(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2012-123793
(P2012-123793A)
(43)公開日　平成24年6月28日(2012.6.28)

(51)Int.Cl.	F I		テーマコード（参考）
G06F 12/00 (2006.01)	G06F	12/00	531M
	G06F	12/00	517

審査請求 未請求　請求項の数 18　ＯＬ　（全 23 頁）

(21)出願番号　特願2011-247465 (P2011-247465)
(22)出願日　平成23年11月11日 (2011.11.11)
(31)優先権主張番号　12/963146
(32)優先日　平成22年12月8日 (2010.12.8)
(33)優先権主張国　米国 (US)

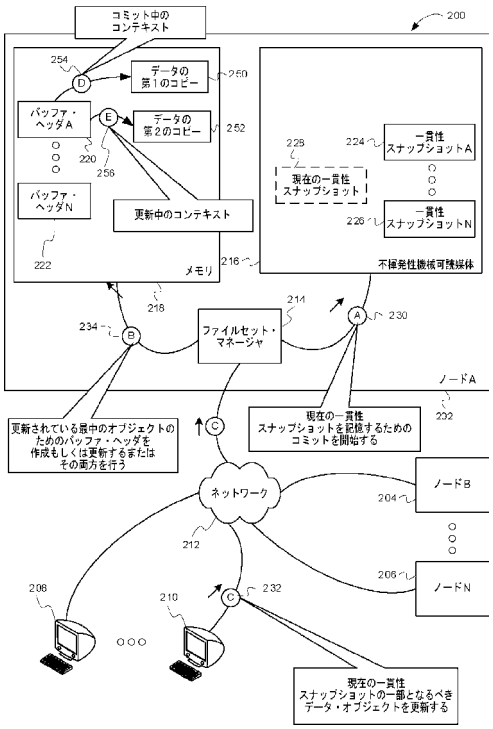
（特許庁注：以下のものは登録商標）
１．ＲＲＡＭ

(71)出願人　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国１０５０４　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
(74)代理人　100108501
弁理士　上野　剛史
(74)代理人　100112690
弁理士　太佐　種一
(74)代理人　100091568
弁理士　市位　嘉宏

最終頁に続く

(54)【発明の名称】リダイレクト・オン・ライト・ファイル・システムにおける複数のコンテキストを提供する方法、装置およびコンピュータ・プログラム

(57)【要約】　（修正有）
【課題】リダイレクト・オン・ライト・ファイル・システムにおける複数のコンテキストを提供する。
【解決手段】リダイレクト・オン・ライト・ファイル・システム内の複数のデータ・オブジェクトの現在の一貫性スナップショットを永続的に記憶するためのコミットを開始する。複数のデータ・オブジェクトの各々は、第１のコンテキストを有する、第１のコピーを有する。現在の一貫性スナップショットを記憶するためにコミットしている最中に、複数のデータ・オブジェクトのうちのデータ・オブジェクトに対する更新を受信すると、第１のコピーからデータ・オブジェクトのデータの第２のコピーを作成する。データの第２のコピーは少なくとも２つのコンテキストのうちの第２のコンテキストを有する。データ・オブジェクトに対する更新の受信に応答して、更新に基づきデータ・オブジェクトのデータの第２のコピーを更新する。
【選択図】図２



【特許請求の範囲】
【請求項１】
リダイレクト・オン・ライト・ファイル・システム内の複数のデータ・オブジェクトの現在の一貫性スナップショットを不揮発性機械可読媒体に記憶するためのコミットを開始することであって、前記複数のデータ・オブジェクトの各々は、コミット中のコンテキストを有する、前記複数のデータ・オブジェクトのデータの第１のコピーを有する、前記開始することと、
他の一貫性スナップショットの世代値に対して一意的である世代値を、前記現在の一貫性スナップショットに付与することと、
前記現在の一貫性スナップショットを記憶するために前記コミットしている最中に、前記複数のデータ・オブジェクトのうちのデータ・オブジェクトに対する更新を受信することと、
前記データ・オブジェクトに対する前記更新を受信することに応答して、
前記データ・オブジェクトのための世代値をインクリメントすることと、
前記データ・オブジェクトの前記世代値から導き出される世代値を前記更新に関連付けることと、
前記データ・オブジェクトの前記データの前記第１のコピーからコピーされる、前記データ・オブジェクトのデータの第２のコピーを作成することであって、前記データ・オブジェクトの前記データの前記第２のコピーは更新中のコンテキストを有する、作成することと、
前記データ・オブジェクトの前記データの前記第１のコピーを更新することとは独立に、前記更新に基づき前記データ・オブジェクトの前記データの前記第２のコピーを更新することと、
を含む方法。
【請求項２】
前記方法は、前記データ・オブジェクトに対する前記更新を受信することに応答して、揮発性機械可読媒体内に、前記データ・オブジェクトに関連付けられるバッファ・ヘッダを作成することをさらに含み、前記バッファ・ヘッダは、前記データ・オブジェクトの前記データの前記第１のコピーを指し示す第１のデータ・ポインタおよび前記データ・オブジェクトの前記データの前記第２のコピーを指し示す第２のデータ・ポインタを含む、請求項１に記載の方法。
【請求項３】
前記現在の一貫性スナップショットは、前の一貫性スナップショットより後の、前記複数のデータ・オブジェクトに対する更新を含む、請求項１に記載の方法。
【請求項４】
前記現在の一貫性スナップショットを記憶するためのコミットを前記開始することは、一貫性スナップショットを作成する定期的作業に応答するものである、請求項１に記載の方法。
【請求項５】
前記データ・オブジェクトの前記世代値から導き出される前記世代値を前記更新に関連付けることが、前記データ・オブジェクトの前記世代値に等しい前記世代値を前記更新に関連付けることを含む、請求項１に記載の方法。
【請求項６】
リダイレクト・オン・ライト・ファイル・システム内の複数のデータ・オブジェクトの現在の一貫性スナップショットを永続的に記憶するためのコミットを開始することであって、前記複数のデータ・オブジェクトの各々は、異なるコンテキストを有する、前記複数のデータ・オブジェクトのデータの複数のコピーを有するように構成可能であり、前記複数のデータ・オブジェクトの各々は、少なくとも２つのコンテキストのうちの第１のコンテキストを有する、前記データの少なくとも２つのコピーのうちの第１のコピーを有する、前記開始することと、

前記現在の一貫性スナップショットを記憶するためにコミットしている最中に、前記複数のデータ・オブジェクトのうちのデータ・オブジェクトに対する更新を受信することと、
前記データ・オブジェクトに対する前記更新の受信に応答して、
前記第１のコピーから前記データ・オブジェクトのデータの第２のコピーを作成することであって、前記データの前記第２のコピーは前記少なくとも２つのコンテキストのうちの第２のコンテキストを有する、前記作成すること、および
前記更新に基づき前記データ・オブジェクトの前記データの前記第２のコピーを更新することと、
を含む方法。

【請求項７】
他の一貫性スナップショットの世代値に対して一意的である世代値を、前記現在の一貫性スナップショットに付与することをさらに含む、請求項６に記載の方法。

【請求項８】
前記データ・オブジェクトに対する前記更新の受信に応答して、
前記データ・オブジェクトのための世代値をインクリメントすること、および
前記データ・オブジェクトの前記世代値から導き出される世代値を前記更新に関連付けること、
をさらに含む、請求項７に記載の方法。

【請求項９】
前記現在の一貫性スナップショットは、前の一貫性スナップショットより後の、前記複数のデータ・オブジェクトに対する更新を含む、請求項６に記載の方法。

【請求項１０】
前記現在の一貫性スナップショットを永続的に記憶するためのコミットを前記開始することは、一貫性スナップショットを作成する定期的作業に応答するものである、請求項６に記載の方法。

【請求項１１】
不揮発性機械可読媒体と、
揮発性機械可読媒体と、
プロセッサと、
前記プロセッサ上で実行するように作動可能なファイルセット・マネージャと、
を含む装置であって、
前記ファイルセット・マネージャは、
リダイレクト・オン・ライト・ファイル・システム内の複数のデータ・オブジェクトの現在の一貫性スナップショットを前記不揮発性機械可読媒体内に記憶するためのコミットを開始することであって、前記複数のデータ・オブジェクトの各々は、異なるコンテキストを有する、前記複数のデータ・オブジェクトのデータの複数のコピーを有するように構成可能であり、前記複数のデータ・オブジェクトの各々は、前記異なるコンテキストのうちの第１のコンテキストを有する、前記データの前記複数のコピーのうちの第１のコピーを有し、前記複数のコピーのうちの前記第１のコピーは前記揮発性機械可読媒体内に記憶されるように構成される、前記開始すること、
前記現在の一貫性スナップショットを記憶するためにコミットしている最中に、前記複数のデータ・オブジェクトのうちのデータ・オブジェクトに対する更新を受信すること、ならびに
前記データ・オブジェクトに対する前記更新の受信に応答して、
前記第１のコピーから前記揮発性機械可読媒体内に前記データ・オブジェクトのデータの第２のコピーを作成することであって、前記データの前記第２のコピーは前記異なるコンテキストのうちの第２のコンテキストを有する、前記作成すること、および
前記更新に基づき前記データ・オブジェクトの前記データの前記第２のコピーを更新すること、

を実行するように構成される、
装置。
【請求項１２】
前記ファイルセット・マネージャは、他の一貫性スナップショットの世代値に対して一意的である世代値を、前記現在の一貫性スナップショットに付与するように構成される、請求項１１に記載の装置。
【請求項１３】
前記ファイルセット・マネージャは、前記データ・オブジェクトに対する前記更新の受信に応答して、
前記データ・オブジェクトのための世代値をインクリメントすること、および
前記データ・オブジェクトの前記世代値から導き出される世代値を前記更新に関連付けること、
をするように構成される、請求項１２に記載の装置。
【請求項１４】
前記現在の一貫性スナップショットは、前の一貫性スナップショットより後の、前記複数のデータ・オブジェクトに対する更新を含む、請求項１１に記載の装置。
【請求項１５】
前記現在の一貫性スナップショットを記憶するための前記コミットの前記開始は、一貫性スナップショットを作成する定期的作業に応答するものである、請求項１１に記載の装置。
【請求項１６】
複数のデータ・オブジェクトのうちのデータ・オブジェクトのための複数のコンテキストを提供するコンピュータ・プログラムであって、コンピュータに、
リダイレクト・オン・ライト・ファイル・システム内の前記複数のデータ・オブジェクトの現在の一貫性スナップショットを不揮発性機械可読媒体内に記憶するためのコミットを開始することであって、前記複数のデータ・オブジェクトの各々は、コミット中のコンテキストを有する、前記複数のデータ・オブジェクトのデータの第１のコピーを有する、前記開始すること、
他の一貫性スナップショットの世代値に対して一意的である世代値を、前記現在の一貫性スナップショットに付与すること、
前記現在の一貫性スナップショットを記憶するために前記コミットしている最中に、前記複数のデータ・オブジェクトのうちの前記データ・オブジェクトに対する更新を受信すること、
前記データ・オブジェクトに対する前記更新を受信することに応答して、
前記データ・オブジェクトのための世代値をインクリメントすること、
前記データ・オブジェクトの前記世代値から導き出される世代値を前記更新に関連付けること、
前記データ・オブジェクトの前記データの前記第１のコピーからコピーされる、前記データ・オブジェクトのデータの第２のコピーを作成することであって、前記データ・オブジェクトの前記データの前記第２のコピーは更新中のコンテキストを有する、前記作成すること、および
前記データ・オブジェクトの前記データの前記第１のコピーを更新することとは独立に、前記更新に基づき前記データ・オブジェクトの前記データの前記第２のコピーを更新すること、
を実行させる、コンピュータ・プログラム。
【請求項１７】
前記データ・オブジェクトに対する前記更新の受信に応答して、揮発性機械可読媒体内に、前記データ・オブジェクトに関連付けられるバッファ・ヘッダを作成することであって、前記バッファ・ヘッダは、前記データ・オブジェクトの前記データの前記第１のコピーを指し示す第１のデータ・ポインタおよび前記データ・オブジェクトの前記データの前

