(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4363676号
(P4363676)
(45)発行日　平成21年11月11日(2009.11.11)　(24)登録日　平成21年8月28日(2009.8.28)

(51)Int.Cl.　　F I
G06F 3/06 (2006.01)　　G06F 3/06 304F
G06F 11/14 (2006.01)　　G06F 11/14 310Z
G06F 12/00 (2006.01)　　G06F 12/00 531J

請求項の数 7　(全 15 頁)

(21)出願番号　特願平9-300916
(22)出願日　平成9年10月31日(1997.10.31)
(65)公開番号　特開平11-134117
(43)公開日　平成11年5月21日(1999.5.21)
審査請求日　平成16年10月13日(2004.10.13)

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100058479
弁理士　鈴江　武彦
(74)代理人　100084618
弁理士　村松　貞男
(74)代理人　100092196
弁理士　橋本　良郎
(74)代理人　100091351
弁理士　河野　哲
(74)代理人　100088683
弁理士　中村　誠
(74)代理人　100070437
弁理士　河井　将次

最終頁に続く

(54)【発明の名称】コンピュータシステム

(57)【特許請求の範囲】
【請求項１】
不揮発性記憶装置への書き込みデータと、その論理アドレスとを一時的に保持するバッファ手段と、
前記バッファ手段によって一時的に保持される書き込みデータの数がｎ－１個に達した場合に、当該ｎ－１個の書き込みデータと、当該ｎ－１個の書き込みデータの論理アドレスとタイムスタンプとを含む１個の管理データとからｎ個のデータ分のサイズを持つストライプを作成し、この作成したストライプを前記不揮発性記憶装置に書き込むと共に、スナップショット採取の指示を受けた場合に、その時点で前記バッファ手段によって一時的に保持されているｎ－１個以下の書き込みデータと、当該ｎ－１個以下の書き込みデータの論理アドレスとタイムスタンプとを含む１個の管理データとから前記ストライプを作成し、この作成したストライプを前記不揮発性記憶装置に書き込むスナップショット採取手段と、
システム起動時に、前記不揮発性記憶装置に書き込まれたすべてのストライプの管理データを用いて、各論理アドレスについて最新のデータが格納された前記不揮発性記憶装置上の物理アドレスを取得するための変換マップを作成する変換マップ作成手段と、
スナップショット参照の指示を受けた場合に、当該スナップショットのタイムスタンプ値よりも小さいタイムスタンプ値を持つストライプの管理データを用いて、各論理アドレスについて当該指示されたスナップショット採取時において最新のデータが格納された前記不揮発性記憶装置上の物理アドレスを取得するための変換マップを作成し、この作成し

た変換マップに基づき、前記不揮発性記憶装置上のデータを参照するスナップショット参照手段と、
を具備したことを特徴とするコンピュータシステム。
【請求項２】
今回のシステム起動が前回のシステム再起動失敗による再起動である旨を検出する検出手段と、
前回のシステム起動時に適用したスナップショットの識別情報を管理する識別情報管理手段と、
システム起動が失敗したときに、前記識別情報管理手段により管理された識別情報に基づき、スナップショットの世代を遡ったシステム再起動を自動的に実行する自動再起動手段と、
をさらに具備したことを特徴とする請求項１記載のコンピュータシステム。
【請求項３】
前記検出手段は、前回のシステム起動時からの経過時間によって前記前回のシステム起動が失敗した旨を検出する請求項２記載のコンピュータシステム。
【請求項４】
システム再起動を予め定められた条件で自動的に実行する自動モードおよびオペレータへの問い合わせを介在させながらシステム再起動を実行する手動モードのいずれかを設定する再起動モード設定手段をさらに具備してなることを特徴とする請求項１、２または３記載のコンピュータシステム。
【請求項５】
前記自動モードは、最新のスナップショットまたは特定のスナップショットのいずれを参照するかを条件として定める請求項４記載のコンピュータシステム。
【請求項６】
最新のスナップショットを参照するとした場合、システムの再起動後、予め定められた時間内に再度システムがダウンしたときに、各世代ごとに何度再起動を試行するか、および何世代まで遡るかの少なくとも一方をさらに定める請求項５記載のコンピュータシステム。
【請求項７】
システム再起動に関する履歴情報を管理する履歴情報管理手段と、前記履歴情報管理手段により管理された履歴情報を表示または通知する履歴情報呈示手段をさらに具備してなることを特徴とする請求項４、５または６記載のコンピュータシステム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
この発明は、たとえばディスク装置などに格納されたファイル群の所定の時点における内容が保持されるスナップショットを適宜に採取するコンピュータシステムに係り、特に任意のスナップショットのディスクイメージでのシステム再起動を容易に実行することのできるコンピュータシステムに関する。
【０００２】
【従来の技術】
従来より、ディスク単体の障害に対しては、冗長化ディスク構成を採用することによって対処が可能であるが、この冗長化技術は、ソフトウェアバグ、誤操作およびウイルス感染などによるプログラムまたはデータの喪失・改変、もしくはディスク記憶装置そのものを失なうような大規模な障害には役立たない。これらの問題に対処するためには、スナップショットやバックアップなどの採取が必須となる。
【０００３】
スナップショットとは、ある時点のファイルやディスクなどのコピーイメージのことであり、定期的にファイルやディスク全体を同じディスク記憶装置や別のディスク記憶装置にコピーして、スナップショットを作成する。そして、プログラムまたはデータが喪失・改

変された場合には、直前に作成されたスナップショット（コピーイメージ）にプログラムまたはデータ、もしくはディスク全体を戻すことによりこれらの障害に対処する。

【０００４】

また、バックアップとは、このスナップショットを別の大容量記憶媒体（磁気テープなど）に保存して保管したものである。これにより、ディスク記憶装置そのものを失った場合（中のデータも全部喪失）であっても、新しいディスク記憶装置にこのバックアップをロードすることによって、スナップショット採取時のプログラムまたはデータを復元することができる。

【０００５】

通常、スナップショットやバックアップの採取時には、それらのファイルを変更する可能性があるアプリケーションをすべて停止する必要がある。そうしないと、コピー作成中にファイルやデータが変更されてしまい、正しいスナップショットやバックアップが採取できないからである。すなわち、不正なスナップショットやバックアップをディスク記憶装置に戻しても、アプリケーションプログラム側でエラーや誤動作を発生させる危険性があるため、先の障害対策として意味をなさない。

