

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2008-547123

(P2008-547123A)

(43) 公表日 平成20年12月25日(2008.12.25)

(51) Int.Cl.		F I		テーマコード(参考)
<b>G06F 12/00</b>	<b>(2006.01)</b>	G06F 12/00	531M	5B065
<b>G06F 3/06</b>	<b>(2006.01)</b>	G06F 3/06	304F	5B082
		G06F 12/00	510B	

審査請求 未請求 予備審査請求 未請求 (全 38 頁)

(21) 出願番号 特願2008-518430 (P2008-518430)  
 (86) (22) 出願日 平成18年6月23日 (2006.6.23)  
 (85) 翻訳文提出日 平成20年2月25日 (2008.2.25)  
 (86) 国際出願番号 PCT/US2006/024478  
 (87) 国際公開番号 W02007/002397  
 (87) 国際公開日 平成19年1月4日 (2007.1.4)  
 (31) 優先権主張番号 60/693,715  
 (32) 優先日 平成17年6月24日 (2005.6.24)  
 (33) 優先権主張国 米国 (US)

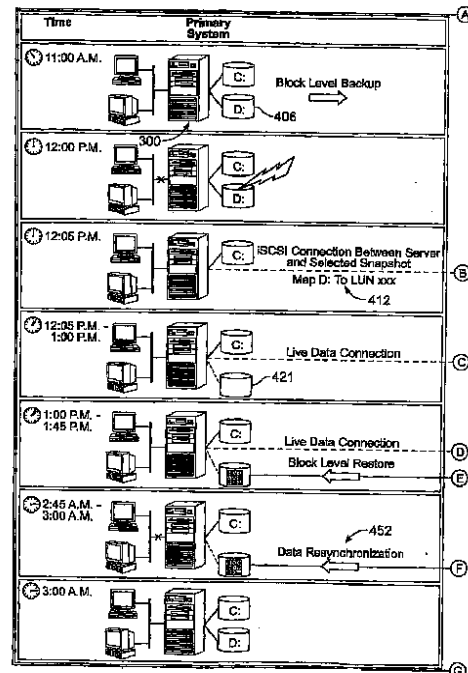
(71) 出願人 507421164  
 シンクソート インコーポレイテッド  
 アメリカ合衆国 ニュージャージー州 O  
 7675 ウッドクリフ レイク タイス  
 ブールヴァード 50  
 (74) 代理人 100082005  
 弁理士 熊倉 禎男  
 (74) 代理人 100067013  
 弁理士 大塚 文昭  
 (74) 代理人 100086771  
 弁理士 西島 孝喜  
 (74) 代理人 100109070  
 弁理士 須田 洋之

最終頁に続く

(54) 【発明の名称】 高性能な企業データ保護のためのシステムおよび方法

(57) 【要約】

高性能な企業レベルのデータ保護システムおよび方法は、第一次ストレージ装置の効率的なブロックレベルのインクリメンタル(incremental)スナップショットと、第一次ストレージ装置の代わりに直接使用することができる瞬時に組み込み可能な形態で当該スナップショットの即時利用性とを提供する。関連システムおよびアプリケーションは、バックアップ画像用の「企業用画像の移動先(enterprise image destination(EID))」と、代替の物理的第一次ファシリティ(facility)を作成し、一方で第一次ソースファイルシステムとして第二次ストレージユニットと協働で動作するミラーリング技法と、部分的に共通のコンテンツ(オペレーティングシステムファイル、共通データベース、アプリケーションエグゼクタブル(application executable)など)を有する複数のシステムがバックアップされるときに、バックアップ画像中の冗長なデータを解消するための技術を含めて、開示する。「高速アプリケーション復元(First Application Restore (FAR))」を本発明の使用と共に示す、複数の例を提供する。



**【特許請求の範囲】****【請求項 1】**

即時利用可能であることを基本とする同一または異なるコンピュータシステム上に組込み可能な第一次システムのバックアップコピーを作成する方法であって、

第二次システム上で記憶された、前記第一次システムのベースレベルのスナップショットを提供し、

前記第二次システムに記憶された、前記第一次システムのブロックレベルのインクリメンタル(incremental)スナップショットを提供し、スナップショットは前のスナップショット以後に変更されたブロックのみを表わすものであり、

組込まれたストレージユニットとして使用することができる前述のインクリメンタル(incremental)スナップショット画像の少なくとも1つから論理ディスク画像を作成する、方法。

10

**【請求項 2】**

アプリケーション特有のメタデータをバックアップすることをさらに含む請求項 1 に記載の方法。

**【請求項 3】**

バックアップを制御するソフトウェアプロセスと、バックアップの開始と最後にアプリケーション特有の対話を仲介するためのソフトと、ソフトウェアを処理するスナップショットが存在し、さらに、

前記制御プロセスがメッセージを前記仲介プロセスに送信し、

20

それを受けて前記ソフトウェアを処理するスナップショットがボリュームのターゲットセットをスナップショットし、且つ変更ジャーナルをフラッシュ(flush)し、

前記ボリュームのターゲットセットをスナップショットしてスナップショット画像を作成し、

バックアップ仕様のためにコンテンツファイルを作成し、

前記コンテンツファイルおよびスナップショット画像を、前記制御プロセスの制御下でストレージ装置に伝送する、請求項 1 に記載の方法。

**【請求項 4】**

前記バックアップ用データは、ディファレンシャル(differential)ブロックレベルの画像を含み、最初にディスク署名を付与され、有効な単一パーティションを有する SCSI ディスク全体としてのボリュームのバックアップを仮想化するための適切な情報を有する、請求項 1 に記載の方法。

30

**【請求項 5】**

前記スナップショットのステップが変更ジャーナルをフラッシュ(flushing)し且つ変更されたブロックを追跡する(tracking)目的のために内部スナップショットした後、外部スナップショットして、一貫性のある(consistent)スナップショット画像を生成する、請求項 1 に記載の方法。

**【請求項 6】**

請求項 1 により作成されたバックアップ画像に関する企業の画像移動先であり、その画像は統合され、近くに並べ、バージョン管理され、

40

前記企業の画像移動先によってアクセス可能な1つ以上のデータストレージユニットと、

前記企業の画像移動先を操作するように適合され、そして複数の永続スナップショットを収容するためにスパー(sparse)ファイルサポートでFSをジャーナル化するサポートを含むソフトウェアと

LUNの作成/削除と、LUNのクローン化と、スナップショット作成および削除をサポートするインタフェースと、

スナップショットIDをブロックアロケーションビットマップおよびインクリメンタル(incremental)ビットマップと関連づけるデータベースと、

を含む、企業の画像移動先。

50

**【請求項 7】**

1つ以上の前記データストレージユニットが、(i)共有ストレージ、(ii)バックアップLUNへのローカルコピー、および(iii)ネットワークストレージから成るグループから選択されるストレージ装置によって提供される、請求項6に記載の企業の画像移動先。

**【請求項 8】**

前記インタフェースが、プラグインインタフェースである、請求項6に記載の企業の画像移動先。

**【請求項 9】**

前記インタフェースがブロックレベルのミラーリングをサポートする、請求項6に記載の企業の画像移動先。

10

**【請求項 10】**

クライアントノードのバックアップ画像を作成するために、請求項6に記載の企業の画像移動先によって開始され、

(1)クライアントノード上の前記対応するソフトウェアとの接続を確立し、クライアントノードにコンテンツファイルへのパスを渡し、

(2)前記クライアントノード上のソフトウェアを用いて、コンテンツファイルのコンテンツを読み出し、そして前記企業の画像移動先のノード上にソフトウェアに戻す、付加的なデータフローを実施するステップを更に含む請求項3に記載の方法。

**【請求項 11】**

前記企業の画像移動先によって使用される1つ以上の前記データストレージユニットが、(i)共有ストレージ、(ii)バックアップLUNへのローカルコピー、および(iii)ネットワークストレージから成るグループから選択されるストレージ装置によって提供される、請求項10に記載の方法。

20

**【請求項 12】**

前記記憶装置がネットワークストレージ装置を含み、さらに、前記クライアントノード上でソフトウェアを使用し、バックアップされるべき装置のローカルスナップショットから割当てられたブロックを読み取り、

ネットワーク中で前記ブロックを前記企業の画像移動先ノードの移動先ソフトウェアに伝送し、

前記企業の移動先ノード上でソフトウェアを使用し、前記ネットワークのチャンネルから読み取り、前記ネットワークストレージ装置上の予め定義されたボリューム上の(sparse)ファイルに読み出された日付を書込む、ことを含む請求項10に記載の方法。

30

**【請求項 13】**

請求項1に従って作成された第一次ストレージユニットのバックアップ画像を含む、機械可読媒体。

**【請求項 14】**

請求項13に記載される機械可読バックアップ画像のバックアップ画像を作成する方法であって、前記媒体の第三次のバックアップコピーを作成することを含み、前記第三次のバックアップコピーは、前記第1次ストレージユニットの十分に完全で自給式のバックアップである、方法。