記第２のコピーを指し示す第２のデータ・ポインタを含む、前記作成することを前記コンピュータに実行させる、請求項１６に記載のコンピュータ・プログラム。
【請求項１８】
揮発性機械可読媒体内に前記データ・オブジェクトの前記データの前記第１のコピーおよび前記データ・オブジェクトの前記データの前記第２のコピーを作成することを前記コンピュータに実行させる、請求項１６に記載のコンピュータ・プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、概してコンピュータの分野に関し、より具体的には、リダイレクト・オン・ライト・ファイル・システムにおけるデータ・バックアップに関する。
【背景技術】
【０００２】
ファイル・システムは、システム・クラッシュが発生した場合にその内部の一貫性を確実にするために種々の方法を用いる。１つのアプローチは、ファイル・システムが数秒ごとにボトム・アップ順でディスク上の新しい位置に変更データを書き込むことである。その内部に記憶されたデータのこれらのビューは一貫性スナップショットと呼ばれる。システム・クラッシュの後、ファイル・システムは、一貫性があることが保証されている、ファイル・システムの最後の一貫性スナップショットの上部から起動する。
【発明の概要】
【発明が解決しようとする課題】
【０００３】
一貫性スナップショットが書き込まれている間に、ユーザがファイル・システムを新しく改変しようとする可能性があろう。一貫性スナップショットが不揮発性機械可読媒体（例えば、ハード・ディスク）上での記憶のためにコミットされるまで、このような改変をブロックすることは容易であろう。しかし、このアプローチは受け入れられるものではない。なぜなら、このようなアプローチは、一貫性スナップショットがユーザに対して非透過的となってしまうだろうからである。特に、このアプローチは、一貫性スナップショットが記憶のためにコミットされている間、全ファイル・システムを数秒ごとにフリーズさせてしまう可能性がある。
【課題を解決するための手段】
【０００４】
本発明による実施形態には、リダイレクト・オン・ライト・ファイル・システム内の複数のデータ・オブジェクトの現在の一貫性スナップショットを不揮発性機械可読媒体に記憶するためのコミットを開始することであって、複数のデータ・オブジェクトの各々は、コミット中のコンテキストを有する、複数のデータ・オブジェクトのデータの第１のコピーを有する、開始することを含む方法が含まれる。方法は、他の一貫性スナップショットの世代値に対して一意的である世代値を、現在の一貫性スナップショットに付与することも含む。方法は、現在の一貫性スナップショットを記憶するためにコミットしている最中に、複数のデータ・オブジェクトのうちのデータ・オブジェクトに対する更新を受信することを含む。方法は、データ・オブジェクトに対する更新を受信することに応答して、データ・オブジェクトのための世代値をインクリメントすることも含む。方法は、データ・オブジェクトに対する更新を受信することに応答して、データ・オブジェクトの世代値から導き出される世代値を更新に関連付けることを含む。同様に、データ・オブジェクトに対する更新を受信することに応答して、方法は、データ・オブジェクトのデータの第１のコピーからコピーされる、データ・オブジェクトのデータの第２のコピーを作成することを含む。データ・オブジェクトのデータの第２のコピーは更新中のコンテキストを有する。同様に、データ・オブジェクトに対する更新を受信することに応答して、方法は、データ・オブジェクトのデータの第１のコピーを更新することとは独立に、更新に基づきデータ・オブジェクトのデータの第２のコピーを更新することを含む。

【０００５】
本発明による実施形態には、リダイレクト・オン・ライト・ファイル・システム内の複数のデータ・オブジェクトの現在の一貫性スナップショットを永続的に記憶するためのコミットを開始することであって、複数のデータ・オブジェクトの各々は、異なるコンテキストを有する、複数のデータ・オブジェクトのデータの複数のコピーを有するように構成可能である、開始することを含む方法が含まれる。複数のデータ・オブジェクトの各々は、少なくとも２つのコンテキストのうちの第１のコンテキストを有する、データの少なくとも２つのコピーのうちの第１のコピーを有する。方法は、現在の一貫性スナップショットを記憶するためにコミットしている最中に、複数のデータ・オブジェクトのうちのデータ・オブジェクトに対する更新を受信することを含む。データ・オブジェクトに対する更新の受信に応答して、方法は、第１のコピーからデータ・オブジェクトのデータの第２のコピーを作成することを含む。データの第２のコピーは少なくとも２つのコンテキストのうちの第２のコンテキストを有する。データ・オブジェクトに対する更新の受信に応答して、方法は、更新に基づきデータ・オブジェクトのデータの第２のコピーを更新することも含む。

【図面の簡単な説明】

【０００６】
【図１】いくつかの実施形態例による、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供するクラスタ化ファイル・システム構成の概念図である。
【図２】いくつかの実施形態例による、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供するクラスタ化ファイル・システム構成のより詳細な概念図である。
【図３】いくつかの実施形態による、クラスタ化ファイル・システム内に記憶されるデータ・オブジェクトのためのバッファ・ヘッダの例を示す図である。
【図４】いくつかの実施形態による、データ・オブジェクトの複数の世代に対して一貫性スナップショットをコミットするタイムラインの例を示す図である。
【図５】いくつかの実施形態例による、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供する作業のフローチャートである。
【図６】いくつかの実施形態例による、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供する作業のフローチャートである。
【図７】コンピュータ・システムの例を示す図である。

【発明を実施するための形態】

【０００７】
添付の図面を参照することによって、本実施形態はより良く理解され、数多くの対象、特徴および利点が当業者に明らかになればよい。

【０００８】
以下に続く記載は、本発明の主題の技法を具体化するシステム、方法、技法、命令シーケンスおよびコンピュータ・プログラムの例を含む。しかし、記載されている実施形態は、これらの具体的詳細を備えずに実施されてもよいことは理解されよう。例えば、例が、ファイル・システムの一部であるデータのためのデュアル・コンテキストに言及していても、いくつかの他の実施形態例はデータのためのコンテキストをいくつでも（例えば、３つ、４つ、５つ等）構成することができる。同様に、ファイル・システムのための一貫性スナップショットが作成されるように記載される一方で、いくつかの他の実施形態例では、他のレベルにおける一貫性スナップショットが作成されることができる。例えば、ユーザは、ファイルのサブセット、特定のファイル等を、ファイル・システムのための定期スナップショットよりも頻繁に一貫性スナップショット内にバックアップされるように構成することができる。それ故、いくつかの実施形態例はこれらの他のレベルの一貫性スナッ

プショットに適用できる。他の例では、記載を分かりにくくしないようにするために、周知の命令インスタンス、プロトコル、構造および技法は詳細に示されていない。

【０００９】

多重コンピュータ・システムまたはノード、ならびに永続性記憶資源を含む資源からクラスタが形成される。クラスタの記憶資源にわたってクラスタ化ファイル・システムが実装される。クラスタ記憶資源は、クラスタのノードによる直接アクセスを許すように結合される。記憶資源はノードに直接ケーブルでつながれるか、もしくはネットワーク（例えば、ストレージ・エリア・ネットワーク）を介してアクセス可能とされるか、またはその両方の方法が用いられることができる。

【００１０】

クラスタが確立されると、管理者が、クラスタのノードのうちの１つをクラスタ・リーダとして動作するように構成する。実施形態は、リーダを自動的に選ぶようにクラスタをプログラムすることもできる。クラスタ・リーダは、ノードがクライアントなのかサーバなのか、それともクライアントとサーバの両方なのかを示すクラスタ役割データを保持する。サーバがクラスタ化ファイル・システム内のファイルセットを管理する。クラスタ・リーダは、どのノードがクラスタ化ファイル・システム・マネージャとして動作するかの指示も保持する。クラスタ化ファイル・システム・マネージャはクラスタ化ファイル・システムのためのメタデータを管理する。実施形態によっては、クラスタ化ファイル・システム・マネージャはクラスタのための唯一のサーバである－フェイルオーバー・サーバの責任は負っていない。実施形態によっては、クラスタ化ファイル・システム・マネージャはクラスタ化ファイル・システム内のファイルセットの管理を、サーバである他のノードに任せる。本願明細書において、用語「ファイルセット」はファイルもしくはディレクトリのセットまたはその両方のセットに言及するために用いられる。どのノードがクラスタ内のサーバであるかの指示とともに、クラスタ・リーダは、サーバまたは「ファイルセット・マネージャ」によって管理されるファイルセットの指示を保持することができる。クラスタ内のノードがクラスタ・リーダおよびクラスタ化ファイル・システム・マネージャとして動作するように構成されることができる。ノードがクラスタ・リーダとして動作するのか、サーバとして動作するのか、クライアントとして動作するのか、その他のものとして動作するのかは、クラスタのユーザに透過的であることができる。ノードがクライアントおよびサーバの両方として動作しても、あるいはクライアントが遠隔ノード上にあっても、ユーザは同じ動作を知覚することになる。

【００１１】

クラスタ化ファイル・システム・マネージャは、クラスタ化ファイル・システムのファイルのためのｉノードの階層としてメタデータを保持することができる。クラスタ化ファイル・システム・メタデータは、クラスタ化された記憶資源の記憶の論理ユニットに関する情報を示す。情報はクラスタ記憶ユニットの位置（例えば、オフセットまたはブロック番号）ならびにエクステントの長さを含むことができる。本記載において、用語「ブロック」は、クラスタ記憶の単位に言及するために用いられる（例えば、４ＫＢブロック）。本記載は、隣接ブロックのセットに言及するために用語「エクステント」も用いる。エクステントの「長さ」に言及するとき、該長さは、エクステントを形成する多数の隣接ブロックに言及している。これらの用語を用いると、４ＫＢブロックを仮定すれば、クラスタ化ファイル・システムは、総計１０ＧＢの記憶資源のプールを０ないし２，６２１，４３９ブロックと見なす。クラスタ・クライアントがクラスタ記憶の論理ユニットに書き込みをするとき、書き込みを果たすために論理ユニット（例えば、ブロック番号）は記憶仮想化層によって物理的位置（例えば、シークおよびオフセット）に変換される。実施形態はブロックおよびエクステントに限定されるものではないが、クラスタ記憶の単位（例えば、可変長ブロック、ビット長等）についてあり得るあらゆる実装を考慮すれば、記載を分かりにくくしてしまうことになろう。