【０００６】

なお、一般に大容量記憶媒体よりもディスク記憶装置への書き込みの方が高速であるので、ファイルやディスクイメージを直接バックアップするよりも、スナップショットを作成し、それを大容量記憶媒体にバックアップする方がアプリケーションの停止期間を短く押さえられる。また、プログラムまたはデータの喪失・改変からの復元もスナップショットの方が短期に行なえる。バックアップが磁気テープなどのシーケンシャルアクセスの媒体の場合、特定のファイルを見つけて読み出すには多くの時間を要してしまい非常に効率が悪い。

【０００７】

また、このスナップショットやバックアップを採取するためにアプリケーションを長時間停止しなければならないといった問題を解決する、スナップショット機能を持ったファイルシステムも開示されている（文献１：「Ｔｈｅ Ｅｐｉｓｏｄｅ Ｆｉｌｅ Ｓｙｓｔｅｍ」，Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ｔｈｅ Ｗｉｎｔｅｒ １９９２ ＵＳＥＮＩＸ Ｃｏｎｆｅｒｅｎｃｅ，ｐｐ．４３－６０，Ｓａｎ Ｆｒａｎｃｉｓｃｏ，ＣＡ、文献２：「Ｆｉｌｅ ＳｙｓｔｅｍＤｅｓｉｇｎ ｆｏｒ ａｎ ＮＦＳ Ｆｉｌｅ Ｓｅｒｖｅｒ Ａｐｐｌｉａｎｃｅ」、Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ｔｈｅ Ｗｉｎｔｅｒ １９９４ＵＳＥＮＩＸ Ｃｏｎｆｅｒｅｎｃｅ，ｐｐ．２３５－２４４，Ｓａｎ Ｆｒａｎｃｉｓｃｏ，ＣＡ）。

【０００８】

ところで、これらのファイルシステム（たとえば文献２）では、ファイルがルート・ノードを起点とするブロック木構造であることを前提にしている。そして、スナップショット作成時には、このルート・ノードのコピーを作成する。このコピーしたルート・ノード（スナップショット・ルートノード）は、スナップショットとしてのブロック木構造を表現することになる。また、コピー元のルート・ノードは、読み出し／書き込みが行なわれるアクティブなファイルシステムのブロック木構造を表現する。スナップショット・ルートノードは、生成されたときにはルート・ノードとまったく同じブロックを指しているので、新しいスナップショットのためにコピーしたルート・ノード以外はまったくディスクスペースを消費しない。

【０００９】

ここで、ユーザがいずれかのデータブロックを変更したとすると、新しいデータブロックをディスク上に書き込み、ルート・ノードで表現するアクティブなファイルシステムが、この新しいブロックを指すように変更される。ディスク上の元のデータブロックは変更されずに残っており、スナップショット・ルートノードは、元のデータブロックをまだ指している。したがって、スナップショット・ルートノードを指定することにより、スナップショット作成時のデータブロックをそのまま参照することができる。

【００１０】
このように、スナップショット機能を持ったファイルシステムを使えば、ルート・ノードをコピーするだけで簡単にスナップショットを作成でき、アプリケーションプログラムを停止することもない。また、スナップショットをアプリケーションプログラムの実行と並行して大容量記憶媒体に保存できるので、バックアップの採取もアプリケーションを停止することなく行なえる。したがって、スナップショットやバックアップのデータ退避のためにアプリケーションプログラムを停止する必要がなくなる。
【００１１】
【発明が解決しようとする課題】
しかしながら、前述した手法においては、専用のファイルシステムを新しく開発する必要があり、既存のコンピュータシステムにそのまま適用できるものではなかった。また、ファイルシステム自身もブロック木構造で構成されていることを前提にしており、たとえばマイクロソフト社のＮＴＦＳなどのエクステントベースのファイルシステムには適用できる技術ではなかった。また、一般にファイルシステムは大きな木構造をしており、実際に前述のデータブロックを更新するには、その経路に位置する中間ノードすべてをコピーする必要があり、スナップショット作成後の更新性能が大きく低下するといった問題もあった。さらに、ファイルシステムという複雑なソフトウェアモジュールにスナップショット機能を付加したため、スナップショットは読み出しだけで非常に硬直なものになってしまっていた。
【００１２】
この発明はこのような実情に鑑みてなされたものであり、既存のコンピュータシステムやファイルシステムにそのまま適用することが可能であって、スナップショットを効率的に採取するとともに、任意のスナップショットのディスクイメージでのシステム再起動を容易に実行することのできるコンピュータシステムを提供することを目的とする。
【００１３】
【課題を解決するための手段】
前述した目的を達成するために、この発明のコンピュータシステムは、不揮発性記憶装置への書き込みデータと、その論理アドレスとを一時的に保持するバッファ手段と、前記バッファ手段によって一時的に保持される書き込みデータの数がｎ－１個に達した場合に、当該ｎ－１個の書き込みデータと、当該ｎ－１個の書き込みデータの論理アドレスとタイムスタンプとを含む１個の管理データとからｎ個のデータ分のサイズを持つストライプを作成し、この作成したストライプを前記不揮発性記憶装置に書き込むと共に、スナップショット採取の指示を受けた場合に、その時点で前記バッファ手段によって一時的に保持されているｎ－１個以下の書き込みデータと、当該ｎ－１個以下の書き込みデータの論理アドレスとタイムスタンプとを含む１個の管理データとから前記ストライプを作成し、この作成したストライプを前記不揮発性記憶装置に書き込むスナップショット採取手段と、
システム起動時に、前記不揮発性記憶装置に書き込まれたすべてのストライプの管理データを用いて、各論理アドレスについて最新のデータが格納された前記不揮発性記憶装置上の物理アドレスを取得するための変換マップを作成する変換マップ作成手段と、スナップショット参照の指示を受けた場合に、当該スナップショットのタイムスタンプ値よりも小さいタイムスタンプ値を持つストライプの管理データを用いて、各論理アドレスについて当該指示されたスナップショット採取時において最新のデータが格納された前記不揮発性記憶装置上の物理アドレスを取得するための変換マップを作成し、この作成した変換マップに基づき、前記不揮発性記憶装置上のデータを参照するスナップショット参照手段と、を具備したことを特徴とする。
【００１４】
この発明によれば、既存のオペレーティングシステムやファイルシステム、あるいはアプリケーションプログラムになんらの改造を伴なうことなく、スナップショットを効率的に採取することが可能となり、また、システムの稼働状況などに応じて選択される任意のスナップショットのディスクイメージでシステムを再起動させることが容易に行なえるため