40

**【請求項 15】**

前記第3のバックアップ画像が、テープ上に作成される、請求項14に記載の第三次バックアップ画像を作成する方法。

**【請求項 16】**

前記第一次バックアップ画像が、tar、cpioまたはpaxから本質的になるグループから選択されるタイプのポータブルフォーマットの論理バックアップに変換される、請求項14に記載の第三次バックアップ画像を作成する方法。

**【請求項 17】**

前記第三次画像が、ワーム(WORM)テープまたはWORMドライブに記録される、請求項14に記載の方法。

50

**【請求項 18】**

前記第一次バックアップ画像の作成時に保存されたファイル履歴を使用しデータを復元することをさらに含む請求項 14 に記載の方法。

**【請求項 19】**

請求項 14 に従って作成される第一次ストレージユニットの第三次バックアップコピーを含む機械可読媒体。

**【請求項 20】**

請求項 6 に記載されている企業の画像移動先を構成するノード内に記憶される第一次バックアップ画像から第三次ノード上に第三次画像を作成または更新する方法であって、前記企業の画像移動先上のソフトウェアを使用して前記第三次ノードにコンタクトをとり、それに前記企業用画像移動先ノードからのデータを要求するメッセージを渡し、前記第三次ノードノード上のソフトウェアを用いて前記データを要求し、前記企業の画像移動先ノードで前記要求を受理し、第三次バックアップ画像を作成または更新する要求として認識し、

保存されたビットマップを用いて適切なスナップショット画像を指定し、適切なスナップショット画像を前記第三次ノードに伝送する、ことを含む、方法。

**【請求項 21】**

請求項 1 によるバックアップ画像の即時利用可能な復元を提供する方法であって、スナップショット内にLUNを作成し公開して復元ボリュームを作成し、前記ローカル名スペースに対して復元ボリュームをマッピングする、方法。

**【請求項 22】**

請求項 2 によるバックアップ画像の即時利用可能な復元を提供する方法であって、スナップショット内にLUNを作成し公開して復元ボリュームを作成し、バックアップデータとアプリケーション特有のメタデータを含み、ローカル名スペースに対して復元ボリュームをマッピングする、方法。

**【請求項 23】**

前記バックアップ画像が、読取り専用画像であって、前記読取り専用バックアップ画像によってバック(back)されるスペース(sparse)ファイルを作成することによって、前記画像をiSCSIアドレス指定可能な読み書きLUNに変換することを含む、請求項 22 に記載の方法。

**【請求項 24】**

変更を集約し、前記バックアップ画像からの元の未変更のデータを維持することができる第一次ストレージ装置を復元する方法であって、請求項 22 に従って作成されるLUNを組込み、前記LUNを第一次ストレージとして使用することを含む、方法。

**【請求項 25】**

アプリケーションまたはデータベースを回復する方法であって、請求項 21 に記載の復元動作を実施し、必要な場合には、ソースボリュームから復元ボリュームにログファイルをコピーし前記アプリケーションまたはデータベースを回復する、方法。

**【請求項 26】**

ボリュームをより初期の状態にロールバックする方法であって、独自のジョブ識別によって前記ボリュームをアドレス指定し、当該アドレスに指定されるボリュームをロックおよび組込み解除し、前記ボリュームのスナップショットを取り、変更ジャーナルをフラッシュ(flush)し、前記マップを組込み可能なボリュームとしてマッピングし、復元される所望の状態の時間以降のすべて変更ジャーナルを論理的にANDで結合し、ビットマップファイルを作成し、前記ビットマップファイルのコンテンツを前記スナップショットに関連するコンテンツファイルに追加し、前記ビットマップファイルを使用し、前記マッピングされたボリュームからブロックのセットのみを復元する、方法。

10

20

30

40

50

**【請求項 27】**

前記ターゲットLUNが、別のノード上でバックエンド化される方法であって、バックエンドストレージが請求項 26 に従ってアドレス指定することを支援するか否かと、復元されるスナップショットおよび復元ターゲットLUNが論理的に関連するか否かを決定し、

バックエンドAPIを復元されているスナップショットの識別子と、ターゲットノード上でボリュームをバックエンドするターゲットエンティティまたはLUNと、一緒に呼出す、請求項 23 に記載の方法。

**【請求項 28】**

仮想化サーバ下で動作し、前記サーバ上で動作する仮想マシン上で、請求項 1 に従って作成されるバックアップ画像を復元する方法であって、

スナップショット内でLUNを作成および公開して復元ボリュームを作成し、前記仮想マシンの仮想サーバのローカル名スペースに対して復元ボリュームをマッピングする、方法。

**【請求項 29】**

第一次ボリュームを第二次ボリューム上にミラー化する方法であって、ブロックのリストを前記第一次ボリュームから第二次ボリュームにコピーし、

第一次および第二次のボリュームの両方に新規のブロックが到着した時にそのブロックを書き込み、

ブロックが書込まれるときに、それらのブロックを前記ブロックのリストから削除し、前記リストをトラバースし、そのようなトラバースの結果として発生するブロックを前記第一次ボリュームから第二次ボリュームにコピーし、

先の2つのステップを、前記リスト上のすべてのブロックがコピーされるまで継続する、方法。

**【請求項 30】**

クラッシュまたは非同期化したミラーを、生成時に残る第一次ミラーから、再ミラー化する方法において、前記第一次ミラーを請求項 29 に記載の方法に従って新たな第二次ミラーに対してミラー化することを含む、方法。

**【請求項 31】**

複数のデータセットをバックアップするために使用されるバックアップサーバに、変更されたブロックベースでバックアップされる第一次データセットの冗長性を低減する方法であって、

前記第一次データセット内で変化した各ブロックのダイジェストを取得し、

前記ダイジェストデータを保存し、

前記ブロックのブロックダイジェストを前記サーバによって以前に保存されたブロックのダイジェストと比較し、既に前記サーバがどのブロックかを所有していないかを決定し、前記サーバに、前記サーバが所有していないと決定したブロックのみを、関連ダイジェストと一緒に伝送し、

前記関連ダイジェストを追加することによって、前記サーバ上に保存されたダイジェストの集合を更新し、

前記サーバがすでに所有していると決定したブロックリストを加えて、バックアップサーバに伝送されたブロックのリストから前記バックアップのカタログを作成する、方法。

**【請求項 32】**

ファイルシステムの冗長性を低減する方法において、

各ブロックのダイジェストを、そのダイジェストがファイルシステムに書込まれる際に取得し、

前記ブロックのダイジェストをファイルシステムに記憶されたブロックのダイジェストと比較し、

同一のコンテンツブロックがファイルシステムに既に存在する場合には、既存のディレクトリポインタを使用するが、複製ブロックを書込まず、

10

20

30

40

50

ファイルが削除されるときには、そのファイルから関連ブロックの割当てを解除し、ファイルがそのブロックを参照しないときには、ブロックを解放する、方法。

【請求項 3 3】

1つ以上のバックアップ画像のセットからアプリケーションをターゲットノードに迅速に復元するための方法であって、

前記バックアップ画像を選択し、

アドレス指定可能なストレージLUNをオンデマンドベースで前記画像から作成し、

前記LUNを前記ターゲットノードに対して見えるようにし、

前記LUNをローカルディスクとしてiSCSIプロトコルを介して復元ターゲット上に付与し、

10

前記アプリケーションを適切なアプリケーション特有のAPIを用いてプログラム可能に回復する、方法。

【請求項 3 4】

故障から回復するために、アプリケーション特有のログファイルを適用し、前記復元動作を故障発生時までロールフォワードすることをさらに含む、請求項 3 3に記載の方法。

【請求項 3 5】

請求項 3 3の方法に従って前記バックアップを前記アプリケーションを実行しているマシンに迅速に回復し、前記アプリケーションが正確に起動および回復しているかを検証する、ことを含む請求項 1に記載に従って作成されたアプリケーションのバックアップを迅速に検証するための方法。

20

【請求項 3 6】

アプリケーション特有の技法を用いて、アプリケーションデータが正しいこと、および前記アプリケーションの論理オブジェクトが適切に機能することを検証する、請求項 3 3に記載の方法。

【請求項 3 7】

第一次システムのファイングレインされた(fine-grained)部分復元を実施する方法であって、

前記システムを請求項 3 3に記載の方法に従って代替ノードに復元し、

アプリケーション特有のツールを使用して所望の部分データを抽出する、方法。

【発明の詳細な説明】

30

【技術分野】

【0001】

本発明は、情報技術の分野に関し、より詳しくは、高性能な企業レベルのバックアップおよび災害時回復システムに関する。

【0002】

(関連技術の説明)

最近の事件は、災害(人災と天災の両方)から迅速に回復する必要が極めて重要であることを証明している。企業レベルのバックアップおよび災害時回復システムは、この必要性に向けられている。最先端の技術では、バックアップ動作で典型的に最終生成されるのがバックアップボリュームであるが、これは稼働使用される前に相対的に冗長な「復元」プロセスを経なければならない。