【００１２】

実施形態によっては、クラスタ化ファイル・システム・マネージャはクラスタ化ファイ

ル・システム・メタデータ（「メタデータ」）をｉノードの階層データ構造内に保持する。クラスタ化ファイル・システム・マネージャはメタデータのためのルートをクラスタ記憶資源（「クラスタ記憶」）内の既知の位置（すなわち、所定位置）に保持する。一貫性スナップショットを支援するクラスタ内では、一貫性スナップショットのルートを、対応する一貫性スナップショットのルート・メタデータとともに記憶するために、クラスタ記憶内の複数の位置が確保または定義される。ルート・メタデータは、一貫性スナップショットを識別すること、および一貫性スナップショットの保全性を確実にすることを助ける。実施形態は、一貫性スナップショットの進行を追跡する、一貫性スナップショットの時間ベースの識別子（例えば、世代値）、およびデータの保全性を検査するためのルート・チェックサムを用いることができる。実施形態は、ノードがルートを書き込み始める際、第１のルート・チェックサム（「ヘッダ・チェックサム」）を書き込み、ルートが永続性クラスタ記憶にうまく書き込まれた後、第２のルート・チェックサム（「トレーラ・チェックサム」）を書き込むことができる。実施形態はヘッダ・チェックサムおよびトレーラ・チェックサムを用いて、一貫性スナップショットのルートの書き込みが妨げられなかったことを確実にすることができる。障害から回復するためには、位置の各々が調べられ、最新の世代値を持つ位置が選択され、選択された位置によって参照されるその一貫性スナップショットから回復が始まることを可能にする。実施形態は、一貫性スナップショットをいくつでも保存するようにクラスタを構成することができる。

【００１３】

いくつかの実施形態例は所与のファイル・システム内のデータの一貫性スナップショットを提供するが、このようなスナップショットは、一貫性スナップショットが記憶のためにコミットされている間、入ってくるファイル・システム・トランザクションをブロックしたりまたは遅らせたりしない。それ故、ファイル・システム内に記憶されているデータに対する更新が、同じファイル・システムの一貫性スナップショットの記憶と同時に行われることができる。以下においてさらに記載されるように、この同時実行が可能とされるために、同じデータ・オブジェクトのための少なくとも２つのコンテキストが保持される。

【００１４】

ファイル・システムの一貫性スナップショットは、一意の世代値に関連付けられる。例えば、世代値は整数値とすることができる。それ故、一貫性スナップショットを記憶するためのコミット（すなわち、該一貫性スナップショットの同期）が始められると、ファイル・システムのための世代はインクリメントされることができる。

【００１５】

実施形態例によっては、ファイル・システム内のオブジェクト（例えば、データ、ファイル等）に対するあらゆる変更がトランザクションに関連付けられる。トランザクションはファイル・システムの世代に関連付けられ、かくして一貫性スナップショットに関連付けられる。デュアル・コンテキスト構成の実施形態例によっては、ファイル・システム内のオブジェクトが、常に、オブジェクトの最大で２つのコピーとともにキャッシュされる。１つのコピーは、更新中のコンテキストのためのものである。特に、オブジェクトのための更新中のコンテキストは、オブジェクトが更新されている最中であった（例えば、ユーザがオブジェクトを更新した）後、オブジェクトを有するファイル・システムの一貫性スナップショットが記憶のためにコミットされている間に作成される。オブジェクトの第２のコピーは、コミット中のコンテキストのためのものである。オブジェクトのこのコピーは、一貫性スナップショットの一部として記憶のためにコミットされている最中の／コミットされることになるオブジェクトのコピーである。実施形態例によっては、２つのオブジェクトは、サイズ２の配列として一緒にキャッシュされることができる。同様に、オブジェクトは、配列内の各オブジェクトに関連付けられる世代を記憶する、２つの要素の配列を有する。

【００１６】

実施形態例によっては、オブジェクトのどのコンテキストを用いるべきかを決定するた

めに、トランザクションに関連付けられる世代値または世代番号が、キャッシュされているオブジェクトの世代番号と比較される。正しいオブジェクト配列要素が選択され、オブジェクトの正しいコンテキスト（例えば、更新中のコンテキストまたはコミット中のコンテキスト）を変更する。

【００１７】

以下においてさらに記載されるように、一貫性スナップショットは定期的に（例えば、５秒ごとに）作成される。これらの一貫性スナップショットは、以前のバージョンのオブジェクトを回復しようと試みるために作成される。例えば、これらの一貫性スナップショットは、システム・クラッシュの後に、ファイル・システム内に記憶されているオブジェクトを回復するために用いられることができる。実施形態例によっては、以前の一貫性スナップショットがその同期（記憶のためのコミット）を終える前に、一貫性スナップショットの間隔が到達されると、新しい一貫性スナップショットはスキップされる。特に、そのときは、キャッシュされているオブジェクトの第３のコピーが必要となろうから、一貫性スナップショットは取られないであろう。

【００１８】

図１は、いくつかの実施形態例による、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供するクラスタ化ファイル・システム構成の概念図を示す。図示のクラスタはノード１０３、１０５、１０７、１０９を含む。クラスタは、直接アクセス可能記憶デバイスのプール１０１、ネットワーク・アクセス可能記憶デバイス１１３、１１５、およびネットワーク・インフラストラクチャ１１１も含む。ノード１０３、１０５、１０７、１０９はネットワーク・インフラストラクチャ１１１を介して通信する。ノード１０３、１０５、１０７、１０９はケーブルを介して記憶デバイス・プール１０１にアクセスし、ネットワーク・インフラストラクチャ１１１を介してネットワーク・アクセス可能記憶デバイス１１３、１１５にアクセスする。図示のクラスタにおいて、ノード１０３、１０５、１０７、１０９のうちのいずれでもクラスタのためのクラスタ化ファイル・システム・マネージャとして構成されることができる。クラスタ化ファイル・システム・マネージャはその内部のクラスタ化ファイル・システムのファイルの記憶の様々な側面を管理することができる。例えば、クラスタ化ファイル・システム・マネージャは、クラスタ化ファイル・システムのファイルのためのｉノードの階層としてメタデータを保持することができる。実施形態例によっては、クラスタ化ファイル・システム・マネージャの作業のうちの一部またはすべては、異なるノード１０３、１０５、１０７、１０９に振り分けられることができる。これらの作業のうちのいくつかは、（以下でさらに記載されるように）クラスタ化ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供することに関連する作業を含む。以下においては、データ・オブジェクトのための複数のコンテキストを提供するこれらの作業は異なるノード１０３、１０５、１０７、１０９にわたって振り分けられるように記載されているが、一方、いくつかの他の実施形態例では、このような作業はクラスタ化ファイル・システム・マネージャによって遂行されることができる。

【００１９】

図２は、いくつかの実施形態例による、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供するクラスタ化ファイル・システム構成のより詳細な概念図を示す。図２は、図１のノード１０３、１０５、１０７、１０９を代表することができるノードＡ　２０２、ノードＢ　２０４およびノードＮ　２０６を含むシステム２００を示す。図２はノードＡ　２０２内の多数の構成要素を示す。図示されていないが、ノードＢ　２０４およびノードＮ　２０６はそれらの内部に同様の構成要素を含むことができる。

【００２０】

実施形態例によっては、システム２００は、データが変更される際、リダイレクト・オン・ライト（ｒｅｄｉｒｅｃｔ－ｏｎ－ｗｒｉｔｅ、ＲＯＷ）を用いるファイル・システムのデータ・オブジェクトを記憶するように構成される。特に、リダイレクト・オン・ラ

イトを用いると、変更データのために新しいブロックが割り付けられる。ファイル・システムは１つ以上のファイルセットを含むことができる。実施形態例によっては、ファイル・システム内の各ファイルはｉノードを含むことができる。ｉノードは、ファイル内に記憶されるデータに関する情報またはメタデータを記憶する別個のファイルまたはデータ構造であることができる。例えば、ファイルの部分（例えば、ブロック）ごとに、ｉノードは、このデータが記憶されるファイルセットのアドレス、ファイルセット識別情報および世代を記憶することができる。特に、ファイルのデータが記憶されるブロックは、異なるファイルセット、およびファイルセットの世代にわたって振り分けられることができる。異なるファイルセット、およびファイルセットの世代は複数の記憶デバイスにわたって振り分けられることができる。図２を参照すると、これらのファイルセットはノードＡ　２０２、ノードＢ　２０４およびノードＮ　２０６のうちのいずれかにおける機械可読媒体内に記憶されることができる。

【００２１】

システム２００は多数のクライアント・デバイス（クライアント・デバイス２０８およびクライアント・デバイス２１０として図示）を含む。システム２００はネットワーク２１２を含み、ノードＡ　２０２、ノードＢ　２０４、ノードＮ　２０６、クライアント・デバイス２０８およびクライアント・デバイス２１０はネットワーク２１２を通じて通信可能に共に結合される。

【００２２】

ノードＡ　２０２は、通信可能に共に結合された、ファイルセット・マネージャ２１４、不揮発性機械可読媒体２１６、およびメモリ（例えば、揮発性機械可読媒体）２１８を含む。ファイルセット・マネージャ２１４はソフトウェア、ファームウェア、ハードウェアまたはそれらの組み合わせであることができる。例えば、ファイルセット・マネージャ２１４は、ノードＡ　２０２内のプロセッサ（不図示）上で実行するオペレーティング・システムの一部であることができる。不揮発性機械可読媒体２１６は、すでに作成された多数の一貫性スナップショット（一貫性スナップショットＡ　２２４および一貫性スナップショットＮ　２２６として図示）を記憶する。不揮発性機械可読媒体２１６は、その内部に記憶するためにコミットされている途中である現在の一貫性スナップショット２２８も記憶している最中である。実施形態例によっては、一貫性スナップショットは定期的に（例えば５秒ごとに）作成される。一貫性スナップショットは、所与の時点におけるファイル・システム内のデータ・オブジェクトのスナップショットを含む。実施形態例によっては、一貫性スナップショットは、最後の一貫性スナップショットが記憶のためにコミットされてより後、不揮発性機械可読媒体２１６に記憶にするためにまだコミットされていない、メモリ２１８内にあるデータ・オブジェクトに対するあらゆる変更（例えば、改変、追加、削除等）を記憶する。これらの一貫性スナップショットは、ファイル・システム内に記憶されている、以前のバージョンのオブジェクトを回復しようと試みるために作成される。例えば、これらの一貫性スナップショットは、システム・クラッシュの後に、ファイル・システム内に記憶されているオブジェクトを回復するために用いられることができる。