、システムの可用性向上とシステム管理の省力化とを図ることが可能となる。
【００１６】
【発明の実施の形態】
以下、図面を参照してこの発明の実施の形態を説明する。
まず、この実施形態におけるディスクスナップショットの管理手法について説明する。図１は、この実施形態のコンピュータシステムでディスクスナップショットの管理を司るディスクスナップショット部の概念構成を示す図である。図１に示すように、この実施形態のディスクスナップショット部２は、ディスクスナップショット制御部３および揮発性メモリ５から構成される。そして、この揮発性メモリ５内には、書き込みの時間的順序を維持するためのタイムスタンプ６、ディスク装置４に書き込むデータをログ構造化して保持する書き込みバッファ７、および書き込みバッファ７内の空き領域ならびに保持されている書き込みデータの論理アドレスの情報を保持するバッファ管理テーブル８が格納される。さらに、ディスク装置４内には、通常のデータのほかにスナップショットを管理するためのスナップショット情報（ＳＳ情報）１０が格納されている。
【００１７】
ディスクスナップショット制御部３は、これらタイムスタンプ６、書き込みバッファ７、バッファ管理テーブル８およびスナップショット情報１０を管理し、ディスク装置４への読み出し／書き込みおよびスナップショットの参照を制御する。
【００１８】
ここで、データの書き込み動作について説明する。ディスク装置４は、それぞれデータブロックサイズの整数倍（ｋ）であるストライプと呼ばれる単位で書き込みを行なう。ディスクスナップショット制御部３は、ファイルシステム１から書き込むデータとその論理アドレスとを受け取り、揮発性メモリ５上の書き込みバッファ７の空き領域にブロック単位に分割して詰める。また、受け取った論理アドレスをブロックごとのアドレスに変換して、バッファ管理テーブル８の対応するエントリに格納する。図２には、この書き込みバッファ７およびバッファ管理テーブル８の内容が例示されている。
【００１９】
そして、ディスクスナップショット制御部３は、ファイルシステム１からの書き込みデータが、１ストライプ分に１ブロック少ない数だけ書き込みバッファ７に溜まった段階で、図３に示すように、最後の１ブロックにストライプ１１上のブロックの論理アドレスとタイムスタンプとを格納して論理アドレスタグブロック１２とし、そのストライプ１１をディスク装置４に書き込む。また、タイムスタンプ６の値は、この書き込みが完了した段階でインクリメントされる。
【００２０】
このように、データの更新に際して、ディスク装置４上の該当する物理領域を書き換えるのではなく、別の空き領域にタイムスタンプとともに書き込んでいくので、ディスク装置４上には同じ論理アドレスをもつデータブロックが複数存在する。通常のデータ参照要求に対しては、最新のデータブロックの物理位置を返す必要があるので、システム起動時に、全ストライプの論理アドレスタグを調べて揮発性メモリ５上に論理アドレスから物理アドレスへの変換マップ９を作成する。変換マップ９は、図４に示すように、各論理アドレスのブロックが格納されているストライプ番号ＳＴ＃と、そのストライプ内のブロック番号ＢＬＫ＃と、そのタイムスタンプＴＳ＃とをテーブル方式で保持している。これにより、各論理アドレスブロックの物理位置が求められる。
【００２１】
次に、スナップショット採取処理について説明する。スナップショット採取の指示を受けると、ディスクスナップショット制御部３は、その時点で書き込みバッファ７に溜まっているデータをディスク装置４に書き込む。このとき、書き込みバッファ７の空き領域に該当するバッファ管理テーブル８のエントリには、論理アドレスとしてヌル（空値：Ｎｕｌｌ　Ｖａｌｕｅ）を設定する。その後、通常の書き込みと同様に、最後の１ブロックとして論理アドレスタグブロック１２を作成し、１ストライプ分を書き込む。さらに、このと

きの論理アドレスタグブロック１２に付加したタイムスタンプ６の値をスナップショット採取時の情報として、ディスク装置４内のディスクスナップショット情報１０に登録し、完了したらタイムスタンプ６の値をインクリメントする。

【００２２】

次に、スナップショットの参照処理について説明する。あるスナップショットのイメージを参照するためには、参照するスナップショットのタイムスタンプ値より小さいタイムスタンプ値をもつストライプに含まれるすべての論理アドレスブロックについて、システム起動時と同様の変換マップを作成し、そのスナップショット参照用変換マップを使ってアクセスすればよい。

【００２３】

このように、この実施形態のディスクスナップショットの管理手法によれば、既存のオペレーティングシステムやファイルシステム、あるいはアプリケーションプログラムになんらの改造を伴なうことなく、スナップショットを効率的に採取することが可能となる。以下では、このように採取したスナップショットの中の任意のスナップショットのディスクイメージでシステムを起動させるための手法について説明する。

【００２４】

図５は、最新のスナップショットをシステム起動時のディスクイメージとするシステムの構成図である。

システム起動時に、ディスクスナップショット制御部３は、スナップショット情報１０を読み込んで、最新のスナップショットのタイムスタンプ（ＴＳ）６を得る。また、ディスク装置４の中の有効なストライプ１１の位置が記録されているストライプ管理情報（ＳＴ）１５を揮発性メモリ５上に読み込む。そして、ディスクスナップショット制御部３は、このストライプ管理情報１５′をもとにディスク装置４から有効なストライプを読み込みバッファ１４に読み込み、ストライプ１１の論理アドレスタグブロック１２のタイムスタンプ１３と最新のスナップショットのタイムスタンプ６とを比較して、タイムスタンプ値１３がタイムスタンプ６より小さい場合のみ、ストライプ１１に含まれるデータブロックの変換マップ９への登録を行なう。このシステムの場合、最新のスナップショットよりも後に書き込まれたストライプは、以後利用されることはないので、タイムスタンプ１３がタイムスタンプ６より大きい場合は、ストライプ管理テーブル１５′の該当するエントリを削除し、そのストライプを無効化する。変換マップ９への登録処理は、論理アドレスタグブロック１２内の各ブロックの論理アドレスを変換マップ９に登録されている論理アドレスと比較し、登録されていなければ登録し、すでに登録されている場合には、タイムスタンプ値１３が登録されているタイムスタンプ値よりも大きい場合のみ、エントリを更新して登録する。ディスク装置４内の有効な全ストライプに対してこの処理を行なって作成した変換マップ９を使用することにより、通常と同じ手順で最新のスナップショット時点のディスクイメージにアクセスできる。そして、この実施形態のディスクスナップショット部２では、これを擬似ディスク装置４′で表す。すなわち、続いて行なわれるファイルシステム１の作成処理において、ディスクスナップショット制御部３がディスク装置４の代わりに擬似ディスク装置４′を見せることにより、擬似ディスク装置４′をディスクイメージとしてアクセスするファイルシステム１が作成され、アプリケーションプログラムは擬似ディスク装置であることを意識することなく実行することができる。