40

【0003】

「ダウンタイムが短い(short downtime)」バックアップおよび回復ソリューションは複数あるが、一般にこれらは高価なサーバクラスタリング能力および複製能力またはいずれか一方を必要とする。

【0004】

本出願に関する最新の技術については、ストレージネットワークインダストリーアソシエーション(Storage Networking Industry Association (SNIA))の出版物に文書化されている。尚、それらの出版物は、オンラインでwww.snia.orgから入手可能である。具体的には、非特許文献 1、非特許文献 2、非特許文献 3、および非特許文献 4を参照されたい。

50

尚、各参照文献については、参照として引用されている。

【0005】

上述で引用される参照文献の検討から理解されるように、最新の技術では、システム故障または災害の場合に迅速に使用するために利用可能なシステムの最近のポイントインタイム(point-in-time)のスナップショットを作成するために、大規模サーバクラスタリングおよび複製またはいずれか一方以外の方法を提供していない。

【0006】

したがって、ある組織が利用可能なプロダクションサーバの最近の自己矛盾のない(self-consistent)画像を常に有することができて、システム故障または災害の場合には、ほぼ瞬間的にそれらの画像をオンラインに導入し積極的な生産を開始することができるような能力を提供する単純なハードウェアを実装したシステムを有することが望ましいと思われる。

【0007】

(関連出願の相互参照)

本願は、2005年6月24日に本願に参照として引用される米国特許仮出願第60/693,715号の出願日利益を主張するものである。また、本願は、本発明者らの共同発明であり本発明の譲受人に譲渡され本願と同日に出願された出願番号第\_\_号の「バックアップイメージを仮想化するためのシステムおよび方法(System And Method for Virtualizing Backup Images)」と称する出願の完全な開示内容を参照として引用する。

【0008】

【非特許文献1】マイケル・ローワン(Michael Rowan)著 「ディスクベースのデータ保護技術(Disk Based Data Protection Technologies) リビビオ社(Revivio Corporation)

【非特許文献2】ジャコブ・ファーマー(Jacob Farmer)著 「バックアップシステムのボトルネックの確認および解消(Identifying and Eliminating Backup System Bottlenecks) ケンブリッジコンピュータ社(Cambridge Computer Corporation)

【非特許文献3】マイケル・フィッシュマン(Michael Fishman)著 「最新のデータ保護に取り組むための技術(Technologies to Address Contemporary Data Protection) EMC社 (EMC Corporation)

【非特許文献4】アンドレア・チアフィッテリ(Andrea Chiaffitelli)著 「次世代ビジネスの継続性(Next Generation Business Continuity)」、AT&T社(AT&T Corp.)

【発明の開示】

【0009】

本発明の一実施態様は、本願の譲受人であるシンクソート社(Synsort Incorporated)のソフトウェア製品であるBackup Express<sup>(R)</sup>(BEX)の一部として利用できるようになっている。その中でも特にBackup Expressに実装されたものとして、本発明は、「ファーストアプリケーションリカバリー(Fast Application Recovery, FAR)」と呼ばれるサービスを提供する。尚、このサービスは、大半の企業のIT予算の十分な範囲内で、単純なハードウェアを使用して故障からほぼ瞬間的に回復させることができる。

【0010】

本発明のある目的は、第一次ストレージ装置の効率的なブロックレベルインクリメンタル(incremental)スナップショットを提供する高性能で企業レベルのデータ保護システムおよび方法と、第一次ストレージ装置の代わりに直接使用することができる直接組み込み可能な形態で当該スナップショットの即時利用性を提供することにある。

【0011】

本発明の別の目的としては、以下が挙げられる。

種々のストレージプラットフォームで本願明細書に記載される方法を容易にするように適合された当該スナップショット用の企業リポジトリを提供する。

リアルタイムで代替の物理的な第一次ファシリティが設置され別のストレージユニットと協働で第一次ファシリティ(the primary)として動作する能力を提供する。

多重バックアップ時におよび単一ファイルシステムでのまたはいずれか一方の冗長性

10

20

30

40

50

をブロックレベルの比較によって解消する能力を提供する。

【0012】

一実施態様では、本発明の即時利用性に関する態様は、

(a)ソース(「第一次」)ファイルシステムの第二次システムに記憶されるベースレベル(base-level)のスナップショットを提供し、

(b)第二次システムに記憶され、以前のスナップショット以降に変更があったブロックのみを表現する第一次システムのブロックレベル(block-level)のインクリメンタル(Incremental)スナップショットを提供し、かつ

(c)そのインクリメンタル(Incremental)スナップショット画像の少なくとも1つから、組込まれるストレージユニットとして直接使用することができる論理ディスク画像を作成する(ただし、ステップ(b)のインクリメンタル(Incremental)スナップショットはこのステップの即時実施を容易にする方法で必要に応じて作成されている)ことによって、提供される。

10

【0013】

本発明のスナップショットおよび即時利用性の特徴は、ストレージハードウェアコンポーネントに関連し、本発明に従って作成されたバックアップイメージ用の「企業の画像移動先(Enterprise Image Destination(EID))」に提供するために使用される。さらに、EIDソフトウェアは、さまざまなベンダーのストレージハードウェア(安価なATAストレージハードウェアを含む)を使用して操作可能であることが特徴である。「二重保護」機構が提供され、それによってEID内のポイントインタイム(point-in-time)画像自体が、選択された媒体にバックアップされてもよいし、その他のEIDに複製されてもよい。

20

【0014】

本発明は、「遅延ミラーリング」と呼ばれる機構を提供し、それによって代替の物理的な第一次ファシリティーが設置され第一次ソースファイルシステムとして第二次ストレージ装置と協働で動作することができる。この機構に従うこの第二次ストレージ装置は、上述に言及される本発明の「即時利用性」に従って予めオンラインで導入される第二次論理ボリュームであってもよい。この場合、代替の物理的な第一次ボリュームが同時に設置されることになる。または、第二次ストレージ装置は、(別の例として)ミラーリングされたストレージシステムの残存ユニットであってもよい。この場合、別のミラーユニットが同時に「再ミラー化」または置換される。

30

【0015】

「遅延ミラーリング」の技術のそれ以外の用途も可能である。さらに、そのような用途では、本発明による「遅延ミラーリング」は、処理時に割り込みを受けずに進行することができることを特徴とする。

【0016】

最後に、本発明は、ブロック比較に基づいて、部分的に共通のコンテンツ(オペレーティングシステムや共通データベースなど)を有する複数のシステムがバックアップされるときに、冗長データを削除することによって、分散したバックアップ動作を大幅に高速化するための技術を提供する。バックアップすべきブロックが既にバックアップセットに存在していると判定される場合、既存のブロックはバックアップのディレクトリまたはカタログに使用され、両方のブロックを記憶しない。同様の技術は、ファイルシステムの冗長ブロックを削除するために使用される。

40

【0017】

本発明のその他の目的および利益は、図面および以下の詳細な説明から明らかになる。

【発明を実施するための最良の形態】

【0018】

以下は、本発明の各種の態様に関する複数の好ましい実施形態であり、システムがどのように構成されて本発明を実行するかの詳細と、当該システムを利用し当該方法を実践するために使用することができるステップを示す。このような実施形態は、あくまでも説明

50



するためのものであり、本発明は、示される特定の例に一切限定されることはない。例えば、特定の好ましい実施形態は、ストレージハードウェアおよびオペレーティングシステムを併用する実施に関連して記載されるが、それに続く開示は、当業者が記載された教示内容を他のストレージハードウェアやオペレーティングシステムに容易に適用できるように意図されていることは理解されよう。具体的な実施例の特定の特徴は、請求することができる範囲を制限するものとして理解しないでいただきたい。

【0019】

(定義)

以下の用語は、本願明細書で使用されるように定義された意味がある。

【0020】

APM(高度保護マネージャ、Advanced Protection Manager)：本発明の実施態様を実施する一連の製品に使用される名称。

【0021】

APM2D(ディスクに対する高度保護マネージャ、Advanced Protection Manager to Disk)：現在利用可能な第二次装置と、永久的なブロックレベルインクリメンタル(incremental)および「即時利用性(Instant Availability)」を提供するシステムでの将来のソリューションを対象とする包括的用語。

【0022】

アプリケーション(application)：バックアップによって保護されるビジネスアプリケーション(通常はデータベース)への大量生産された(すなわち、商用認可された)バックエンド。これは、エンドユーザアプリケーションとは異なる(また、混用すべきではない)。

【0023】

アプリケーションインスタンス(application instance)：物理的マシン上でその他のインスタンスと共存するアプリケーションの論理的に別のインカーネーション(incarnation)。アプリケーションインスタンスは、FAR用のターゲットである。

【0024】

バックアップクライアント(backup client)：高速バックアップ用のブロックレベルのインクリメンタル(incremental)バックアップをその他の動作に実質的に影響させずに提供するクライアントソフトウェア。ディスクに直接アクセスし、極度に高速で効率的な画像ベースのバックアップ用にファイルシステムをバイパスする。バックアップクライアントは、Exchange 2000/2003およびSQLサーバ2000のデータベースのブロックレベルのインクリメンタル(incremental)バックアップ用にも提供される。

