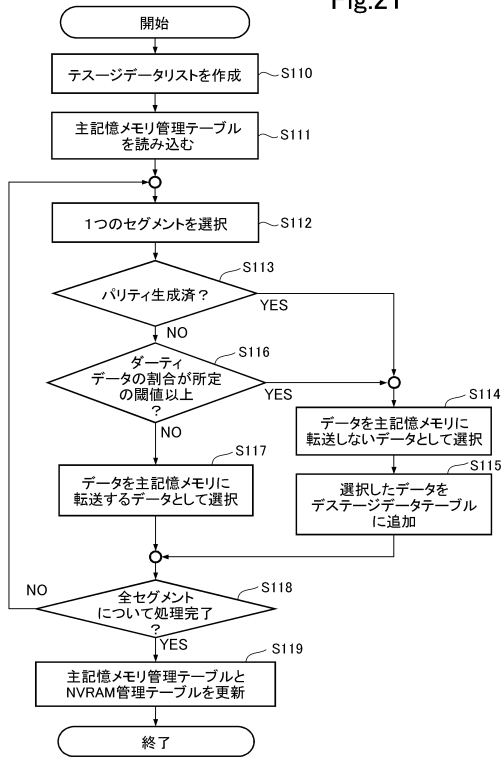


【図20】



【図 2 1】

Fig.21



【図 2 2】

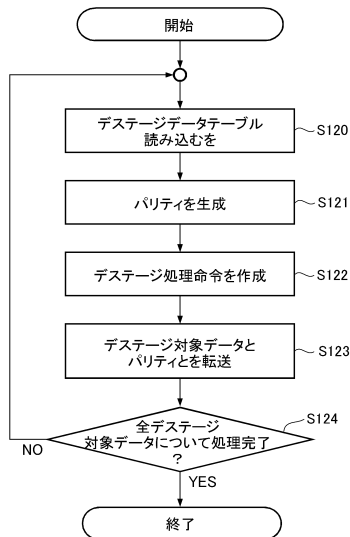
Fig.22

デステージデータテーブル

81	82	83	84	85
スロット管理 ポインタ	スロット ID	セグメント ID	データ データサイズ	データデータ 先頭アドレス
0x124a5440	2	2	0x400	0xb0a0110aa
0x3fe4a9000	1341	2	0x123	0xb0a00000
...
0x48663adc	2	44	0xffff	0xb0a0110bb

【図 2 3】

Fig.23



【図 2 4】

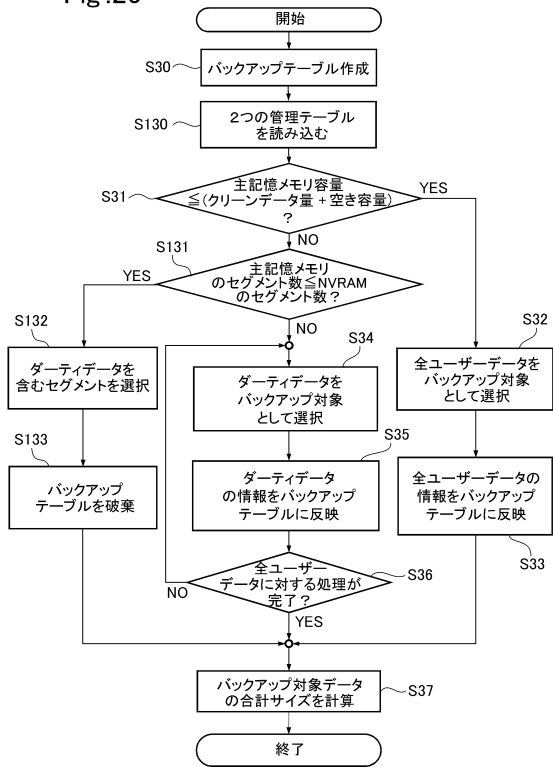
Fig.24

NVRAM管理テーブル

54		55	
MRU情報テーブル		LRU情報テーブル	
54A	54B	55A	55B
アクセス 履歴情報	スロット位置 ポインタ	アクセス 履歴情報	スロット位置 ポインタ
1	0xc00138888	1	0xd0131111
2	0x12375681	2	0xcc12ff12
...
0xffff	0xade9924	0xffff	0x508796111