【００２５】

また、図５に示したシステムの構成に、システムダウンしたときにディスクイメージを戻したい日時が予め指定できる機能と、スナップショット情報１０からキーとなるタイムスタンプ６を得る処理においてその指定された日時以前で最も新しいスナップショットのタイムスタンプ１３を選択する機能とを付け加えることによって、指定した日時のディスクイメージまで戻してシステム再起動することが可能になる。

【００２６】

さらに、図５に示したシステムの構成に、スナップショット情報１０を揮発性メモリ５内に読み込んだ後、有効なスナップショット１１の日時を画面表示する機能と、その中から

オペレータが選択したスナップショット１１のタイムスタンプ１３を変換マップ６の作成時のキーとなるタイムスタンプ６として設定する機能とを付け加えることによって、オペレータの指定したスナップショット時点のディスクイメージでシステム起動することが可能になる。なお、このとき必ず選択画面を表示するというオプション情報を設けてチェックするようにすれば、システムダウンした場合だけでなく、通常の立ち上げ時にも任意のスナップショットを選択してブートさせることが可能となる。

【００２７】

図６は、システム起動が失敗した場合にスナップショットの世代を遡った自動再起動が可能なシステムのブート時の処理の流れを示すフローチャートである。図５に示したシステム構成に付加する必要がある要素は、前回の起動が失敗したことを検出する機能と、前回の起動で使用したスナップショットの識別情報を保存する機能とである。

【００２８】

この場合、ディスクスナップショット制御部３は、まず、前回システムが正常終了したかどうかを判定する（ステップＡ１）。この処理は、予め決まった場所にある情報を参照することによって行なう。次に、ディスクスナップショット制御部３は、この再起動が初めてのものであるか、または再起動失敗に伴なう再起動かを判定する（ステップＡ２）。ここでは、保存されている前回のリスタート開始時刻と現在の時刻とを比較して、前回システム再起動した時刻から予め指定された時間または所定の時間を経過しているかどうかを調べると同時に、リスタート開始時刻を現在の時刻で更新する。

【００２９】

再起動失敗の後の再起動の場合であった場合（ステップＡ２のＮ）、ディスクスナップショット制御部３は、参照するスナップショットを前回使用したスナップショットより一世代旧いものにし、使用するスナップショットの識別情報を更新する（ステップＡ３）。具体的には、前回の再起動時に参照したスナップショットのタイムスタンプが保存されているので、スナップショット情報１０を検索してその一世代前のスナップショット１１を選択するとともに、このスナップショット１１のタイムスタンプ１３の値でスナップショット情報１０のタイムスタンプ６を更新する。次に、ディスクスナップショット制御部３は、選択されたスナップショット１１のタイムスタンプ１３をキーとして、このスナップショット１１の参照用変換マップ９を作成する（ステップＡ３）。この後は、ファイルシステム作成から通常の立ち上げ処理に入る（ステップＡ５，ステップＡ６）。

【００３０】

このように、スナップショットの世代を遡った再起動を自動的に実行することが可能となる。

次に、システム再起動を予め定められた条件で自動的に実行する自動モードおよびオペレータへの問い合わせを介在させながらシステムの再起動を実行する手動モードのいずれかによるリスタートを実現する手法について説明する。図７は、このリスタートを実現するコンピュータシステムの構成図である。

【００３１】

図７に示すように、このコンピュータシステム１００は、端末装置１０１を備えており、この端末装置１０１は、オペレータがコンピュータシステム１００に対して指示を行なうための入力装置と処理結果やその他の情報を表示するための出力装置とからなっている。

【００３２】

このコンピュータシステム１００は、前述のシステムが実現する機能に加え、さらにきめ細かな再起動指定を可能にするために、リスタート情報設定モジュール（ＲＳＴ）２１を用意し、リスタート自動／手動、自動のときのディスクイメージの日時指定、再起動失敗を判断する際の前回再起動からの経過時間Ｔ、スナップショット一世代ごとの再起動リストライ回数Ｎ、およびスナップショットを遡るときの世代数の制限値Ｍなどのリスタート設定を、予めオペレータに設定させ、リスタート設定情報（ＲＳＴ）１６としてディスク装置４に保存する。

【００３３】

一方、これらの指定にしたがってリスタートを行なうためには、現在何回目のリスタートであるか、再起動失敗後なのか、などのリスタート履歴を保存し、参照する必要があるので、さらにリスタート履歴情報（ＲＬＯＧ）１７をディスク装置４に保存する。そして、図６のステップＡ１で示した、システムが前回正常終了したかどうかを判定するリスタート判定モジュール３１を設けて、リスタート回数Ｋ（正常終了からの）を設定するために用いる。また、カレントの世代のスナップショットでのリトライ回数ｎ、現在何世代さかのぼったかを示すカレント世代数ｍ、リスタート開始時刻、および起動に使用したスナップショットのタイムスタンプなどもリスタート履歴情報１７として保存する。

【００３４】

ブート時に使用するスナップショットを選択するシナップショット選択モジュール３４は、リスタート判定モジュール３１、リスタート設定情報参照モジュール３２、およびリスタート履歴情報参照・更新モジュール３３を利用して指定に応じたスナップショットを選択し、ディスクスナップショット制御部３に指示してそのスナップショットを参照するための変換マップ９を作成させ、擬似ディスク装置４′をディスクイメージとしてファイルシステム１に見せる。

【００３５】

オペレータが起動に使うスナップショットを選択する指定では、ファイルシステム１が立ち上がる前のブート処理の段階で、その選択画面の表示を行なう必要があるので、スナップショット検索・一覧モジュール３５を用意してスナップショット情報１０を取得し、オペレータ問い合わせモジュール３６を介してスナップショットの一覧を端末装置１０１に出力し、オペレータの選択したスナップショットのタイムスタンプをスナップショット選択モジュール３４に通知する。

【００３６】

なお、これらのリスタート管理のために用意した制御モジュールをリスタート管理モジュール３０と呼び、一方、リスタート管理のためのアプリケーション群をリスタート管理ユーティリティ２０と呼ぶことにする。

【００３７】

図８は、ブート時にリスタート管理モジュール３０が行なう処理の流れを示すフローチャートである。

リスタート管理モジュール３０は、まず、システムが前回正常に終了したかどうかを判定する（ステップＢ１）。この処理は、ある決まった領域にハードウェアまたは基本システムモジュールが設定する情報を参照することによって行なわれる。システムダウンだった場合（ステップＢ１のＮ）、リスタート管理モジュール３０は、リスタート履歴情報１７の中のリスタート回数Ｋをインクリメントする（ステップＢ２）。このリスタート回数Ｋは、システムのシャットダウン（正常終了処理）の際にリスタート履歴参照・更新モジュール３３によってクリアされる。