【0025】

BAR(復元後のバックアップ、backup after restore)：復元後の最初のバックアップも、インクリメンタル(incremental)バックアップ(incremental)であり、オリジナルのベースに関連付けられる。

【0026】

EID(企業の画像移動先、enterprise image destination)：アプリケーション認識の(application-aware)永久画像インクリメンタル(incremental)用のニアライン(nearline)の移動先およびリポジトリ。

【0027】

EOFM：Windows用のセントバーナード(St. Bernard)から入手されるスナップショットドライバの - OEMバージョン。

【0028】

アプリケーション用ERF(最終高速フェールバック)：アプリケーションをFAR用のターゲットノードからアプリケーション用の元のまたは新規に指定されたホームノードにフェールバックすることが望ましいことがある。これは、高速に、途切れることなく、最低限のアプリケーションダウンタイムで実施される。

【0029】

エクスプレスDR(express DR)：日常のバックアップから単純で強固なワンステップペア

10

20

30

40

50

メタルリカバリ(one-step bare metal recovery)をクライアントノード用に提供する。完全なシステム画像を複数のマシンに配布するために使用することもできる。

【0030】

エクスプレスイメージ(express image)：テープまたはストレージから独立したディスクへのシステムの高性能バックアップ用にブロックレベルの技術を利用する。小さなファイルが多数有る状態で(with many small files)大容量のバックアップ用に極めて優秀な性能ゲインを提供する。

【0031】

FAR(高速アプリケーション回復、Fast Application Recovery)：高速アプリケーション回復は、NAS装置のバックアップ画像から生成される仮想ストレージに帰属させることによって、待機サーバまたは元のサーバ上でアプリケーションをオンラインに迅速に移行する能力である。

10

【0032】

ファイラ(filer)：NAS装置。

【0033】

永久画像増分(never image incremental、「永久増分(never incremental)」および「永久的なブロックレベルのインクリメンタル(incremental)」も含める)：ベースレベルをシードバックし(to seed base level back)、その後で永久的にインクリメンタル(incremental)のブロックレベルのバックアップをスケジュールするための能力。

20

【0034】

即時利用性(instant availability)：バックアップデータセットを読み/書きボリュームとして迅速に組込むことを可能にする。重要なアプリケーションおよびデータを、データを転送せずにほぼ瞬間的に回復させる。

【0035】

iSCSI：TCP/IPベースのストレージプロトコル。認証されたイニシエータノードにアクセス可能なIPネットワーク上でリモートストレージを形成するためにファイバーチャネルに代わる低コストのもの。

【0036】

iSCSIマッピングおよびアンマッピング(iSCSI mapping and unmapping)：フィルタへのiSCSIロケインプロセスで、フィルタ上のLUNが復元ターゲットノード上のローカルストレージとして視覚化される。iSCSIログオフはこの処理を取り消し、これらディスクを削除する。

30

【0037】

LAR(復元後の寿命、life after store)：これは、FARボリュームを保護する価値がある場合には、FARボリュームのERFおよびバックアップの組合せである。

【0038】

LUNクローニング(LUNクローニング)：スナップショットがバックアップされたLUNをバックアップ中のスナップショットおよびノーマルLUNへの移行から解放することができるNASファイラの機能。この処理が完了する間、LUNをアプリケーションで使用することができる。その後、スナップショットを削除することができ、LUNは独立して存在する。

40

【0039】

LUN作成(LUN creation)：スナップショットに記憶されるバックアップイメージから仮想ストレージを形成するNASの機能。さらに、このようなLUNを復元ターゲット上で読み書き用に組込むことができる。読取りデータが、スナップショットから充足される一方で、書込みデータは別の永続領域に送信される。元のバックアップ画像は変化しない。

【0040】

オンライン/バックグラウンド復元(online/background restore)：FAR後のiSCSIドライブからローカルディスクに画像データを自動的にバックグラウンドでコピーするが、アプリケーションはオンラインのままである。アプリケーションが立ち上がって稼働する間、これは、バックグラウンドで目立たないように実行される。アプリケーションが休止ま

50

たは再起動され、iSCSIドライブがアンマップされる際に、短時間の同期化が最後に必要となる。処理が終わると、全てのデータはローカルになる。データ転送は発生するが、アプリケーションの機能停止またはダウンタイムという意味では、不利益は被らない。

【0041】

PIT画像(PIT image)：アプリケーションボリュームのポイントインタイム(point in time)画像であり、バックアップ時にフリーズされる(frozen at the time of backup)。

【0042】

プロトコルディレクタ(protocol director)：ブロックレベルのアプリケーションコンシステント(application-consistent)プロトコルを使用するジョブの実施を制御および管理する。

【0043】

第二次ストレージ(secondary storage)：第一次ストレージ(生成データが存在する場所)とは異なり、これは、バックアップ用の移動先である他、仮想マシンディスクを形成するLUN用の基盤でもある。ストレージを追加する必要しかないため、バックアップに必要なもの以外は、第二次ストレージにはほとんど必要がない。この第二次ストレージは、法的な要求事項に適用させるために変更不可の内容保持をサポートする追記型(write once read many(WORM))であってもよい。

【0044】

専用バックアップソフトウェア(specialized backup software)：これは、インクリメンタル(incremental)の変更を取得し第二次ストレージ上で過去のポイントインタイム(points in time)を確保するバックアップイメージを作成する。バックアップソフトウェアは、アプリケーションコンシステント(application consistent)なイメージを作成し、永続状態および揮発状態を含むマシンコンフィギュレーションを取得する。

【0045】

アプリケーションマネージャ(application manager)：使用し易いブラウザベースのGUIからすべてのブロックレベルのアプリケーションコンシステント(application consistent)なバックアップ動作を管理する。NAS装置に加えて、Windows、Unix、およびLinuxのノードのバックアップを支援する。また、GUIのSQLおよびExchangeのボリュームならびにデータベースを選択可能なバックアップおよび復元用に表示もする。すべてのバックアップおよびその他の動作は、単一のカタログに記録される。

【0046】

待機ノード(stand-by node) / 代替ノード(alternate node) / 予防的セットアップ(preventive setup)：高い可用性と検証上の理由からFARのターゲットとなり得る最低限のハードウェアおよびデフォルトアプリケーションをインストールしたマシン。業務の必要性に応じて、このモードは、永久的にアプリケーションを動作することができる強力なマシン(machine)にもなり得る。

【0047】

ボリューム(volume)：バックアップのユニット。ブロックレベルでバックアップされる多数のファイルおよびディレクトリを含む単一のファイルシステム。

【0048】

(企業の画像移動先を使用したエンドツーエンド保護)

企業の画像移動先(EID)は、APM(高度保護マネージャ)の一連の製品の一部である。この特徴は、ソフトウェア内で実施され、一旦ノードにインストールされた後は、そのノードがアプリケーションウェア(application aware)の永久画像インクリメンタル(incremental)用のニアライン(nearline)移動先として機能することができるようになる。このEIDノードは、さまざまな方法(ローカルディスク、iSCSIストレージなど)で構成され、さまざまな程度の保護および信頼性を提供する。各種ノードからの画像バックアップは、この仕組みで統合、ニアライン化(nearlined)、およびバージョン化(versioned)される。ファイルシステムおよびアプリケーションの即時利用性は、このようなバージョン化された画像から影響を受けない。

10

20

30

40

50

## 【 0 0 4 9 】

図 1 は、本発明の一実施形態の典型的な企業用配置を示す図であり、安価な SATA ディスクドライブを利用する第二次ストレージサーバ 107 と、さらにそれに順次接続されているサーバ 103 のアレ、NAS デバイス 104、リモート第二次ストレージ装置 105、およびテープストレージ 106 を示す。このバックアップ用配列は、小規模リモートサイト 101、大規模リモートサイト 102 の両方を含むネットワーク用のバックアップおよび回復をリモートで管理するために使用される。ブロックレベルのバックアップクライアントは、参照番号 111、112、114、115 で指示されるブロックレベルのバックアップ動作を実施するために使用される。第三次ストレージ 113 (第二次サーバ 107 も第三次ストレージとして役立つ) とテープ 116 (テープドライブ 106 へ) への複製も示す。図 1 に示す種々の要素およびバックアップステップを、以下に続くこの開示のセクションでさらに説明する。

10

## 【 0 0 5 0 】

(アーキテクチャ：)

(ベーシック(Basic)：)

EID ノードは、ホットプラグ可能な RAID として構成されるようなローカルに取付けられる SATA ドライブを有する。このストレージは、画像用のリポジトリとして使用される。バージョン化は、システム (Win 2003 用 VSS または Linux 用 LVM/EVMS) 上で利用可能なスナップショット経由で実施される。画像は、バンドルされた iSCSI ターゲットソフトウェア経由で読み書き LUN としてエクスポートされる。

20

## 【 0 0 5 1 】

(シン(Thin)：)