【００２３】

メモリ２１８は多数のバッファ・ヘッダ（バッファ・ヘッダＡ　２２０、バッファ・ヘッダＮ　２２２等）を記憶する。以下においてさらに記載されるように（図３の記載を参照）、バッファ・ヘッダは、ファイル・システム内に記憶されるデータ・オブジェクトに関する様々なメタデータを記憶する。データ・オブジェクトがアクセスされている、変更されている等の最中であれば、ファイルセット・マネージャ２１４はメモリ２１８内のデータ・オブジェクトのためのバッファ・ヘッダを作成する（その中にまだ作成されていなければ）。例えば、ファイルセット・マネージャ２１４は、現在の一貫性スナップショット２２８を作成するためにデータ・オブジェクトがアクセスされている最中、何らかのクライアント・デバイス要求に基づきデータ・オブジェクトが変更されている最中などに、バッファ・ヘッダを作成することができる。メモリ２１８のサイズおよびアクセスされて

いるデータ・オブジェクトの数に基づき、ファイルセット・マネージャ２１４は、関連データ・オブジェクトがアクセスされている最中でないバッファ・ヘッダのうちのいくつかをフラッシュするよう求められてよい。それに応じて、ファイルセット・マネージャ２１４は、データ・オブジェクトのアクセスが生じると、メモリ２１８内のデータ・オブジェクトのためのバッファ・ヘッダを再作成するよう求められてよい。以下においてさらに記載されるように、バッファ・ヘッダ内のメタデータは、所与のデータ・オブジェクトのために作成されるデータの異なるコピーのためのデータ・ポインタを記憶する。本例では、バッファ・ヘッダＡ　２２０は、データの第１のコピー２５０を指し示す第１のデータ・ポインタ、およびデータの第２のコピー２５２を指し示す第２のデータ・ポインタを有する。メモリ２１８内に記憶される異なるバッファ・ヘッダのために同様のデータ・ポインタが作成されることができる。

【００２４】

実施形態例によっては、ファイル・システム内の同じデータ・オブジェクトのためのデータの複数のコピーが作成される。データの複数のコピーの各々は、異なるコンテキストに関連付けられることができる。実施形態例によっては、データ・オブジェクトが、デュアル・コンテキスト構成のためにそのデータのコピーを２つ有することができる。例として、メモリ２１８は同じデータ・オブジェクトのためのデータの２つのコピー（データの第１のコピー２５０およびデータの第２のコピー２５２）を記憶する。ファイル・システム内に記憶されるいずれかまたはすべてのデータ・オブジェクトはこのマルチ・コピー、マルチ・コンテキスト構成を含むことができる。図示のように、データの第１のコピー２５０はコミット中のコンテキスト２５４を有し、データの第２のコピー２５２は更新中のコンテキスト２５６を有する。同じデータ・オブジェクトのための２つのコンテキストはファイル・システム内のデータの一貫性スナップショットを提供するが、このようなスナップショットは、一貫性スナップショットが記憶のためにコミットされている間、入ってくるファイル・システム・トランザクションをブロックしたりまたは遅らせたりしない。それ故、ファイル・システム内に記憶されているデータに対する更新が、同じファイル・システムの一貫性スナップショットの記憶と同時に行われることができる。具体的には、コミット中のコンテキスト２５４は、現在の一貫性スナップショット２２８内にこの特定のデータ・オブジェクトを作成するために用いられるデータのコピーに関連付けられる。更新中のコンテキスト２５６は、現在の一貫性スナップショット２２８が、（不揮発性機械可読媒体２１６内に作成される）記憶のためにコミットされている間、データ・オブジェクトに対する更新（例えば、ユーザがデータに変更を行っている）を受け付けるために用いられるデータのコピーに関連付けられる。

【００２５】

図２は多数の作業（作業２３０、作業２３２および作業２３４）も示す。本例では、ファイルセット・マネージャ２１４は作業２３０を遂行し、該作業２３０において、ファイルセット・マネージャ２１４は現在の一貫性スナップショットを記憶するためのコミットを開始する。特に、ファイルセット・マネージャ２１４は現在の一貫性スナップショット２２８の作成を開始する。作業の一部として、ファイルセット・マネージャ２１４は、以前の一貫性スナップショットが記憶のためにコミットされて以来、どのようなデータ・オブジェクトが変更されたのかを判定することができる。次に、ファイルセット・マネージャ２１４は、変更されたデータ・オブジェクトを不揮発性機械可読媒体２１６内の新しい位置にボトム・アップ順で書き込むことができる。実施形態例によっては、現在の一貫性スナップショット２２８内に記憶されている最中のデータ・オブジェクトごとに、ファイルセット・マネージャ２１４はメモリ２１８内の関連バッファ・ヘッダを作成もしくは更新するまたはその両方を行うことができる（作業２３４として図示）。メモリ２１８内にデータ・オブジェクトのための関連バッファ・ヘッダがない場合、ファイルセット・マネージャ２１４は、このようなデータが、現在の一貫性スナップショット２２８内に記憶するためにアクセスされている最中に、バッファ・ヘッダを作成する。以下において図４～６を参照してさらに記載されるように、各データ・オブジェクトのためのバッファ・ヘッ

ダは様々なメタデータ（例えば、世代、コンテキスト、位置、データ・ポインタ）を含む。ファイルセット・マネージャ２１４は、メモリ２１８内にバッファ・ヘッダを作成することの一部としてこのメタデータを更新する。代替的に、バッファ・ヘッダが所与のデータ・オブジェクトのためにメモリ２１８内にすでにインスタンス化されている場合には、ファイルセット・マネージャ２１４はその中のメタデータを更新することができる。例えば、ファイルセット・マネージャ２１４は、データ・ポインタによって参照されている複数のデータのための世代およびコンテキストを定義する様々なフィールドを更新することができる（以下においてさらに記載されるように）。

【００２６】

同様に、現在の一貫性スナップショット２２８を記憶するためのコミットの完了より前に、現在の一貫性スナップショット２２８内に含まれるべきデータ・オブジェクトが変更される。本例では、クライアント・デバイス２１０が、現在の一貫性スナップショット２２８の一部であるデータ・オブジェクトのための更新要求を、ネットワーク２１２を通じて送信し、該更新要求はファイルセット・マネージャ２１４によって受信される（作業２３２として図示）。この状況において、ファイルセット・マネージャ２１４は、データの第１のコピーからコピーされる、データ・オブジェクト内のデータの第２のコピーを作成する（例えばデータの第１のコピー２５０およびデータの第２のコピー２５２を参照）。同様に、データの第２のコピーは、データの第１のコピーのために定義されるコンテキストとは別個且つ異なるコンテキストを有する。実施形態例によっては、データの第２のコピーは、デュアル・コンテキストを提供するために第２のコピーが必要になるまでは作成されない。例えば、同じデータ・オブジェクトを記憶する一貫性スナップショットが作成されている途中である間は、ファイルセット・マネージャ２１４は、データ・オブジェクトに対する更新が要求されるまで第２のコピーを作成しない。同様に、ファイルセット・マネージャ２１４はメモリ２１８内のこのデータ・オブジェクトのためのバッファ・ヘッダを作成もしくは更新するまたはその両方を行う。例えば、ファイルセット・マネージャ２１４は、データの第２のコピーを指し示すためにバッファ・ヘッダ内の第２のデータ・ポインタを更新することができる。同様に、ファイルセット・マネージャ２１４は、データの２つの異なるコピーが２つの異なるコンテキストを有するように、コンテキストを更新する。ファイルセット・マネージャ２１４の、データ・オブジェクトのための複数のコンテキストを提供する作業のより詳細な記載が、以下において図５～６のフローチャートを参照して説明される。

【００２７】

図３は、いくつかの実施形態による、クラスタ化ファイル・システム内に記憶されるデータ・オブジェクトのためのバッファ・ヘッダの例を示す。バッファ・ヘッダ３００は、クラスタ化ファイル・システム内に記憶されるデータ・オブジェクトに関連する多数のフィールドを含む。上述されたように、データ・オブジェクトのためのバッファ・ヘッダが、まだそれがメモリ内にない場合に且つ、データ・オブジェクトへのアクセスに応答して、メモリ内に作成される。例えば、ファイルセット・マネージャ２１４は、一貫性スナップショット内のデータ・オブジェクトを記憶するためにデータ・オブジェクトにアクセスすることができる。別の例では、ファイルセット・マネージャ２１４は、データ・オブジェクトを更新する何らかのアプリケーション（例えば、クライアント・デバイス２０８、２１０）に応答してデータ・オブジェクトにアクセスすることができる。バッファ・ヘッダを作成することに加えて、ファイルセット・マネージャ２１４はその内部のフィールド（３０２～３１６）にデータを格納することもできる。フィールド３０２～３０４はこのデータ・オブジェクトのための２つの異なる世代値を定義する。最後にコミットされた世代（Ｌａｓｔ　Ｃｏｍｍｉｔｔｅｄ　Ｇｅｎｅｒａｔｉｏｎ、ＬＣＧ）フィールド３０２は、このデータ・オブジェクトが一貫性スナップショット内の記憶のためにコミットされた最後の時のこのデータ・オブジェクトのための世代値を定義する。最後に更新された世代（Ｌａｓｔ　Ｕｐｄａｔｅｄ　Ｇｅｎｅｒａｔｉｏｎ、ＬＵＧ）フィールド３０４は、このデータ・オブジェクトが更新されていた最後の時のこのデータ・オブジェクトのため

の世代値を定義する。データ・オブジェクトの世代値は、データ・オブジェクトが最初に更新される度に、ただし、データ・オブジェクトが一貫性スナップショットの一部として永続的記憶のためにコミットされるより前に、インクリメントされる。例えば、データ・オブジェクトの現在の世代値が１５であるとする。データ・オブジェクトが一貫性スナップショットの一部として永続的記憶のためにコミットされた後に何らかのアプリケーションがデータ・オブジェクトを更新しようとすると、世代値は１６にインクリメントされる。このデータ・オブジェクトのこの世代値は、データ・オブジェクトが一貫性スナップショットの一部として永続的記憶のためにコミットされるまで、１６にとどまる。

【００２８】

フィールド３０６～３０８はこのデータ・オブジェクトのための２つの異なるコンテキスト値を定義する。これらのコンテキスト値は０または１のいずれかに設定される。特に、データ・オブジェクトのためのコンテキストは２つの値の間で反転する（デュアル・コンテキストの一部であるため）。最後にコミットされたコンテキスト（Ｌａｓｔ Ｃｏｍｍｉｔｔｅｄ Ｃｏｎｔｅｘｔ、ＬＣＸ）フィールド３０６は、このデータ・オブジェクトが一貫性スナップショット内の記憶のためにコミットされた最後の時のこのデータ・オブジェクトのためのコンテキストを定義する。最後に更新されたコンテキスト（Ｌａｓｔ Ｕｐｄａｔｅｄ Ｃｏｎｔｅｘｔ、ＬＵＸ）フィールド３０８は、このデータ・オブジェクトが更新されていた最後の時のこのデータ・オブジェクトのためのコンテキストを定義する。例えば、データ・オブジェクトが一貫性スナップショットの一部として永続的記憶のためにコミットされた後であるが、データ・オブジェクトに対する更新よりも前において、ＬＣＸフィールド３０６およびＬＵＸ３０８は両方とも同じ値（例えば１）に設定される。その後、何らかのアプリケーションがデータ・オブジェクトを更新しようとすると、ＬＵＸフィールド３０８は０の値に反転される。その後、このデータ・オブジェクトが再び一貫性スナップショットの一部として永続的記憶のためにコミットされる際、ＬＣＸフィールド３０６は０の値に反転される。フィールド３０２～３０８の使用は、以下において図５～６のフローチャートを参照してさらに記載される。