【００３８】

次に、リスタート管理モジュール３０は、システムダウンの場合の再起動の指定が自動か手動かを判断する（ステップＢ３）。この処理は、リスタート設定情報１８の中の該当する情報をリスタート設定情報参照モジュール３２が参照することによって行なわれる。

【００３９】

手動の場合は（ステップＢ３のＮ）、前述したようにオペレータにスナップショットを選択させた後（ステップＢ１０）、後述するステップＢ８へと進む。一方、自動の場合は（ステップＢ３のＹ）、まず、起動ディスクイメージの日時指定があるかどうかを判定する（ステップＢ４）。この処理もステップＢ３と同様に、リスタート設定情報１８を参照することによって行なわれ、指定があった場合には（ステップＢ４のＹ）、スナップショット選択モジュール３４でその日時以前の中で最新のスナップショットを選択した後（ステップＢ１２）、後述するステップＢ８へと進む。また、日時指定がない場合（ステップＢ４のＮ）、さらに前回の起動から一定時間が経過しているかどうかを調べる（ステップＢ５）。これは、正常実行後初めての再起動なのか、再起動が失敗した直後の再起動なのか

を判定するためであり、図６のステップＡ２と同様に、リスタート履歴情報参照・更新モジュール３３でリスタート履歴情報１７の中のリスタート開始時刻の参照・更新を行なう。この判断に使用する前回リスタートからの経過時間の基準は、既定値またはリスタート設定情報１８から取得する。

【００４０】

ここで、前回のリスタートから指定時間が経過しており、一回目の再起動と判断した場合は（ステップＢ５のＹ）、カレントリトライ回数ｎを１に、カレント世代数ｍを０にセットした後（ステップＢ６）、最新のスナップショットを選択する（ステップＢ７）。

【００４１】

また、前回のリスタートから指定時間内に再びシステムダウンした場合は（ステップＢ５のＮ）、スナップショット選択処理（ステップＢ９）に進み、ここで条件にあったスナップショットを選択する。

【００４２】

そして、これまでの処理で選択されたスナップショットから起動用のディスクイメージ（擬似ディスク装置４′）を作成する（ステップＢ８）。具体的には、その選択されたスナップショットのタイムスタンプをキーとして、変換マップ９へ論理アドレスの登録を行ない、擬似ディスク装置４′へのアクセス用の変換マップ９を作成する処理となる。

【００４３】

図９は、スナップショット選択処理（図８のステップＢ９）の流れを示すフローチャートである。

このスナップショット選択処理では、まず、カレントリトライ回数ｎとリスタート設定情報１６にある最大リトライ回数Ｎ（一世代のスナップショットに関して連続して何回リスタートを試みるか）とを比較し（ステップＣ１）、ｎ＜Ｎならば（ステップＣ１のＹ）、前回と同じスナップショットを選択するのでカレントリトライ回数をインクリメントして（ステップＣ４）終了する。そうでなければ（ステップＣ１のＮ）、一世代前のスナップショットを選択するために、カレント世代数ｍとリスタート設定情報１６の中の最大世代数Ｍ（何世代さかのぼって自動リスタートを行なうか）とを比較し（ステップＣ２）、指定されている世代数を越えていたら（ステップＣ２のＮ）、立ち上げを中断する（ステップＣ５）。この場合は、その旨のメッセージを表示または通知し、リスタート履歴情報１７にログを残す。一方、ｍ＜Ｍならば（ステップＣ２のＹ）、カレント世代数ｍをインクリメントし、カレントリトライ回数ｎを１に設定して、前回使用したスナップショットよりも一世代旧いスナップショットを選択する（ステップＣ３）。この処理は、スナップショット情報１０に有効なスナップショット１１とタイムスタンプ６とが格納されているので、リスタート履歴情報１７に保存している、前回使用したスナップショット１１のタイムスタンプ１３よりも小さい中で最大のものを選び、この選んだスナップショット１１のタイムスタンプ１３でタイムスタンプ６を更新する。

【００４４】

以上の構成要素により、システム再起動を予め定められた条件で自動的に実行する自動モードおよびオペレータへの問い合わせを介在させながらシステム再起動を実行する手動モードのいずれかによるリスタートを実現できるが、最新のスナップショット以外のディスクイメージで再起動した場合は、起動に使用したスナップショットよりも後に採取されたスナップショットと再起動後のディスクイメージとの関係を考えなければならない。たとえば、図１０に示すように、スナップショット３（ＳＳ３）の後にシステムがダウンし、オペレータがスナップショット１（ＳＳ１）を使用して再起動したとすると、再起動後にスナップショット４（ＳＳ４）を採取するのであれば、その前にスナップショット２（ＳＳ２）およびスナップショット３（ＳＳ３）は無効化しなければならない。スナップショット４（ＳＳ４）を採取した後に、スナップショット２（ＳＳ２）およびスナップショット３（ＳＳ３）が残っていると、そこでシステムダウンした場合にディスクイメージの整合性がとれなくなってしまうからである。そこで、リスタート履歴情報１７に、図８のステップＢ９，Ｂ１１，Ｂ１２で最新のスナップショット以外を選択した場合にオンとする

フラグを設け、ディスクスナップショット制御部３によるスナップショット採取処理でそのフラグをチェックし、オンの場合には、該当するスナップショット（図１０の場合には、スナップショット２（ＳＳ２）およびスナップショット３（ＳＳ３））を削除する。スナップショットの削除は、スナップショット情報１０に登録されているスナップショットの中でリスタートに使用したスナップショットのタイムスタンプよりも大きいタイムスタンプをもっているスナップショットのエントリを削除するとともに、該当するストライプも無効化する。この処理は、ストライプの物理的な位置とタイムスタンプとが登録されているストライプ管理情報１５を検索し、リスタートに使用しているスナップショットのタイムスタンプより大きく、リスタート開始時刻よりも小さいタイムスタンプをもつストライプを無効化すればよい。

【００４５】

なお、この例では、スナップショットの無効化は再起動後の最初のスナップショット採取直前に行なわれるので、スナップショット採取を行なわないように設定する手段があれば、スナップショット２（ＳＳ２）およびスナップショット３（ＳＳ３）は再起動後も残すことができる。これにより、たとえば実験などのために、およびスナップショット１（ＳＳ１）やスナップショット０（ＳＳ０）のイメージで起動して実行し、次回の起動では、スナップショット２（ＳＳ２）またはスナップショット３（ＳＳ３）を使って起動することも可能になる。図１１には、このスナップショット採取のタイミングを変更できるシステムの概念図が示されている。