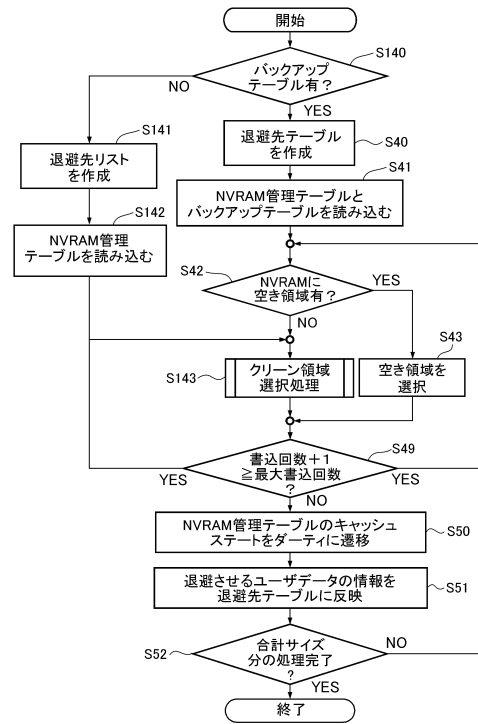
【図25】

Fig.25



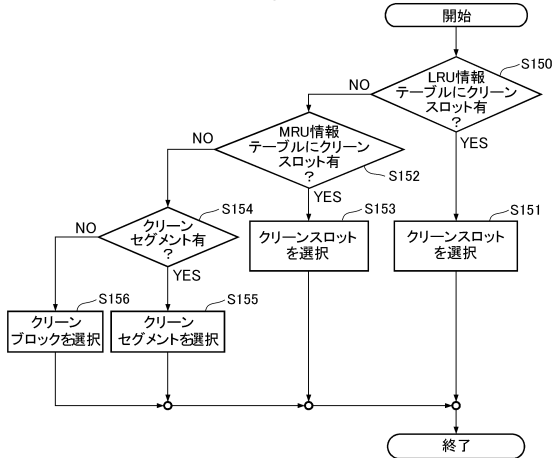
【図26】

Fig.26



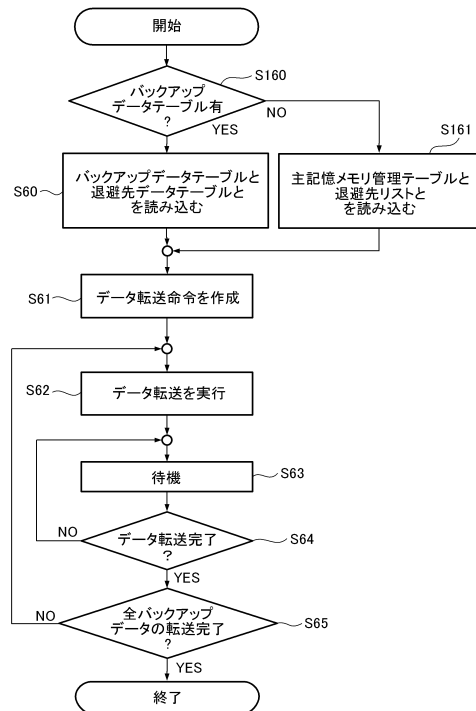
【図27】

Fig.27



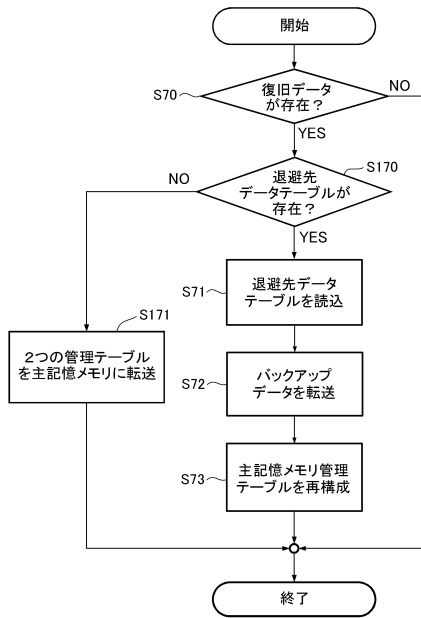
【図28】

Fig.28



【図29】

Fig.29



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 12/16 3 4 0 Q
G 0 6 F 12/08 5 5 3 B

(72)発明者 工藤 晋太郎
東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
(72)発明者 水野 真喜夫
東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

審査官 滝谷 亮一

(56)参考文献 特開2009-075759(JP,A)
特開2008-108026(JP,A)

(58)調査した分野(Int.Cl., DB名)
G 0 6 F 1 2 / 0 8 6 8
G 0 6 F 1 2 / 0 8
G 0 6 F 3 / 0 6
G 0 6 F 1 2 / 1 6