EID ノードは、小型ローカルドライブ (理想的にはミラー化されている) のみを有し、OS および EID ソフトウェアを保持する。バックエンド iSCSI ストレージアレ (または同様のネットワークインテリジェンス) は、バックアップイメージ用の実際の移動先として使用される。ストレージアレは、LUN 作成、スナップショット作成、LUN クローニング、LUN マスキング / アンマスキング (un-masking) 機能をエクスポートしてバンドルされた EID ソリューションに参加させる候補にするために必ず必要である。VSS/VDS または SMI-S API を使用して EID ソフトウェアと外部ストレージの間のインタフェース上で標準化してもよい。

30

## 【 0 0 5 2 】

(シンシェアード(Thin Shared)：)

これは、上述の変更形態であり、ネットワークストレージアレが、ソースマシンと EID ノードとで共有される場合である。ソースと移動先とでスナップショットを共有することによって、バックアップをこの構成で最適化することができる。EID ノードは、この構成ではバックアップヘッドとして作用する。

## 【 0 0 5 3 】

(二重保護を使用する EID：)

バックアップをテープまたはディスクのさらなるバックアップによって保護する必要がある。これは、二重保護と呼ばれている (二重保護に関する文書を参照のこと)。EID ノード上でのディスクへの第一のバックアップを、SAN 上のテープ装置または第一のバックアップが存在するストレージとは異なるその他のディスクに送ることができる。これは、SAN に存在するかまたはある種のリモートアプライアンスに取付けられる第二層または第三層のストレージである (可能であれば、別の EID ノード)。このため、EID は多層ストレージに基づくデータ保護に対するエンドツーエンドソリューションにとって成功へのカギとなる。

40

## 【 0 0 5 4 】

(構成)

(APM クライアントノード：)

このようなノードは、複数のスナップショットプロバイダの APM クライアントおよびサ

50

ポートを使用して構成される(利用可能な場合)。APMクライアントは、EIDターゲットを第二次ストレージにバックアップすることができる。また、第二次ストレージは、ベンダー供給ストレージハードウェアまたは汎用ATAストレージハードウェアにバックアップすることができる。スナップショットサポートは、ベーシック(basic、バンドルされたEOFM)またはコンプレックス(complex)であってよい。尚、各ボリュームは、別のスナップショットプロバイダを有することができる(複数のプロバイダが存在する時は、それらの使用が予め構成されるか、またはEIDノードによって指示されなければならない)。アプリケーションサポートが実施されるときには、第二次ターゲットおよびEIDターゲットの両方に同時に利用できる。

【 0 0 5 5 】

(APMサーバ - EIDノード : )

このノードは、バックエンドiSCSIストレージ(ある場合)に依存して、ストレージ特有のプラグインを使用してインストールされるEIDソフトウェアを有する。プラグイン構成は、ライセンス情報と共にインストール中にハードワイヤードされる。基本構成は、2つの異なるOSすなわち、Windows 2003と、LVM/EVMSを使用するext3fs/xfsを備えたLinux 2.6である(2003/NTFS and Linux 2.6 with ext3fs/xfs with LVM/EVMS)。複数の永久スナップショットを使用するスペースファイルサポートによる64ビットのジャーナルファイルシステムが本質的に必要である。このような基準に適合するシステムは、EIDノードの候補となり得る(圧縮や暗号化などのファイルシステムの付加的なプロパティは、必須ではないが、付加的なコンプレキシティ(complexity)およびオーバーヘッドまたはいずれか一方に付加的な特徴を加えるために使用してもよい)。

【 0 0 5 6 】

(バックアップフロー : )

図 2 は、ポイントインタイム(point-in-time)スナップショット、ブロックレベルのインクリメンタル(incremental)バックアップ、およびポイントインタイム(point-in-time)のフルボリューム画像の作成の他、ファイルレベルの復元動作も図式的に示す図である。

【 0 0 5 7 】

(スナップショットフェーズ : )

プロトコルディレクタ(protocol director)は、APPH(アプリケーションヘルパ、アプリケーション(SQLサーバ、Exchangeなど)特有のバックアップ開始時および終了時の対話を仲介する)とBACKUP\_PREPAREとのコンタクトを取る。APPHは、スナップショットハンドラとコンタクトを取り、スナップショットハンドラは、スナップショットコードとインクリメンタル(incremental)ブロックトラッキングインタフェース(block tracking interface)とをカプセル化し、一連のボリュームのスナップショットを作成し、変更ジャーナルをフラッシュ(flush)する。スナップショットハンドラは、ファイルシステムディスクバリアの一環としてDISCOVER\_LUNSを実行する。複数のLUNが、支援されるiSCSI(またはFCP(ファイバチャネルプロトコル、fibre channel protocol))ベンダーによってバックエンド化されるのを検出すると、これは、ベンダー特有の方法を呼び出し、同じiSCSIストレージエンティティ上に存在するボリュームセット(例えば、ストレージ装置上の一連のLUNを含むボリューム)のスナップショットを取る。専用プロバイダは、この機能性を提供するストレージベンダーごとに存在するが、ストレージベンダーから入手可能な場合には、VSSもしくはSMI-Sプロバイダを使用してもよい。付加的な構成情報は、この機能性(データベースから取得する必要がある)を提供するためにバックエンドストレージノードごとに必要とされる(この情報はローカル構成ファイルの一部としてキャッシュまたはセーブしてもよい)。大部分の外部プロバイダが、外部のどちらにも(またはVSS仲介で)変更ジャーナルの支援を提供しないため、バンドルされたEOFMスナップショットを取る必要がある。EOFMスナップショットは、変更ジャーナルのフラッシュ(flush)および変更されたブロックのトラッキングに単独で使用される。外部スナップショットは、現実のバックアップインスタンスまたはリモートコピー用の一貫性のあるソースを表現する。EOFMスナップショットを最初にとった後に外部スナップショットを取って、一貫性のある画像を生成する。プロ

10

20

30

40

50

ックが変わる場合には、小さなウィンドウが、両方のスナップショットの間に存在する。アプリケーションは、すでに休止されているので(アプリケーションの状態は、アプリケーションがバックアップを開始し、それによってそのトランザクションをディスクにフラッシュ(flush)したことを認識するように、APPH経由で仲介されている)、I/Oはそのため生成されるべきではない。ファイルシステムメタデータは、どちらも変更すべきではない(ファイルシステムは常時クラッシュ一貫性のある状態に回復することができる)。独立したファイルは、次のインクリメンタル(incremental)まで捕捉されない変更されたブロックを有してもよい。このウィンドウは小さいものであり、一貫性のない状態を有するサポートされないアプリケーションの残りのウィンドウは、極端に小さいことに、留意されたい。

10

## 【0058】

APPHは、プロセスの最後にバックアップ詳細用のコンテンツファイルを作成する。このファイルは、EOFM、VSS、または第三者のプロバイダーによって作成されたローカルのスナップショットボリュームと一緒に、場合により論理名および永久的なスナップショットIDを有するベンダー特有の情報をを用いて増補される。

(データ転送：)

## 【0059】

SVHは、EIDソフトウェアとCREATE\_RELATIONSHIPメッセージ(第一のバックアップ用)とのコンタクトを取り、ソースパスとしてコンテンツファイルを渡す。

20

## 【0060】

次に、EIDノード上でEIDソフトウェアは、ソースノード上でそれに対応するソフトウェア(「ノードウェア」との接続を確立し、コンテンツファイルパスを渡す。その後、ソース側のノードソフトウェアがコンテンツファイルの内容を読み取り、EIDノード上のEIDソフトウェアに戻す。

(変更形態I：共有スナップショット＝バックアップ)

## 【0061】

EIDソフトウェアは、ベンダー特有のスナップショット情報を試験し、ベンダーがサポートおよび認可されているかを決定する。回答がイエスであれば、EIDソフトウェアは、ローカルクエリ経由で共有ストレージにスナップショットが存在していると決定し、それが共有スナップショットをバックアップとして使用できると決定した場合には、プロセスが完了する。アロケーションビットマップもこの時点で取得される。EIDソフトウェアは、その関係、すなわち、ソースノードとソースドライブ(または独自のID)と移動先ノードとlun名の組合せを、そのローカルデータベースに記憶する。アロケーションビットマップも保存され、スナップショットIDによってインデックスされる。

30

## 【0062】

(EIDノード上のスナップショット：)

SVHからのCREATE\_SNAPSHOTは、以前のステップで共有スナップショットと共に戻る。

## 【0063】

(エラー回復：)

このシナリオには必要ない。

40

## 【0064】

(キャンセル後の再起動：)

バックアップは非常に高速であるべきなので必要ない。

## 【0065】

(ファイル履歴：)

ファイル履歴は、バックアップLUNを使用してEIDノード上で生成される(任意)。このファイル履歴は、ある種の実施特有の方法でバックアップエクスプレスマスタ(backup express master)サーバに伝達されるべきである。

## 【0066】

(インクリメンタル(incremental)バックアップ：)