【００２９】

物理的位置フィールド３１０はファイル・システム内のデータ・オブジェクトの物理的位置（例えばブロック番号）を定義する。論理的位置フィールド３１２は、データ・オブジェクトが記憶される論理的位置を、このデータ・オブジェクトのための関連ｉノードの位置に基づき定義する。例えば、論理的位置は、このデータ・オブジェクトが記憶される、ｉノードに加えていくらかのオフセットの物理的位置を含むことができる。

【００３０】

データ・ポインタ０・フィールド３１４は、メモリ２１８内のデータ・オブジェクトのデータの第１のコピーを指し示している第１のデータ・ポインタ（データ・ポインタ０）を記憶する。データ・ポインタ１・フィールド３１６は、メモリ２１８内のデータ・オブジェクトのデータの第２のコピーを指し示している第２のデータ・ポインタ（データ・ポインタ１）を記憶する。上述されたように、データ・オブジェクトのデータの第２のコピーは、データ・オブジェクトのための第２のコンテキストが必要となるまで作成されない。例えば、データ・オブジェクトが一貫性スナップショットの一部として永続的記憶のためにコミットされた後であるが、データ・オブジェクトに対するいかなるその後の更新よりも前においては、データ・オブジェクトのデータのコピーは１つだけ提供されることができる。この状況において、データ・ポインタ０・フィールド３１４（データの第１のコピーを指し示す）はデータの第１のコピーを指し示し、データ・ポインタ１・フィールド３１６（データの第２のコピーを指し示す）は位置を指し示さない（例えばヌル（ＮＵＬＬ））。データの第２のコピーは、データ・オブジェクトのために第２のコンテキストが必要になった後、データの第１のコピーのコピーから作成される。例えば、データ・オブジェクトが一貫性スナップショット内に記憶されている最中であり、同時に、クライアント・デバイスがデータ・オブジェクトに対する更新を要求しているとする。この状況において、データ・オブジェクトの第２のコピーが作成される。同様に、データ・ポインタ０

・フィールド３１４（データの第１のコピーを指し示す）はなおデータの第１のコピーを指し示し、データ・ポインタ１・フィールド３１６（データの第２のコピーを指し示す）は、今度は、データ・オブジェクトのデータの第２のコピーを指し示すように変更される。フィールド３１４～３１６の使用は以下において図５～６のフローチャートを参照してさらに記載される。

【００３１】

図４は、いくつかの実施形態による、データ・オブジェクトの複数の世代に対して一貫性スナップショットをコミットするタイムラインの例を示す。タイムライン４００は左から右へ時間が増大する。時点４０２は、データ・オブジェクトのための世代Ｎが終了した時刻である。時点４０４は、同じデータ・オブジェクトのためのより後の世代（世代Ｎ＋１）が終了した、より後の時刻である。時点４０６は、同じデータ・オブジェクトのためのより後の世代（世代Ｎ＋２）が終了した、より後の時刻である。期間４０８は、（データ・オブジェクトを含む）一貫性スナップショットが永続的記憶のためにコミットされている期間である。期間４０８は、世代Ｎが終了した後の時点４０２において開始される。コミットの一部として上述されたように、ファイルセット・マネージャ２１４はデータ・オブジェクトの階層をボトム・アップ順で横断し、子データ・オブジェクトのブロック番号およびチェックサムを収集する。図示のように、期間４０８内には２つのサブ期間－期間４１０および期間４１２－がある。期間４１０は、データ・オブジェクトの１つのコピーまたはバージョンがメモリ内に存在する期間を含む。例えば、この期間は、データ・オブジェクトが永続的記憶のためにコミットされている最中の時間であって、データ・オブジェクトはまだ変更されていない（例えば、クライアント・デバイス上で実行するアプリケーションによって）、時間を含むことができる。期間４１２は、データ・オブジェクトの２つのコピーまたはバージョンがメモリ内に存在する期間を含む。期間４１２は、世代Ｎのための一貫性スナップショットのコミットが依然、行われている間にデータ・オブジェクトが変更されるのに応答して開始される。例えば、この期間は、データ・オブジェクトが永続的記憶のためにコミットされている最中且つ、データ・オブジェクトが変更されている（例えば、クライアント・デバイス上で実行するアプリケーションによって）最中の時間を含むことができる。換言すると、データ・オブジェクトの第１のバージョンは、発行されている最中の世代Ｎの一貫性スナップショットの一部として存在する。データ・オブジェクトの第２のバージョンは、世代Ｎの一貫性スナップショットの発行の完了より前に現在の世代Ｎ＋１内でデータ・オブジェクトへの書き込みが行われる場合に備えてもしくはそれが行われる結果、またはその両方の故に存在する。

【００３２】

図５～６は、いくつかの実施形態例による、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供する作業のフローチャートを示す。図５はフローチャート５００を示し、図６はフローチャート６００を示す。フローチャート６００はフローチャート５００の続きであり、ポイントＡにおいて移行する。フローチャート５００～６００は、分散型構成で行われるものとして記載されており、そこでは、ファイルセット・マネージャ２１４がその内部の作業を遂行する。他の実施形態例によっては、フローチャート５００～６００の作業は集中型構成で行われ、そこでは、ファイル・システム・マネージャがこのような作業を遂行することができる。フローチャート５００～６００は、データ・オブジェクトのためのデュアル・コンテキストが必要とされる状況の例を示す。特に、この状況例では、特定のデータ・オブジェクト（「データ・オブジェクトＡ」と呼ばれる）を含む一貫性スナップショットが不揮発性機械可読媒体内の記憶のためにコミットされている最中となっている。なぜなら、データ・オブジェクトＡは、前の一貫性スナップショットが記憶のためにコミットされてより後に、変更されているからである。この一貫性スナップショットが記憶のためにコミットされるのと同時に、データ・オブジェクトＡをさらに変更する作業がある。例えば、クライアント・デバイス上で実行するアプリケーションがデータ・オブジェクトＡを変更することができる。図１～３を参照してフローチャート５００～６００の作業が記載される。フロー

チャート５００がまず記載され、その後にフローチャート６００の記載が続く。

【００３３】

ファイルセット・マネージャ２１４が、ファイル・システム内の多数のデータ・オブジェクトを含む現在の一貫性スナップショットを不揮発性機械可読媒体に記憶するためのコミットを開始する（５０２）。実施形態例によっては、ファイルセット・マネージャ２１４は現在の一貫性スナップショットを記憶するために定期的にコミットする（例えば、３秒、５秒、１０秒等）。それ故、この作業は一貫性スナップショットを作成する定期的作業のうちの１つであることができる。図２を参照すると、ファイルセット・マネージャ２１４は現在の一貫性スナップショット２２８を記憶するためのコミットを開始する。実施形態例によっては、現在の一貫性スナップショット２２８は、前の一貫性スナップショットより後に変更されたデータ・オブジェクトを含むことになる。データ・オブジェクトに対するそれらの改変はメモリ２１８内に常駐することができ、そのため、改変は不揮発性機械可読媒体２１６内の記憶のためにまだコミットされていない。フローチャート５００の作業は５０４において続く。

【００３４】

ファイルセット・マネージャ２１４は、現在の一貫性スナップショット内に記憶されるべきデータ・オブジェクトのためのバッファ・ヘッダがメモリ内にあるかどうかを判定する（５０４）。図２を参照すると、ファイルセット・マネージャ２１４は、現在の一貫性スナップショット２２８内に記憶されるべきデータ・オブジェクトのためのバッファ・ヘッダがメモリ２１８内にあるかどうかを判定する。特に、実施形態例によっては、データ・オブジェクトがアクセスされる（読み出される、書き込まれる等）度に、メモリ２１８内に関連バッファ・ヘッダが作成される。現在の一貫性スナップショット２２８内に記憶されるべきデータ・オブジェクトごとのバッファ・ヘッダがメモリ内にすでにある場合は、フローチャート５００の作業は５０８において続く。さもなければ、フローチャート５００の作業は５０６において続く。

【００３５】

ファイルセット・マネージャ２１４は、（メモリ内にバッファ・ヘッダをまだ有していないデータ・オブジェクトのために）メモリ内にバッファ・ヘッダを作成し更新する（５０６）。図２を参照すると、ファイルセット・マネージャ２１４は、メモリ内にバッファ・ヘッダを有していないこれらのデータ・オブジェクトのためにメモリ２１８内にバッファ・ヘッダを作成する。ファイルセット・マネージャ２１４はバッファ・ヘッダのフィールドを更新することもできる。図３を参照すると、ファイルセット・マネージャ２１４はこれらのデータ・オブジェクトごとのバッファ・ヘッダのためのこれらのフィールドの値を設定する。ファイルセット・マネージャ２１４はＬＣＧフィールド３０２およびＬＵＧフィールド３０４の両方をデータ・オブジェクトのための現在の世代値に設定する。例えば、もし最後にコミットされた一貫性スナップショットが５の値を有していれば、ファイルセット・マネージャ２１４はＬＣＧフィールド３０２およびＬＵＧフィールド３０４を５に設定することになろう。コンテキスト・フィールド（３０６、３０８）は、２つのコンテキスト（コミット中のコンテキストおよび更新中のコンテキスト）を区別するために０または１のいずれかに設定される。従って、第２のコンテキストが必要とされる場合は、これらの２つのコンテキスト・フィールド３０６、３０８は逆の値を有することになろう。１つのコンテキストだけが必要とされる場合は、これらの２つのコンテキスト・フィールド３０６、３０８は同じ値を有することになろう。この状況においては、データ・オブジェクトのための１つのコンテキストだけが必要である。従って、ファイルセット・マネージャ２１４はＬＣＸフィールド３０６およびＬＵＸフィールド３０８を同じ値（例えば１）に設定する。ファイルセット・マネージャ２１４は、ファイル・システム内のデータ・オブジェクトの位置（例えばブロック番号）に基づき物理的位置フィールド３１０を設定する。ファイルセット・マネージャ２１４は、このデータ・オブジェクトのための関連ｉノードの位置に基づき論理的位置フィールド３１２を設定する。例えば、論理的位置は、このデータ・オブジェクトが記憶される、ｉノードに加えていくらかのオフセットの

物理的位置を含むことができる。ファイルセット・マネージャ２１４は、データの第１のコピーが配置されるメモリ２１８内の位置を指し示すためにバッファ・ヘッダ３００内のデータ・ポインタ０・フィールド３１４を更新する。この状況は複数のコンテキストを要求していないので、第２のデータ・オブジェクトは必要ない。それ故、ファイルセット・マネージャ２１４はデータ・ポインタ１・フィールド３１６を、ヌルを指し示すように更新する。フローチャート５００の作業は５０８において続く。