【００４６】

ここでは、スナップショット採取のタイミングの制御は、タイミング制御モジュール１８が行なう。通常、決められた一定時間間隔でディスクスナップショット部２に対して実行を指示するが、次のような機能を用意することにより、タイミングを変更可能にする。すなわち、システム管理者が、スナップショット採取タイミングの指定を行なうためのスナップショット採取モジュール２３を用意することにより、タイミング制御モジュール１８にスナップショット採取実行や採取間隔の指示を行なう。このタイミング制御モジュール１８は、指定された時間ごとにディスクスナップショット部２に対してスナップショット採取実行を指示するほかに、スナップショット採取モジュール２３からの指示で即時的に実行を指示する。また、採取間隔が０に設定されたときは採取を行なわない。

【００４７】

さらに、採取を行なう指定で実行していたシステムがシステムダウンした場合に、採取を行なわない指定で再起動したいときのために、図７で示したオペレータ問い合わせモジュール３６に相当するモジュールでスナップショット採取に関する指定を可能にし、システム実行が開始される前にその情報をタイミング制御モジュール１８に通知する。

【００４８】

また、システムの再起動を予め定められた条件で自動的に実行する自動モードおよびオペレータへの問い合わせを介在させてシステムの再起動を実行する手動モードのいずれかによるリスタートを実現するような、きめ細かなリスタート制御を可能とする、リスタート履歴情報１７を有するシステムでは、リスタート管理ユーティリティ２０の一部として、リスタート履歴情報１７を表示したり通知したりするモジュールを用意することにより、リスタートの履歴をシステムの診断に利用することが可能となる。

【００４９】

【発明の効果】

以上詳述したように、この発明によれば、たとえばディスク装置などの不揮発性記憶装置とファイルシステムとの間に、ディスクレベルのスナップショットを採取するスナップショット管理手段を介在させる構造を採用するため、既存のオペレーティングシステムやファイルシステム、あるいはアプリケーションプログラムになんらの改造を伴なうことなく、スナップショットを効率的に採取することが可能となり、また、システムの稼働状況などに応じて選択される任意のスナップショットのディスクイメージでシステムを再起動させることが容易に行なえるため、システムの可用性向上とシステム管理の省力化とを図る

ことが可能となる。

【図面の簡単な説明】

【図１】この発明の実施形態に係るコンピュータシステムでディスクスナップショットの管理を司るディスクスナップショット部の概念構成を示す図。

【図２】同実施形態の書き込みバッファおよびバッファ管理テーブルの内容を例示する図。

【図３】同第実施形態の最後の１ブロックにストライプ上のブロックの論理アドレスとタイムスタンプとを格納して論理アドレスタグブロックとし、そのストライプをディスク装置に書き込む様子を示す図。

【図４】同実施形態の変換マップの記憶形式を示す図。

【図５】同実施形態の最新のスナップショットをシステム起動時のディスクイメージとするシステムの構成図。

【図６】同実施形態のシステム再起動が失敗した場合にスナップショットの世代を遡った自動再起動が可能なシステムのブート時の処理の流れを示すフローチャート。

【図７】同実施形態のシステム再起動を予め定められた条件で自動的に実行する自動モードおよびオペレータへの問い合わせを介在させながらシステム再起動を実行する手動モードのいずれかによるリスタートを実現するコンピュータシステムの構成図。

【図８】同実施形態のブート時にリスタート管理モジュールが行なう処理の流れを示すフローチャート。

【図９】同実施形態のスナップショット選択処理の流れを示すフローチャート。

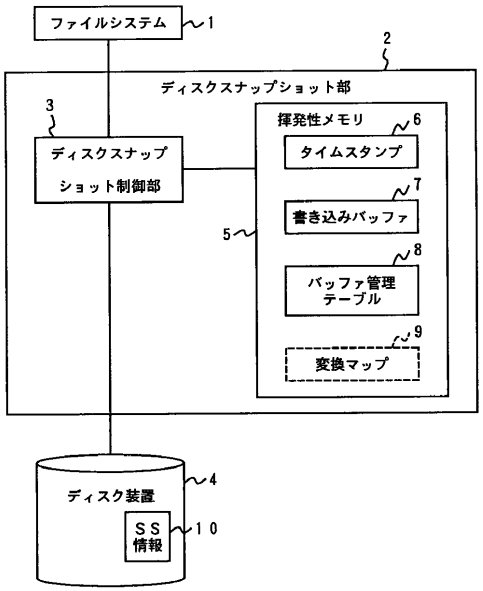
【図１０】同実施形態の最新のスナップショット以外のディスクイメージで再起動した場合の考慮点を説明するための図。

【図１１】同実施形態のスナップショット採取のタイミングを変更できるシステムの概念図。

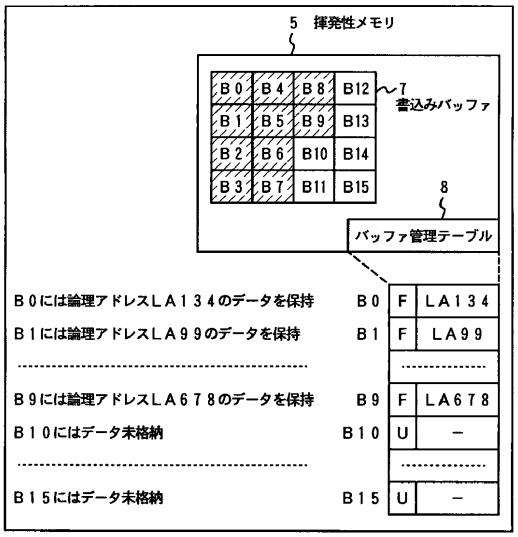
【符号の説明】

１…ファイルシステム、２…ディスクスナップショット部、３…ディスクスナップショット制御部、４…ディスク装置、５…揮発性メモリ、６…タイムスタンプ、７…書き込みバッファ、８…バッファ管理テーブル、９…変換マップ、１０…スナップショット情報、１１…ストライプ、１２…論理アドレスタグブロック、１３…タイムスタンプ、１４…読み込みバッファ、１５…ストライプ管理情報、２０…リスタート管理ユーティリティ、２１…リスタート情報設定モジュール（ＲＳＴ）、３０…リスタート管理モジュール、３１…リスタート設定情報参照モジュール、３３…リスタート履歴情報参照・更新モジュール、３４…スナップショット選択モジュール、３５…スナップショット検索・一覧モジュール、３６…オペレータ問い合わせモジュール、１００…コンピュータシステム。