50

これは、変更ジャーナルがEIDノードにそのまま渡されるという事実を除き、ベースバックアップと同じ方法で進行する。尚、さらに、EIDノードは変更ジャーナル(CJ)をスナップショットIDによってインデックスされるローカルデータベースに格納する。

【0067】

(チェックサム：)

チェックサムをLUN画像上のすべての割り当てられたブロックについて算出し、スナップショットIDによってインデックスされるEIDデータベースに保存してもよい。チェックサムは、以下の3つの理由で重要である。

1. 書込み後の検証能力。
2. 信頼性のあるチェックポイント再起動の支援。
3. インクリメンタル(incremental)バックアップをブロックレベルトラッキングと共に実施する能力(ただし、コストは増加する)。

10

【0068】

(APPS：)

APPSボリュームは、スナップショットが取られた後のライブファイルシステム上でPOST\_BACKUPイベントの一部として生成されるファイルとを含む。これらのファイルは、共有スナップショットには存在しない。これらのファイルは、独立してバックアップされなければならない。変更形態IIの「バックアップLUNへのローカルコピー」は、このケースで使用しなければならない。APPSは、バーチャルボリュームとして出現するが、APPSのバックアップは、ファイル全体をコピーすることによって達成され(ファイルごとのバックアップ)、ボリューム指向のブロックコピーでは達成されない。

20

(変更形態II：バックアップLUNへのローカルコピー)

【0069】

EIDソフトウェアが共有スナップショットを使用できないと判定した場合、EIDソフトウェアは、iSCSIストレージ上にまたはローカルにバックアップLUNを作成し、それにソースノードとドライブIDを組合わせた独自の名前を付ける。ホスト名(hostname)とポートID(portid)とターゲット名(targetname)とルニド(lunid)が初期ハンドシェイクの一部としてソースEIDソフトウェアに戻される。

【0070】

次に、ソース側のノードソフトウェアは、MPA\_LUN(iSCSIロゲインを間接的に使用する)を、EIDノードから渡された情報と一緒に呼び出す。MAP\_LUNは、ローカル名スペースにマップされる装置をエクスポートする。ノードソフトウェアは、割り当てられたブロックを装置のローカルスナップショットからiSCSIマップされた装置にコピーし始める。このプロセス中には、ノードソフトウェアは、ステータス/チェックサム/進行状況をすでに確立されたチャンネル経由でEIDソフトウェアに渡す。

30

【0071】

(EIDノード上でのスナップショット：)

EIDソフトウェアは、バックアップLUNまたはある種のカバーリングエンティティのスナップショットを取り、スナップショットIDを戻す。

【0072】

(エラー回復：)

データ転送用のiSCSI接続に信頼性があり、内蔵されたりカバリアンドエラー(recovery and error)接続を有するので、必要にならないであろう。EIDソフトウェアは、DMAにトランスペアレントな制御接続部上でエラーから回復することができる。

40

【0073】

(キャンセル後の再起動)

これは、実施しなければならない。EIDソフトウェアは、書込みが成功した最新のブロックを記憶して、初期ハンドシェイク時にこれが中止されたバックアップを再起動する一貫であることを示しながら、そのブロックを渡さなければならない。

【0074】

50

## (ファイル履歴)

ファイル履歴は、バックアップLUNを使用してEIDノード上で生成される(任意)。このファイル履歴は、ある種の実施に特有の方法でバックアップエクスプレスマスタ(backup express master)に伝達されるべきである。

## 【0075】

(インクリメンタル(incremental)バックアップ：)

このようなバックアップは、変更ジャーナルがローカルで使用されて、変更ブロックのみをプロセスの一部としてバックアップLUNにコピーする事実を除いて、ベースバックアップと同じ方法で進行する。

## 【0076】

(チェックサム：)

チェックサムは、LUN画像上のすべての割当てられブロックに対して算出され、スナップショットIDによってインデックスされるEIDデータベースに保存される。

## 【0077】

(APPS：)

APPSボリュームは、スナップショットが取られた後のライブファイルシステム上にPOST\_BACKUPイベントとの一部として生成されるファイルを含む。このようなファイルは、バックアップスナップショットには存在しない。APPS LUNがローカルにマップされた後、ローカルで認識されたファイルシステムとしてフォーマットされなければならない。次に、APPSディレクトリ/ファイルは、APPHが指示した位置(スナップショットからではない)から、その全体が(ファイルごとに)APPSバックアップLUNにコピーされる。インクリメンタルバックアップ(incremental backup)時にAPPS LUNは、クリアにされて、新規のAPPSファイルセットがコピーされなければならない(古いスナップショットは、以前のバージョンのAPPファイルを維持する)。

## 【0078】

(変更形態III：ネットワークコピー)

変更形態IIと同様に、EIDソフトウェアが共有スナップショットを使用できないと決定した場合には、EIDソフトウェアは、バックアップLUNをiSCSIストレージ上にまたはローカルで作成し、それにソースノードとドライブIDを組合わせた独自の名称を付ける。LUN作成は、このノード(実際には基本構成)に支援されない場合には、履行されない。もし、不履行が起これば、ホストネーム(hostname)とポートID(portid)とターゲットネーム(targetname)とLUNID(lunid)は、ソースノードソフトウェアに、初期ハンドシェイクの一部として戻されず、変更形態IIIが指示される。

## 【0079】

変更形態IIIが指示されるか、または、iSCSIもしくはその他にソースノード上にLUNマッピングサポート手段がない場合には、ソース側のノードソフトウェアは、装置のローカルスナップショットから割当てられたブロックの読取を開始し、ネットワーク経由で移動先のEIDソフトウェアにその読取を伝送する。移動先のEIDソフトウェアは、チャンネルから読取り、スパーファイルを移動先のある所定のボリュームに書込む。このプロセスのどちらのエンドがチェックサムを生成しても良い。

## 【0080】

(EIDノード上のスナップショット：)

EIDソフトウェアは、バックアップイメージファイルを含むボリュームのスナップショットを取り、スナップショットIDをDMAに戻す。

## 【0081】

(エラー回復：)

移動先に維持されるチェックポイント経由でネットワークの停止から回復するために必要である。

## 【0082】

(再起動/キャンセル後の再起動)



これは、実施する必要がある。EIDソフトウェアは書込みに成功した最新のブロックを記憶し、初期のハンドシェイク時にこれが中止されたバックアップを再起動する一貫であることを示しながら、このブロックを渡す必要がある。

【0083】

(ファイル履歴)：ファイル履歴は、バックアップ画像を使用してEIDノード上で生成される(任意)。

【0084】

(インクリメンタル(incremental)バックアップ：)

このバックアップは、変更ジャーナルがローカルで使用されて変更ブロックのみを讀取ってから、ネットワーク上にそのブロックを転送し、移動先上のバックアップ画像を更新するという事実を除き、ベースバックアップと同じ方法で進行する。

10

【0085】

(チェックサム：)

チェックサムは、バックアップ画像上のすべての割当てられた/変更されたブロックに対して算出され、スナップショットIDによってインデックスされるEIDデータベースに保存される。

【0086】

(APPS：)

APPSディレクトリ/ファイルは、APPHに指示された位置から(スナップショットからではない)その全体が(ファイルごとに)讀取られ、ネットワークを渡って移動先のEIDソフトウェアにコピーされ、そのEIDソフトウェアでは、ディレクトリ構造が(所定のバックアップディレクトリ位置に基づいて)作成され、そのソースのファイルと同一のコピーを反映する。インクリメンタル(incremental)バックアップ中に、APPSディレクトリは、クリアにされて、新規のAPPSファイルセットがソースから転送され、再度作成されなければならない(古いスナップショットは、以前のバージョンのAPPSファイルを維持する)。

20

(外部LUN/スナップショット管理用のプラグインアーキテクチャ)

【0087】

EIDバックアップは、スナップショット作成、LUN作成、LUNクローニングなどに依存する。バックアッププロセスのソース側もEID側も、このようなサービスの消費者である。容易にアーキテクチャを分離し各種のベンダーをプラグインすることができるためには、関連するベンダー特有のプロバイダとのインタフェース(DLLまたは共有ライブラリの形態)を実装する必要がある。デフォルト実装には、EIDノード上のバンドルされたiSCSIプロバイダを使用するが、保証があればベンダー特有の実装と置き換えてもよい。インタフェースは、汎用LUN作成/削除、LUNクローニング、スナップショット作成/削除の機能性を提供する。増補バージョンのインタフェースは、ブロックレベルのミラーリング用の機能性やその他の注目すべき機能(例えば、第二次から第三次への複製機能など)を追加することもできよう。それらの機能を、効率的な/洗練された二重保護の方法論を支援するための利益に活用してもよい。

30

【0088】

(EIDデータベース：)

EIDノード上の小規模データベースは、構成情報(バックエンドのiSCSIストレージなど)、認可情報、スナップショットID情報、チェックサム情報などを維持するために必要である。このデータベースは、EIDノードがある種のiSCSI/共有SANストレージをバックエンド化している場合に、特に必要である。バックアップエクスプレス(Backup Express)インフラストラクチャは、独自のスナップショットIDを扱っているが、EIDソフトウェアは、ローカルデータベース経由でスナップショットIDを逆参照することによって、これを正確なネットワークエンティティに翻訳しなければならない。