【００３６】

ファイルセット・マネージャ２１４は、ファイル・システム内のデータ・オブジェクトＡ（現在の一貫性スナップショットの一部となるべきデータ・オブジェクトの一部である）を更新するトランザクションを受信する（現在の一貫性スナップショットを記憶するためのコミットがまだ行われている間に）（５０８）。図２を参照すると、ファイルセット・マネージャ２１４は、クライアント・デバイス２０８、２１０のうちの１つから、データ・オブジェクトＡを更新するトランザクションを受信する。例えば、クライアント・デバイス２０８、２１０のうちの１つの上で実行するアプリケーションがデータ・オブジェクトＡを更新することができる。フローチャート５００の作業は５１０において続く。

【００３７】

ファイルセット・マネージャ２１４は、メモリ内にデータ・オブジェクトＡのためのバッファ・ヘッダがあるかどうかを判定する（５１０）。図２を参照すると、ファイルセット・マネージャ２１４は、メモリ２１８内にデータ・オブジェクトＡのためのバッファ・ヘッダがあるかどうかを判定する。特に、実施形態例によっては、データ・オブジェクトがアクセスされる（読み出される、書き込まれる等）度に、メモリ２１８内に関連バッファ・ヘッダが作成される。メモリ２１８内にデータ・オブジェクトＡのためのバッファ・ヘッダがすでにある場合は、フローチャート５００の作業は継続点Ａにおいて続く（５１８）。さもなければ、フローチャート５００の作業は５１２において続く。

【００３８】

ファイルセット・マネージャ２１４はデータ・オブジェクトＡのためにメモリ内にバッファ・ヘッダを作成する（５１２）。図２を参照すると、ファイルセット・マネージャ２１４はデータ・オブジェクトＡのためにメモリ２１８内にバッファ・ヘッダを作成する。なぜなら、メモリ２１８内にはデータ・オブジェクトＡのための関連バッファ・ヘッダがないからである。ファイルセット・マネージャ２１４はバッファ・ヘッダのフィールドにデータを格納することもできる（以下の作業によってさらに記載されるように）。フローチャート５００の作業は５１４において続く。

【００３９】

ファイルセット・マネージャ２１４はデータ・オブジェクトＡのためのバッファ・ヘッダ内のデータ・ポインタ０・フィールドを更新する（５１４）。図２～３を参照すると、ファイルセット・マネージャ２１４は、データの第１のコピーが配置されるメモリ２１８内の位置を指し示すために、バッファ・ヘッダ３００内のデータ・ポインタ０・フィールド３１４を更新する。フローチャート５００の作業は５１６において続く。

【００４０】

ファイルセット・マネージャ２１４はデータ・オブジェクトＡのためのバッファ・ヘッダ内の物理的位置、論理的位置、ＬＣＧフィールドおよびＬＣＸフィールドも更新する。図２～３を参照すると、ファイルセット・マネージャ２１４はバッファ・ヘッダ３００のための物理的位置フィールド３１０、論理的位置フィールド３１２、ＬＣＧフィールド３０２、ＬＣＸフィールド３０６を更新する。ファイルセット・マネージャ２１４は、ファイル・システム内のデータ・オブジェクトの位置（例えばブロック番号）に基づき物理的位置フィールド３１０を設定する。ファイルセット・マネージャ２１４は、このデータ・オブジェクトのための関連ｉノードの位置に基づき論理的位置フィールド３１２を設定する。例えば、論理的位置は、このデータ・オブジェクトが記憶される、ｉノードに加えていくらかのオフセットの物理的位置を含むことができる。ファイルセット・マネージャ２１４はＬＣＧフィールド３０２をデータ・オブジェクトＡのための現在の世代値に設定す

る。例えば、もし最後にコミットされた一貫性スナップショットが５の値を有していれば、ファイルセット・マネージャ２１４はＬＣＧフィールド３０２を５に設定することになろう。コンテキスト・フィールド（３０６、３０８）は、２つのコンテキスト（コミット中のコンテキストおよび更新中のコンテキスト）を区別するために０または１のいずれかに設定される。従って、第２のコンテキストが必要とされる場合は、これらの２つのコンテキスト・フィールド３０６、３０８は逆の値を有することになろう。１つのコンテキストだけが必要とされる場合は、これらの２つのコンテキスト・フィールド３０６、３０８は同じ値を有することになろう。ファイルセット・マネージャ２１４はＬＣＸフィールド３０６を１に設定すると仮定する。ＬＵＸフィールド３０８の設定は以下において記載される。フローチャート５００の作業は継続点Ａ（５１８）において続く。

【００４１】

継続点Ａ（５１８）はフローチャート６００の継続点Ａ（６０２）において続く。継続点Ａ（６０２）より、作業は６０３において続く。

【００４２】

ファイルセット・マネージャ２１４は、データ・オブジェクトＡのためのバッファ・ヘッダ内のＬＣＧフィールドまたはＬＵＧフィールドの値がトランザクションの世代値と一致するかどうかを判定する（６０３）。図２～３を参照すると、ファイルセット・マネージャ２１４は、バッファ・ヘッダ３００内のＬＣＧフィールド３０２の値またはＬＵＧフィールド３０４の値がトランザクションの世代値と一致するかどうかを判定する。トランザクションの世代値は、トランザクションがいつ作成されたかに基づき一貫性世代に設定される。従って、ファイルセット・マネージャ２１４は、トランザクションに関連付けられるこの世代が、最後にコミットされた世代または最後に更新された世代に等しいかどうかを判定する。一致がなければ、作業は６０４において続く。さもなければ、作業は６１６において続く（以下においてさらに記載される）。

【００４３】

ファイルセット・マネージャ２１４はデータ・オブジェクトＡの第１のコピーからデータ・オブジェクトＡの第２のコピーを作成する（６０４）。図２を参照し、データの第１のコピー２５０はデータ・オブジェクトＡの第１のコピーであるとすると、ファイルセット・マネージャ２１４はデータの第１のコピー２５０をメモリ２１８内の異なる位置－データの第２のコピー２５２－にコピーする。フローチャート６００の作業は６０６において続く。

【００４４】

ファイルセット・マネージャ２１４は、データ・オブジェクトＡの第２のコピーを指し示すようにバッファ・ヘッダ内の第２のデータ・ポインタを更新する（６０６）。図２～３を参照すると、ファイルセット・マネージャ２１４は、メモリ２１８内のデータ・オブジェクトＡの第２のコピーを指し示すようにデータ・ポインタ１・フィールド３１６を更新する。フローチャート６００の作業は６０８において続く。

【００４５】

ファイルセット・マネージャ２１４はバッファ・ヘッダ内のＬＵＸフィールドを、ＬＣＸフィールドの値とは逆の値を有するように更新する（６０８）。図２～３を参照すると、ファイルセット・マネージャ２１４はＬＵＸフィールド３０８を、バッファ・ヘッダ３００内のＬＣＸフィールド３０６の値とは逆である値を有するように更新する。上述されたように、ＬＣＸフィールド３０６およびＬＵＸフィールド３０８の値は２つの値のうちの１つであることができる。デュアル・コンテキスト状況が生じると（この場合のように）、ＬＣＸフィールド３０６およびＬＵＸフィールド３０８の値は互いの逆になる。フローチャート６００の作業は６１０において続く。

【００４６】

ファイルセット・マネージャ２１４は、トランザクションのための世代値に基づきバッファ・ヘッダ内のＬＵＧフィールドのための世代値を設定する（６１０）。図２～３を参照すると、ファイルセット・マネージャ２１４は、トランザクションのための世代値に基

づきＬＵＧフィールド３０４のための世代値を更新する（上記の６０３の記載におけるトランザクションのための世代値の記載を参照）。フローチャート６００の作業は６１４において続く。

【００４７】

ファイルセット・マネージャ２１４は、このトランザクションに基づきデータ・オブジェクトＡの第２のコピーを更新する（６１４）。図２～３を参照し、データの第２のコピー２５２はデータ・オブジェクトＡの第２のコピーであるとすると、ファイルセット・マネージャ２１４は、データ・ポインタ１・フィールド３１６内のポインタ値に基づきデータの第２のコピー２５２を更新する。フローチャート６００の作業は、フローチャート６００のこの経路に沿い完結している。

【００４８】

６０３に戻り、一致があるとすると（はいの判定）、ファイルセット・マネージャ２１４は、データ・オブジェクトＡのためのバッファ・ヘッダ内のＬＵＸフィールドに関連付けられる第１のデータ・ポインタを用いてデータ・オブジェクトＡのコピーを更新する（６１６）。この状況においては、トランザクションのための世代がＬＵＧフィールド３０４と一致するであろうから、６０３において一致があった。図２～３を参照し、第１のデータ・ポインタがデータの第１のコピー２５０を指し示すとすると、ファイルセット・マネージャ２１４は、データ・ポインタ０・フィールド３１４内のポインタ値に基づきデータの第１のコピー２５０を更新する。フローチャート６００の作業は、フローチャート６００のこの経路に沿い完結している。

【００４９】

ファイル・システム内の同じまたは異なるデータ・オブジェクトに対する追加の更新が引き続き行われることができる。同様に、一貫性スナップショットのコミットの完了後、ファイルセット・マネージャ２１４は、（一貫性スナップショットを永続的に記憶するためにコミットするための定期的な間隔に基づき）追加の一貫性スナップショットをコミットすることができる。

【００５０】

当業者によって理解されるように、本発明の主題の態様はシステム、方法またはコンピュータ・プログラムとして具体化されればよい。それ故、本発明の主題の態様は、完全にハードウェアの実施形態、完全にソフトウェアの実施形態（ファームウェア、常駐ソフトウェア、マイクロ・コード等を含む）あるいはソフトウェアおよびハードウェアの態様を組み合わせた実施形態という形をとってよく、本願明細書においてそれらはすべて広く「回路」、「モジュール」または「システム」と呼ばれればよい。さらに、本発明の主題の態様は、１つ以上のコンピュータ可読媒体（単数または複数）であって、その上に具体化されるコンピュータ可読プログラム・コードを有する、コンピュータ可読媒体内に具体化されるコンピュータ・プログラムという形をとってもよい。

【００５１】

１つ以上のコンピュータ可読媒体（単数または複数）の任意の組み合わせが利用されてよい。コンピュータ可読媒体はコンピュータ可読信号媒体またはコンピュータ可読記憶媒体であればよい。コンピュータ可読記憶媒体は、例えば、電子的、磁気的、光学的、電磁的、赤外線または半導体システム、装置またはデバイス、あるいは上述のものの任意の適当な組み合わせであればよい。ただし、それらに限定されるものではない。コンピュータ可読記憶媒体のより具体的な例（限定的なリスト）としては以下のもの：１本以上のワイヤを有する電気接続、ポータブル・コンピュータ・ディスケット、ハード・ディスク、ランダム・アクセス・メモリ（ｒａｎｄｏｍ ａｃｃｅｓｓ ｍｅｍｏｒｙ、ＲＡＭ）、リード・オンリー・メモリ（ｒｅａｄ－ｏｎｌｙ ｍｅｍｏｒｙ、ＲＯＭ）、消去可能プログラマブル・リード・オンリー・メモリ（ｅｒａｓａｂｌｅ ｐｒｏｇｒａｍｍａｂｌｅ ｒｅａｄ－ｏｎｌｙ ｍｅｍｏｒｙ、ＥＰＲＯＭまたはフラッシュ・メモリ）、光ファイバ、ポータブル・コンパクト・ディスク・リード・オンリー・メモリ（ｃｏｍｐａｃｔ ｄｉｓｃ ｒｅａｄ－ｏｎｌｙ ｍｅｍｏｒｙ、ＣＤ－ＲＯＭ）、光学式記憶デバイス