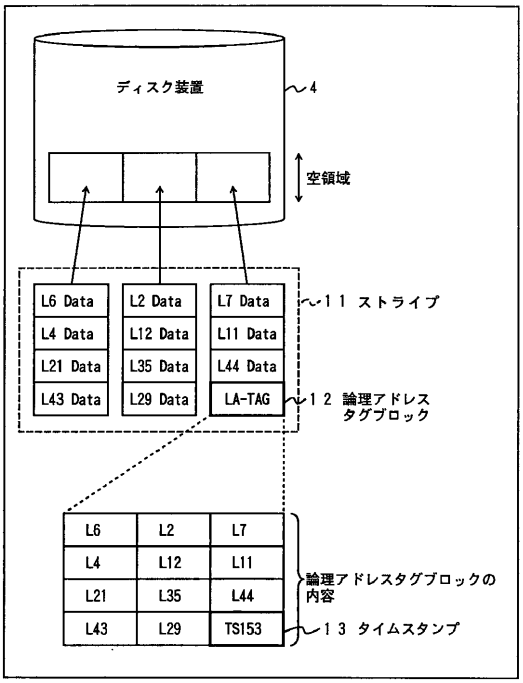
【図１】


ファイルシステム
1
2
ディスクスナップショット部
3
ディスクスナップ
ショット制御部
揮発性メモリ
6
タイムスタンプ
7
書き込みバッファ
5
8
バッファ管理
テーブル
9
変換マップ
4
ディスク装置
SS
情報
10

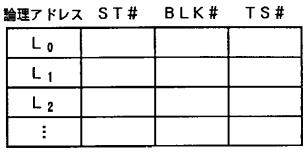
【図２】


5 揮発性メモリ
B0 B4 B8 B12
B1 B5 B9 B13
B2 B6 B10 B14
B3 B7 B11 B15
7
書込みバッファ
8
バッファ管理テーブル
B0には論理アドレスLA134のデータを保持
B1には論理アドレスLA999のデータを保持
B9には論理アドレスLA678のデータを保持
B10にはデータ未格納
B15にはデータ未格納
B0 F LA134
B1 F LA999
B9 F LA678
B10 U －
B15 U －

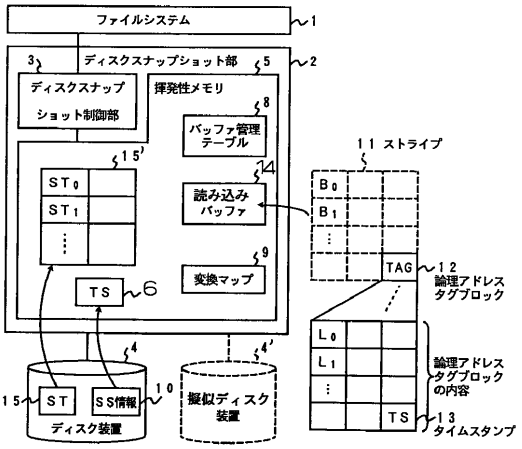
【図３】


ディスク装置
4
空領域
L6 Data
L4 Data
L21 Data
L43 Data
L2 Data
L12 Data
L35 Data
L29 Data
L7 Data
L11 Data
L44 Data
LA-TAG
11 ストライプ
12 論理アドレス
タグブロック
L6
L2
L7
L4
L12
L11
L21
L35
L44
L43
L29
TS153
論理アドレスタグブロックの
内容
13 タイムスタンプ

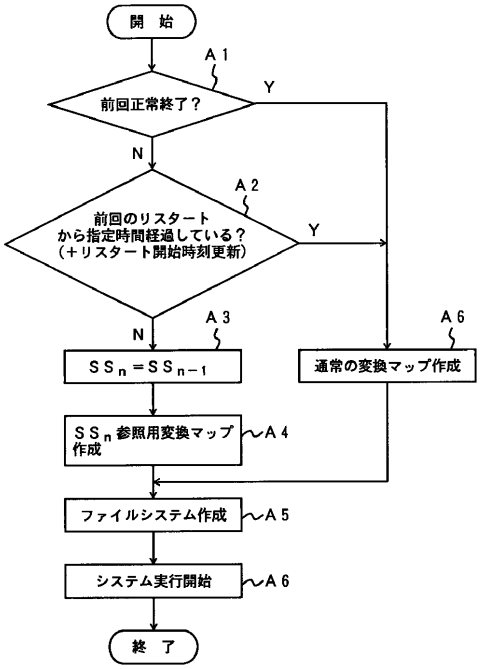
【図４】


論理アドレス ST# BLK# TS#
L0
L1
L2
⋮

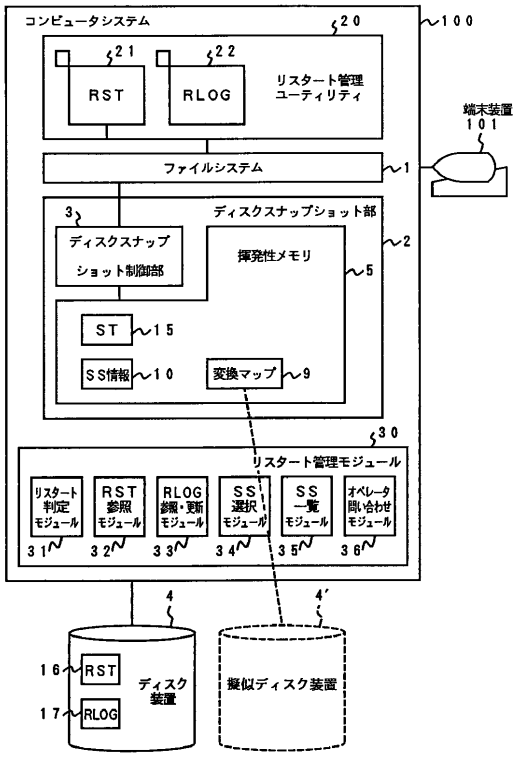
【図５】


ファイルシステム
1
3
ディスクスナップショット部
5
2
ディスクスナップ
ショット制御部
揮発性メモリ
8
バッファ管理
テーブル
15'
14
ST0
ST1
読み込み
バッファ
9
変換マップ
TS
6
11 ストライプ
B0
B1
TAG
12
論理アドレス
タグブロック
L0
L1
論理アドレス
タグブロック
の内容
TS
13
タイムスタンプ
4
4'
15
ST
SS情報
10
ディスク装置
擬似ディスク
装置