40

【0089】

単純な実装は、ブロックアロケーションビットマップ、インクリメンタル(incremental)ビットマップ、チェックサム、ファイル履歴などを含むスナップショットIDを使用して

50

名付けられたディレクトリのセットであってもよい。

【0090】

(テープに対する二重保護：)

これは、SSSVHからジョブハンドラにリダイレクトされる正規のNDMP(ネットワークデータ管理プロトコル、Network Data Management Protocol)バックアップ経路で実施される(別の二重保護の説明を参照のこと)。テープに対する二重保護について留意すべき重要な点は、最初のバックアップの十分な/完全な画像が、テープ上に作成されることである。それに続くテープのバックアップは、他の最初のバックアップインスタンスの十分なコピーである。インクリメンタル(incremental)に関する概念、またはあるテープのバックアップ画像を別の画像に関連付ける方法に関する概念は、この設計の一部ではない。

10

【0091】

(ディスクに対する二重保護：)

ディスクに対する二重保護(DP2D)は、元の/最初のバックアップからディスク上に別のバックアップを作成することによって、ディスク上でのバックアップイメージの寿命を延長する。このケースのあらゆる努力は、インクリメンタル(incremental)データを転送し第三次のバックアップを更新することによって、後続のバックアップを作成することに向けられている。以下に、様々なシナリオを挙げる。

【0092】

(EIDノードに見える多層ストレージ：)

このシナリオでは、第三次ディスクストレージは、EIDノードからアクセス可能である(第二次および第三次のストレージは、一定の単一ベンダーインタフェース(Hitachi Tagma Store)を経由してアクセスされる大規模な多層ストレージ配置の一部であってもよい)。このケースでは、二重保護バックアップは、適切な第三次の位置が選択されて、LUNがローカルEIDノード上でアンマスクまたは組込まれた後で、EIDソフトウェアによって履行されるローカルのブロックレベルのインクリメンタル(incremental)コピーを経由して進行する。

20

【0093】

(単一ベンダーのノード間のブロックミラーリング)

ベンダーが、第二次ノードと第三次ノードの間でデータを転送するために効率的でアプライアンスを実装したブロックミラーリング法を有する場合には、EIDソフトウェアは、ベンダー特有のAPIセット経路で画像転送/更新を始動し、二重保護バックアップを作成する。

30

【0094】

(EIDノード~EIDノード：)

第三次ストレージが、EIDノードから物理的に離れているときには、リモートのEIDノードは、「ネットワークコピー」経路でバックアップを始動し、ローカルEIDノードからデータを引き出す。

【0095】

(EIDノード~第二次ノード：)

データをEIDノードと第二次ノードの間で転送しなければならないときには、適用可能なバックアップクライアント転送法を使用する。すなわち、第二次ノードに連絡しEIDノードからデータを引出すように要求する。EIDソフトウェアは、DP2Dバックアップを認識し、保存されたビットマップを使用して第二次画像を適切な(通常は最新の)スナップショットから更新する。

40

【0096】

(バックアップメカニズム：)

一旦、二重保護ジョブが、最初のバックアップを保護するために作成されると、プロトコルディレクタは、スナップショットフェーズをスキップする以外は、正規のEIDバックアップとほとんど同様にEIDバックアップを始動する。

【0097】

50

CREATE\_RELATIONSHIPは、移動先のEIDソフトウェアに送られる(これは、移動先に相関するEIDノード、リモートノード、または別の二次タイプであってもよい)。EIDソフトウェアは、それがバックアップ用のソースノードであることを検出した場合には、適切なメカニズムを使用して、第三次移動先にローカルに画像をコピーする(バックアップと共に保存された割当てられたビットマップまたはインクリメンタル(incremental)ビットマップを使用する)か、または、ベンダー特有の方法を呼出してこの転送に影響を与えるかのいずれかを行う。EIDソフトウェアが、ソースがリモートであることを検出する場合、EIDソフトウェアは、既に説明されたメカニズムを用いて正規のEIDバックアップを始動する。バックアップは、正規のEIDバックアップと同様に移動先のEIDノード上に保存されるが、これはこのプロセスが無限にカスケード可能であることを含意する。

10

## 【0098】

スナップショットIDは、第二次スナップショット作成に対するNOTIFY応答から戻り、二重保護バックアップの一部としてカタログに登録され、元の最初のバックアップにリンクされる(詳細な説明については、二重保護に関する別の説明を参照のこと)。

## 【0099】

(二重保護バックアップからの復元：)

二重保護に関する記載を参照：

## 【0100】

(復元ブラウズ(restore browse)：)

ファイル履歴が、EIDノード上でバックアップの最後に生成され、バックアップエクスプレス(backup express)データベースに取込まれるときに、ブラウジングが通常はデータベースのカタログブラウズ機能を介して発生する。ファイル履歴が生成されないとき(生成中のファイル履歴がコンピュータ的に集中的であるか、または過度に多いストレージ領域を必要とするとき)、EIDソフトウェアと連絡を取ることによってNDMPディレクトリブラウズ機能を使用することができる。ブラウジングは、EIDノード上にバックアップLUNを組込んでから、既存の「スナップdirリスト(snap dir list)」メカニズムを使用してファイルシステムをブラウジングするか、または、ブラウジングが必要なときに画像ファイルから「rawtoc」を生成することによって、提供されてもよい。テープに対する二重保護では、認識可能なファイルシステムとしてLUNを組込むオプションが利用できない場合には、ファイル履歴が二重保護動作中に画像の一部としてまたはファイルごとのアーカイバルフォーマットを構築するように生成する必要がある。

20

30

(復元フロー)

## 【0101】

(ディレクトリ/ファイルの復元：)

一旦、復元選択(restore selection)が(インスタンス用のバックアップドキュメントがAPPHによってアプリケーションオブジェクトからファイルに翻訳された後に、ユーザまたはプロトコルディレクタによって)生成されて、コンテンツファイルが作成されると、SSSVHは、復元ターゲット上のノードソフトウェアとコンタクトを取り、それにコンテンツファイル(EIDノードはこのコンテンツファイルからデータを取る)、ノード上のパスおよびスナップショットのIDを渡す。次に、復元ターゲット上のノードソフトウェアは、EIDソフトウェアとコンタクトを取り、復元パスとスナップショットを渡す。一旦EIDノードは、この情報を検査すると、スナップショットIDとボリュームの組み合わせを復元ターゲット上でLUNとしてエクスポートできるかどうかを決定する。これが可能である場合、(バックアップとほぼ同様に)LUNがEIDノードによってローカルにまたは共有SANストレージ上のいずれかで作成され、ホスト名前(hostname)とポートID(portid)とターゲット名前(targetname)とLUNID(lunid)が復元ターゲットに渡される(注：ホスト名はEIDノードと同じであってもよい)。一旦復元ターゲット上のノードソフトウェアがこのLUNをマップすることができると、ハンドシェイクは完了する。即時利用性については、これが本質的に復元プロセスを完了する。それ以外の場合には、ノードソフトウェアがマップされたLUNから復元ターゲット位置にファイル/ディレクトリのローカ

40

50

ルコピーを実施する(注：これは、APPSファイルが論理的にバックアップされる方法とまったく同様である)。

【0102】

(フォールバック：)

EIDノードは、LUNが(セキュリティの理由などから)要求中のノードにエクスポートできない、または、初期ハンドシェイクが完了後に要求中のノードがLUNをマップすることができないと決定することは可能である。この状況(低い優先順位)では、従来の復元が進行し、EIDソフトウェアが要求されたファイルをバックアップ画像から読取り、そのファイルをネットワーク上に送り、復元ターゲット上のノードソフトウェアが受信したデータからローカルにファイルを再度作成する。この状況では、「rawtoc」は、ポストバックアッププロセスから既存のものか、または、復元用にオンザフライ(on the fly)で作成された(さらに望む場合にはキャッシュされた)ものが要求される。

10

【0103】

(エラー回復/再起動性)

これは、LUNマップされた/IAスタイルの復元に不要であるが、(実装される限りは)従来の復元に有用になる場合がある。

【0104】

(即時利用性復元：)

その他のブロックレベルの復元と同様に、MAP\_LUNは選択されたスナップショットから復元ターゲット上にiSCSIまたはFCP経由でボリュームセットをマップするために(スナップショットハンドラに実装されるものとして)呼出される。スナップショットハンドラは、EIDノード上のCREATE\_LUN\_FROM\_LUNを呼出して、スナップショット内のLUNを作成しエクスポートする。次に、APPSボリュームは、ローカルiSCSIマウントまたはネットワークマウントのいずれかを介してローカル名スペースに同様にマップされる。一旦このステップが完了すると、SSSVHがAPPHにその復元を完了するよう指示する。APPHは、必要な場合にAPPSボリュームからIAボリュームにログファイルをコピーしてアプリケーションまたはデータベースを回復させる。EIDソフトウェアはIA復元に関してまったくコンタクトされていないことに留意されたい。