、磁気記憶デバイス、あるいは上述のものの任意の適当な組み合わせ、が挙げられよう。本文書の文脈において、コンピュータ可読記憶媒体とは、命令実行システム、装置またはデバイスによって用いられるまたはそれらと連係して用いられるプログラムを包含または記憶することができる任意の有形媒体であればよい。

【００５２】

コンピュータ可読信号媒体とは、例えば、基底帯域内にまたは搬送波の一部として、内部にコンピュータ可読プログラム・コードが具体化される伝搬データ信号を含むものであればよい。このような伝搬信号は、電磁気的、光学的、またはそれらの任意の適当な組み合わせを含む、ただしそれらに限定されるものではない、様々な形態のいずれを取ってもよい。コンピュータ可読信号媒体とは、コンピュータ可読記憶媒体ではない、命令実行システム、装置またはデバイスによって用いられるまたはそれらと連係して用いられるプログラムを伝達、伝搬または輸送することができる任意のコンピュータ可読媒体であればよい。

【００５３】

コンピュータ可読媒体上に具体化されるプログラム・コードは、無線、有線、光ファイバケーブル、ＲＦ等、または上述のものの任意の適当な組み合わせを含む、ただしそれらに限定されるものではない、任意の適切な媒体を用いて送信されればよい。

【００５４】

本発明の主題の態様のための作業を行うためのコンピュータ・プログラム・コードは、Ｊａｖａ（Ｒ）、Ｓｍａｌｌｔａｌｋ（Ｒ）、Ｃ＋＋または同様のもの等のオブジェクト指向プログラミング言語、ならびに「Ｃ」プログラミング言語または同様のプログラミング言語等の従来の手続き型プログラミング言語を含む、１つ以上のプログラミング言語の任意の組み合わせで書き込まれればよい。プログラム・コードは、スタンド・アロン・ソフトウェア・パッケージとして完全にまたは一部分はユーザのコンピュータ上で実行するか、一部分はユーザのコンピュータ上で且つ一部分はリモート・コンピュータ上で実行するか、または完全にリモート・コンピュータもしくはサーバ上で実行すればよい。後者のシナリオでは、リモート・コンピュータは、ローカル・エリア・ネットワーク（ｌｏｃａｌ ａｒｅａ ｎｅｔｗｏｒｋ、ＬＡＮ）またはワイド・エリア・ネットワーク（ｗｉｄｅ ａｒｅａ ｎｅｔｗｏｒｋ、ＷＡＮ）を含む、任意の種類のネットワークを通じてユーザのコンピュータに接続されてもよいし、あるいは接続は、（例えば、インターネット・サービス・プロバイダを利用しインターネットを通じて）外部のコンピュータになされてもよい。

【００５５】

本発明の主題の実施形態による方法、装置（システム）およびコンピュータ・プログラムのフローチャート図もしくはブロック図またはその両方を参照しながら本発明の主題の態様が記載されている。フローチャート図もしくはブロック図またはその両方の各ブロック、ならびにフローチャート図もしくはブロック図またはその両方におけるブロックの組み合わせは、コンピュータ・プログラム命令によって実装されることができることは理解されよう。これらのコンピュータ・プログラム命令は、汎用コンピュータ、専用コンピュータ、または他のプログラム可能なデータ処理装置のプロセッサに提供されて機械を作り出せばよく、それにより、命令は、コンピュータまたは他のプログラム可能なデータ処理装置のプロセッサを介して実行し、フローチャートもしくはブロック図またはその両方のブロックまたはブロック群において特定される機能群／動作群を実装する手段を生み出す。

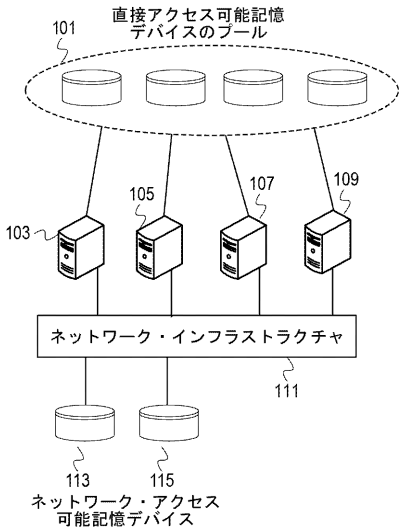
【００５６】

これらのコンピュータ・プログラム命令は、コンピュータ、他のプログラム可能なデータ処理装置または他のデバイスを特定の様式で機能するように仕向けることができるコンピュータ可読媒体内に記憶されてもよく、それにより、コンピュータ可読媒体内に記憶された命令は、フローチャートもしくはブロック図またはその両方のブロックまたはブロック群において特定される機能／動作を実装する命令を含む製造品を作り出す。

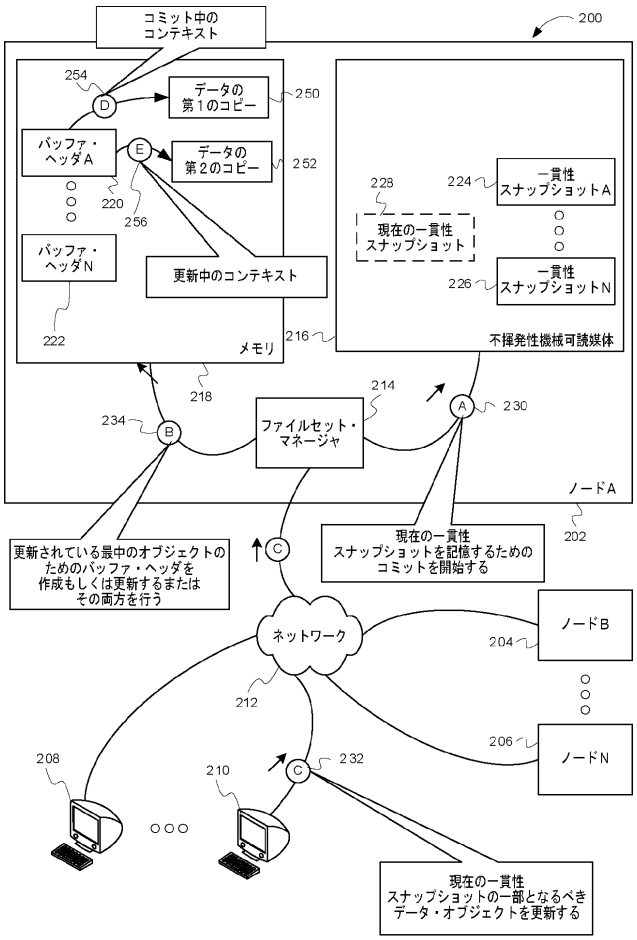
【００５７】
コンピュータ・プログラム命令は、コンピュータ、他のプログラム可能なデータ処理装置または他のデバイス上にロードされ、一連の作業ステップをコンピュータ、他のプログラム可能な装置または他のデバイス上で遂行させ、コンピュータ実装プロセスを作り出してもよく、それにより、コンピュータまたは他のプログラム可能な装置上で実行する命令は、フローチャートもしくはブロック図またはその両方のブロックまたはブロック群において特定される機能群／動作群を実装するためのプロセス群を提供する。
【００５８】
図７はコンピュータ・システムの例を示す。コンピュータ・システムはプロセッサ・ユニット７０１（場合により、複数のプロセッサ、複数のコア、複数のノードを含む、もしくはマルチ・スレッド等を実装する、またはその両方の態様を有する）を含む。コンピュータ・システムはメモリ７０７を含む。メモリ７０７はシステム・メモリ（例えば、キャッシュ、ＳＲＡＭ、ＤＲＡＭ、ゼロ・キャパシタＲＡＭ、ツイン・トランジスタＲＡＭ、ｅＤＲＡＭ、ＥＤＯ　ＲＡＭ、ＤＤＲ　ＲＡＭ、ＥＥＰＲＯＭ、ＮＲＡＭ、ＲＲＡＭ、ＳＯＮＯＳ、ＰＲＡＭ等のうちの１つ以上）であってもよいし、または先にすでに記載された、機械可読媒体の実現可能なもののうちのいずれか１つ以上であってもよい。コンピュータ・システムは、バス７０３（例えば、ＰＣＩ、ＩＳＡ、ＰＣＩ－Ｅｘｐｒｅｓｓ、ＨｙｐｅｒＴｒａｎｓｐｏｒｔ（Ｒ）、ＩｎｆｉｎｉＢａｎｄ（Ｒ）、ＮｕＢｕｓ等）、ネットワーク・インターフェース７０５（例えば、ＡＴＭインターフェース、イーサネット（Ｒ）・インターフェース、フレーム・リレー・インターフェース、ＳＯＮＥＴインターフェース、無線インターフェース等）、および記憶デバイス（単数または複数）７０９（例えば、光学的記憶、磁気的記憶等）も含む。コンピュータ・システムは、リダイレクト・オン・ライト・ファイル・システム内のデータ・オブジェクトのための複数のコンテキストを提供するファイルセット・マネージャ７２５も含む。これらの機能性のうちのいずれか１つは一部（または完全に）、ハードウェアの形で、もしくはプロセッサ・ユニット７０１上で、またはその両方の態様で実装されてよい。例えば、該機能性は、特定用途向け集積回路を用いて、プロセッサ・ユニット７０１内に実装される論理で、周辺デバイスまたはカード上のコプロセッサ内に、等の態様で実装されればよい。さらに、現実のものは、含まれる構成要素がもっと少なくてもよいし、または図７には示されていない追加の構成要素（例えば、ビデオ・カード、オーディオ・カード、追加のネットワーク・インターフェース、周辺デバイス、等）を含んでもよい。プロセッサ・ユニット７０１、記憶デバイス（単数または複数）７０９およびネットワーク・インターフェース７０５はバス７０３に結合される。バス７０３に結合されるように示されているが、メモリ７０７はプロセッサ・ユニット７０１に結合されてもよい。
【００５９】
実施形態は、様々な実装および利用を参照しながら記載されているが、これらの実施形態は例示的なものであること、および本発明の主題の範囲はそれらに限定されないことは理解されよう。多くの変形、変更、追加および改良が可能である。
【００６０】
本願明細書では単一の例として記載されている構成要素、作業または構造に複数の例が提供されてもよい。最後に、様々な構成要素同士、作業同士およびデータ格納同士の間の境界は、いくぶんは自由に決めてよいものであり、特定の作業は特定の例示的構成を背景として示されている。機能性の他の割り振りが想定され、本発明の主題の範囲に含まれてもよい。一般に、構成例において別個の構成要素として提示されている構造および機能性は、組み合わせられた構造または構成要素として実装されてもよい。同様に、単一の構成要素として提示されている構造および機能性は、別個の構成要素として実装されてもよい。これらおよび他の変形、変更、追加および改良は本発明の主題の範囲に含まれてよい。
【符号の説明】
【００６１】
２００　システム