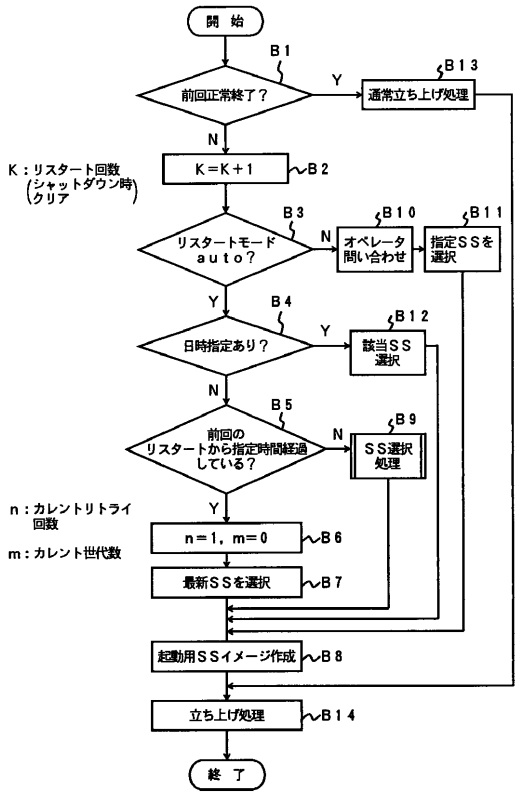
【図６】


開 始
A 1
前回正常終了？
Y
N
A 2
前回のリスタート
から指定時間経過している？
（＋リスタート開始時刻更新）
Y
N
A 3
SSn＝SSn−1
A 6
通常の変換マップ作成
SSn 参照用変換マップ作成
A 4
ファイルシステム作成
A 5
システム実行開始
A 6
終 了

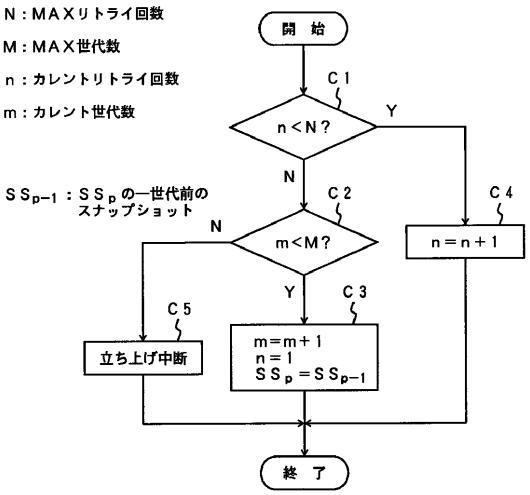
【図７】


コンピュータシステム
20
100
21
RST
22
RLOG
リスタート管理
ユーティリティ
端末装置
101
ファイルシステム
1
3
ディスクスナップショット部
ディスクスナップ
ショット制御部
揮発性メモリ
2
5
ST
15
SS情報
10
変換マップ
9
30
リスタート管理モジュール
リスタート
判定
モジュール
RST
参照
モジュール
RLOG
参照・更新
モジュール
SS
選択
モジュール
SS
一覧
モジュール
オペレータ
問い合わせ
モジュール
31
32
33
34
35
36
4
4′
16
RST
17
RLOG
ディスク
装置
擬似ディスク装置

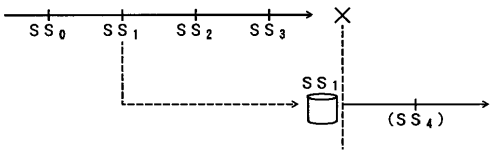
【図８】


開 始
B 1
前回正常終了？
Y
B 13
通常立ち上げ処理
N
K：リスタート回数
（シャットダウン時
クリア）
K＝K＋1
B 2
B 3
リスタートモード
ａｕｔｏ？
N
B 10
オペレータ
問い合わせ
B 11
指定SSを
選択
Y
B 4
日時指定あり？
Y
B 12
該当SS
選択
N
B 5
前回の
リスタートから指定時間経過
している？
N
B 9
SS選択
処理
Y
n：カレントリトライ
回数
m：カレント世代数
n＝1，m＝0
B 6
最新SSを選択
B 7
起動用SSイメージ作成
B 8
立ち上げ処理
B 14
終 了

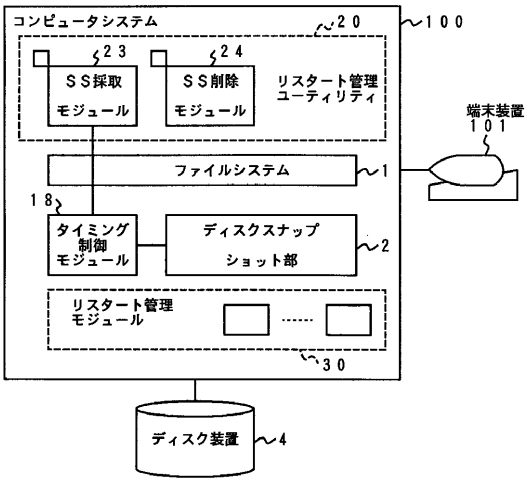
【図９】


N：MAXリトライ回数
M：MAX世代数
n：カレントリトライ回数
m：カレント世代数
SSp−1：SSpの一世代前の
スナップショット
開 始
C 1
n＜N？
Y
N
C 2
m＜M？
N
Y
C 4
n＝n＋1
C 3
m＝m＋1
n＝1
SSp＝SSp−1
C 5
立ち上げ中断
終 了

【図１０】


SS0
SS1
SS2
SS3
SS1
(SS4)

【図１１】


コンピュータシステム
20
100
23
24
SS採取
モジュール
SS削除
モジュール
リスタート管理
ユーティリティ
端末装置
101
ファイルシステム
1
18
タイミング
制御
モジュール
ディスクスナップ
ショット部
2
リスタート管理
モジュール
30
ディスク装置
4

フロントページの続き

(72)発明者　鹿目　奈美子
東京都青梅市末広町２丁目９番地　株式会社東芝青梅工場内

審査官　横山　佳弘

(56)参考文献　特開平０７－２６１９８９（ＪＰ，Ａ）
特開平０５－２８９８５４（ＪＰ，Ａ）
特開平０７－２８１９３４（ＪＰ，Ａ）
特開平０６－１１０６１８（ＪＰ，Ａ）
特開平０６－３２４８１７（ＪＰ，Ａ）
特開昭５８－０９７７２４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F　3/06
G06F　11/14
G06F　12/00
G06F　1/00
G06F　9/06