２０２　ノードＡ
２０４　ノードＢ
２０６　ノードＮ
２０８、２１０　クライアント・デバイス
２１２　ネットワーク
２１４　ファイルセット・マネージャ
２１６　不揮発性機械可読媒体
２１８　メモリ
２２０　バッファ・ヘッダＡ
２２２　バッファ・ヘッダＮ
２２４　一貫性スナップショットＡ
２２６　一貫性スナップショットＮ
２２８　現在の一貫性スナップショット
２３０、２３２、２３４　作業
２５０　データの第１のコピー
２５２　データの第２のコピー
２５４　コミット中のコンテキスト
２５６　更新中のコンテキスト

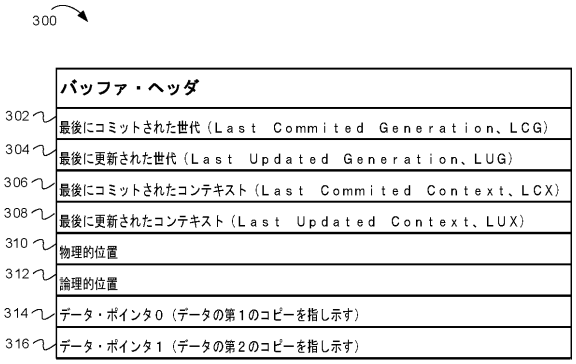
【図１】



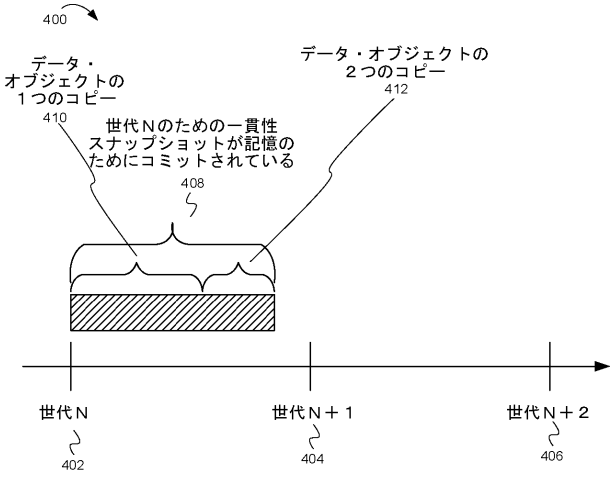
【図２】



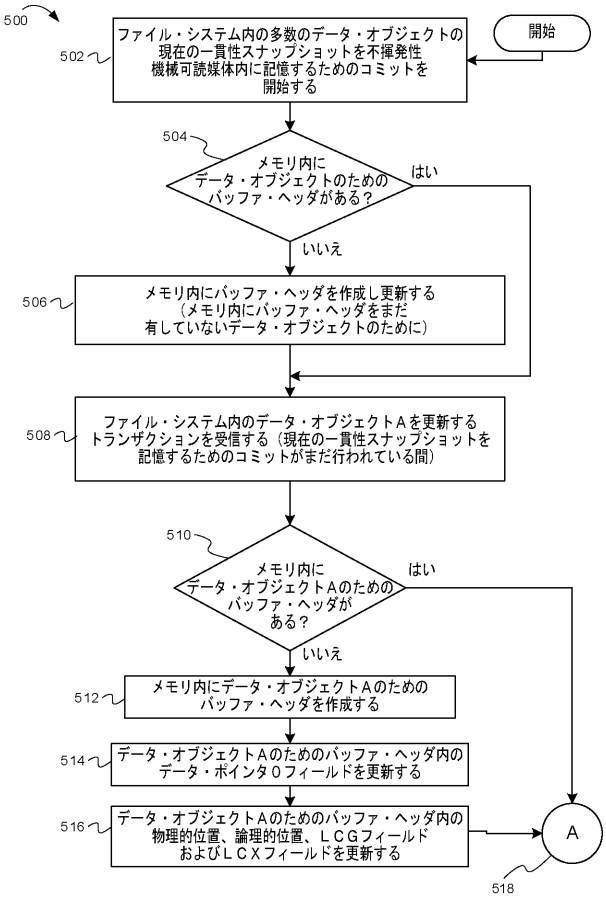
【図３】


300
バッファ・ヘッダ
302
最後にコミットされた世代（Last Commited Generation、LCG）
304
最後に更新された世代（Last Updated Generation、LUG）
306
最後にコミットされたコンテキスト（Last Commited Context、LCX）
308
最後に更新されたコンテキスト（Last Updated Context、LUX）
310
物理的位置
312
論理的位置
314
データ・ポインタ０（データの第１のコピーを指し示す）
316
データ・ポインタ１（データの第２のコピーを指し示す）

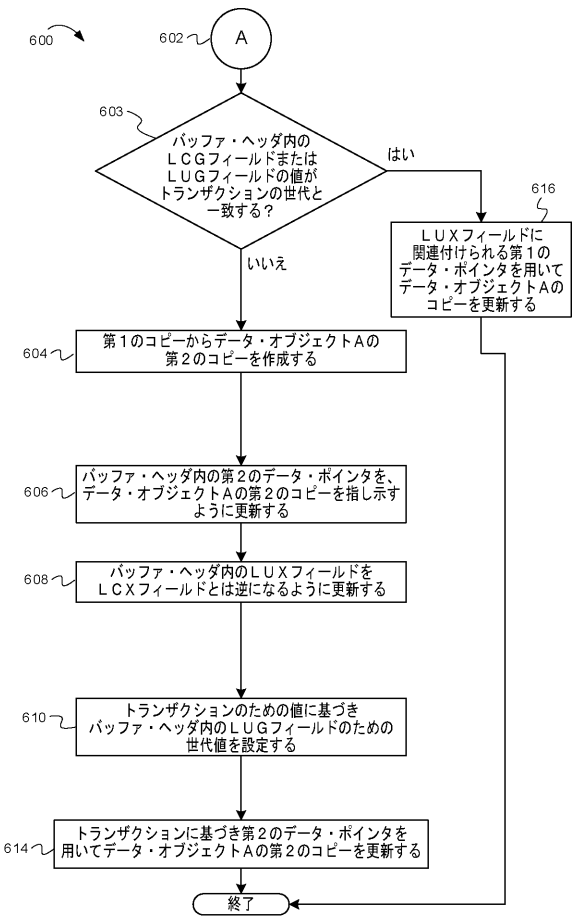
【図４】


400
データ・
オブジェクトの
１つのコピー
410
データ・オブジェクトの
２つのコピー
412
世代Ｎのための一貫性
スナップショットが記憶の
ためにコミットされている
408
世代Ｎ
402
世代Ｎ＋１
404
世代Ｎ＋２
406

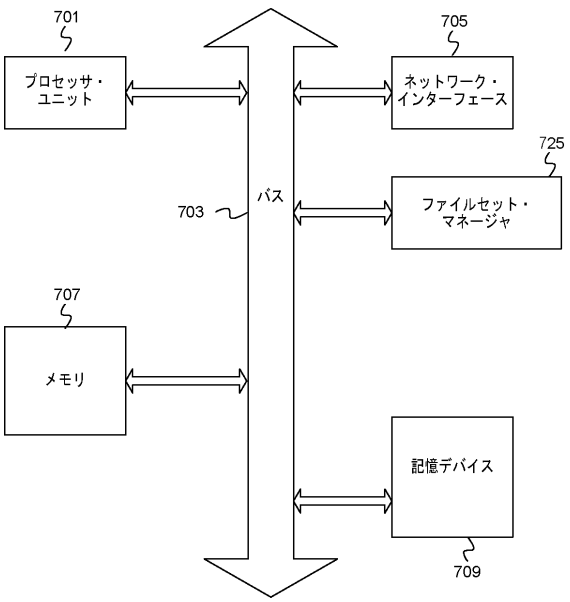
【図５】


500
開始
502
ファイル・システム内の多数のデータ・オブジェクトの
現在の一貫性スナップショットを不揮発性
機械可読媒体内に記憶するためのコミットを
開始する
504
メモリ内に
データ・オブジェクトのための
バッファ・ヘッダがある？
はい
いいえ
506
メモリ内にバッファ・ヘッダを作成し更新する
（メモリ内にバッファ・ヘッダをまだ
有していないデータ・オブジェクトのために）
508
ファイル・システム内のデータ・オブジェクトＡを更新する
トランザクションを受信する（現在の一貫性スナップショットを
記憶するためのコミットがまだ行われている間）
510
メモリ内に
データ・オブジェクトＡのための
バッファ・ヘッダが
ある？
はい
いいえ
512
メモリ内にデータ・オブジェクトＡのための
バッファ・ヘッダを作成する
514
データ・オブジェクトＡのためのバッファ・ヘッダ内の
データ・ポインタ０フィールドを更新する
516
データ・オブジェクトＡのためのバッファ・ヘッダ内の
物理的位置、論理的位置、ＬＣＧフィールド
およびＬＣＸフィールドを更新する
A
518

【図６】


600
602
A
603
バッファ・ヘッダ内の
ＬＣＧフィールドまたは
ＬＵＧフィールドの値が
トランザクションの世代と
一致する？
はい
いいえ
616
ＬＵＸフィールドに
関連付けられる第１の
データ・ポインタを用いて
データ・オブジェクトＡの
コピーを更新する
604
第１のコピーからデータ・オブジェクトＡの
第２のコピーを作成する
606
バッファ・ヘッダ内の第２のデータ・ポインタを、
データ・オブジェクトＡの第２のコピーを指し示す
ように更新する
608
バッファ・ヘッダ内のＬＵＸフィールドを
ＬＣＸフィールドとは逆になるように更新する
610
トランザクションのための値に基づき
バッファ・ヘッダ内のＬＵＧフィールドのための
世代値を設定する
614
トランザクションに基づき第２のデータ・ポインタを
用いてデータ・オブジェクトＡの第２のコピーを更新する
終了

【図７】


701
プロセッサ・
ユニット
705
ネットワーク・
インターフェース
725
703
バス
ファイルセット・
マネージャ
707
メモリ
記憶デバイス
709

フロントページの続き

(72)発明者　アンドリュー・ソロモン
アメリカ合衆国７８７５８　テキサス州　オースティン　バーネット　ロード　１１５０１　エムディー　９５４１
(72)発明者　デイビット・ジョーンズ・クラフト
アメリカ合衆国７８７５８　テキサス州　オースティン　バーネット　ロード　１１５０１　ジップ　９０５
(72)発明者　マノージ・エヌ・クマール
アメリカ合衆国７８７５８－３４００　テキサス州　オースティン　バーネット　ロード　１１５０１　４エイチ－００４　９０５
(72)発明者　ジャネット・エリザベス・アドキンス
アメリカ合衆国７８７５８－３４００　テキサス州　オースティン　バーネット　ロード　１１５０１
(72)発明者　ジュン・チャン
アメリカ合衆国７８７３３　テキサス州　オースティン　ノース　ウェストン　エルエヌ　９０１