20

【0105】

差分ブロックレベルの画像の一部としてネットワークを渡って伝送されるバックアップデータは、最初に付けられたディスク署名を有し、単一の有効なパーティションを使用してSCSIディスク全体としてボリュームのバックアップを仮想化するために適切な情報を有する。

30

【0106】

復元中に、この読取り専用画像は、スナップショット内の画像によってバック(back)されるスパーファイルを作成することによってiSCSIアドレス可能な読み書きLUNに変換される。LUNファイルは、永続的であり、バックアップ画像からの変更や元の未変更データを集約する第一次ストレージとして機能することができる。LUNは共に、独立型ディスクまたはRAIDセットの一部として組込むことができる。

【0107】

(エラー回復/再機動性)

該当なし。

40

【0108】

(ローカルボリュームロールバックを介する復元：)

ボリュームロールバックは、バックアップ時間が存在するため、復元が元の位置およびすべての変更ジャーナルに対して行われる場合に限り可能である。このような基準が作られない場合には、フルボリューム復元を始動することができる(これは、いずれにしても、ボリュームロールバックの悪化したケースである)か、または、復元ジョブが履行されない(IA復元の機能性を考慮すれば、これはまったく実装される必要がない)。

【0109】

50

あるオプションは、ボリュームロールバックが所望されていることを指示するが、その場合、VOLUME\_ROLLBACKメッセージがプロトコルディレクタによってスナップショットハンドラに(MAP\_LUNとほぼ同様に)伝送される。このメッセージは、バックアップjobid(バックアップのポイントインタイム(point-in-time)を独自に特定する)と、問題のボリュームを含む。ボリュームロールバックが可能な場合、スナップショットハンドラは、問題のボリュームをロックし、ディスマウントし(ボリュームによってホストされるアプリケーションはAPPHによってシャットダウンされるかまたはオフラインにされる)、さらに、スナップショットを取って変更ジャーナルをフラッシュ(flush)する。スナップショットが復元されている時以降のすべての変更ジャーナルは、論理的にANDされて、ビットマップファイルを作成し、ビットマップファイルはプロトコルディレクタに戻される(ファイル名のみ)。プロトコルディレクタは、ビットマップファイルをコンテンツファイルに追加し、これをEIDソフトウェアに渡す。EIDソフトウェアは、ビットマップファイルを使用してマップされたLUNからのまたはネットワークを渡るブロックのセットのみを復元する。

10

【0110】

従来のフルボリューム復元が実装される場合、アロケーションビットマップは、割当てられたブロックのみをコピーするようにEIDノードから復元ターゲット上のノードソフトウェアに渡されなければならない。ネットワークコピーが使用される場合は、EIDノードがすでにどのブロックを送るべきかを認知している。

【0111】

復元完了後、ボリュームはローカル名スペース内にアンロックおよびリマップされ、アプリケーション/データベースが再起動されオンラインされる。

20

【0112】

(シンシエードコンフィギュレーションでのボリュームロールバックを介する復元)

この復元モードは、単一ファイルのバックエンドストレージのサポートまたはLUNロールバックを必要とする。

【0113】

ボリュームロックおよびアプリケーションシャットダウンは、上述とまったく同じにスナップショットハンドラおよびAPPHによって仲介される復元ターゲットノード上で発生する。

【0114】

ボリュームロールバックに対する初期ハンドシェイク時に、復元ターゲットはターゲットボリュームに関して対象となる情報(例:D:ファイラA, vol3/lun2)をEIDソフトウェアに渡す。バックエンドストレージがこの特徴を支援すること、およびスナップショットと復元ターゲットLUNが論理的に関連することを決定する時に、EIDソフトウェアは、バックエンドAPI(プラグインインタフェースの一部)を2つの引数と共に呼出す(2つの引数とは、復元中のスナップショットと、復元ターゲットノード上でボリュームをバックエンド化するターゲット論理エンティティまたはLUNである)。

30

【0115】

バックエンドストレージ上のボリュームロールバックは非同期に起こるが、ライブファイルシステムとスナップショットの相違によっては時間がしばらくかかることもある(しかし、ローカルコピーしか関連しないので、早急なほうがよい)。一旦これが完了すると、両方の復元エンドおよびアプリケーションを再起動することができる(このシナリオの例は、NAS装置上の単一ファイルLUNスナップショットのリバート(revert)である)。

40

【0116】

(エラー回復/再起動性:)

フルボリューム復元: 大規模なフルボリューム復元に限り重要である。バックアップに似た再起動メカニズムによって実施されるが、チェックポイントは復元ターゲットノードソフトウェアによって追跡され、再接続時に通信される。復元にDMAによるキャンセル後の再起動が必要であるか否かは、この文書の範囲外である。

【0117】

50

(ローカルボリュームロールバック：)

復元にローカルコピーが関わるので、エラー回復は不要であるべきである。キャンセル / 一時停止後の再起動性が望まれる場合もある。

【0118】

アプリケーション支援ボリュームロールバック：エラー復元は、不要であるべきであるが、バックエンドストレージが再起動を支援する場合には再起動性を備えるべきである。

【0119】

(エクスプレスDR(express DR)復元：)

これは、復元ターゲットがLinuxを稼動している場合のフルボリューム復元の特殊なケースである。Linuxノードソフトウェアは、修正バージョンのjndmpcによって駆動され、上述とまったく同じに動作し、カスタムLinuxカーネル上で利用できる場合には、SCSIイニシエータを活用する。エラー回復 / 再起動性は、この状況では必要不可欠である。さらに標準的なメカニズムが、所定のノードのエクスプレスDR(express DR)バックアップ用のスナップショットをブラウジングするために存在する必要がある。これは、EIDソフトウェアまたはEIDノード上のスナップショットハンドラによってエクスポートされるインタフェース上の一部であるべきである。スナップショットディレクトリのリスト化は、スナップショット用に予め定義された命名法で十分なこともあるが、適切なインタフェースが適合するスナップショットを列挙するために定義されなければならないこともある。

【0120】

(エラー復元 / 再起動性)

これは、大規模の復元に非常に望ましく、フルボリューム復元と同様に実施されるべきである。

【0121】

(セキュリティ / 仮想化 / コンプライアンス / セルフプロビジョンド(self-provisioned)復元)

ニアライン化されたデータは、オフライン媒体(テープ)上のデータよりも、安全でなければならない。この理由は、データが所定の許可を与えられているネットワーク上でまたは小規模のアカウントのセットが生きておりアクセス可能であるからである。あるいは、小規模のアカウントセットの情報が漏れる場合もある。一つのオプションは、ニアラインストレージに存在するデータを暗号化することである(利用可能な場合には、ネイティブファイルシステムの暗号化を使用してもよい)。これは、即時利用性の復元をシャットダウンするが、付加されるセキュリティは、その価値を高めることができる。ディスクおよびテープまたはいずれか一方に対する二重保護は、特に、長期にわたる場合には、アーカイブ上の理由が暗号化の最有力な候補である。

【0122】

少数のユーザアカウント(バックアップエクスプレス(backup express)adminおよびrootまたはEIDノード上の管理者)は、単一のEIDノード上で一元管理される複数のマシンのバックアップを保護してはいるが、多くの企業にとって十分に安全ではない。バックアップ画像のセットに関して責任 / 権利を各々有する複数のadminが、より望ましい場合もある(この状況では、スーパーユーザは必ずしもすべてのバックアップ画像に対して権利を有するとは限らない)。ある形式のRBAC(役割ベースのアクセス制御、Role based access control)は、Windows 2003またはLinux 2.6上の既存のセキュリティメカニズムを使用することによって、実施されてもよい。

【0123】

アプリケーションサーバの完全な画像は、EIDノード上にバックアップイメージとして保存されるので、これらの(過去の様々な異なるポイントインタイム(points-in-time)での)画像のセットは、仮想化の最有力候補である。各クライアントノードまたはアプリケーションサーバは、それがある種の既製のまたはOS依存の仮想化ソフトウェアを使用して過去のあるポイントインタイム(point-in-time)に現れる際に、仮想化することができる

10

20

30

40

50

。マシン状態を安全に仮想化する見込みがあれば(その場合には、権限のある人間のみがマシンデータにアクセス権を有する)、企業は、少ない管理者による復元、コンプライアンス、分析も、またはそれ以外の企業上の重要な理由から、ジャストインタイムの仮想化を実施することができる。

#### 【 0 1 2 4 】

規制コンプライアンスまたはリティゲーションディスカバリ(litigation discovery)は、EIDパラダイム(paradigm)の重要な用途であり、その場合、ほとんどコストを追加せずにコンプライアンス検査のために、EIDノード上のデータを過去のあるポイントインタイム(point-in-time)まで仮想化することができる。第二次WORMストレージまたはWORMテープと同様に、専用のコンプライアンスアプライアンスでターゲットにされるディスクまたはテープに対する二重保護によって、バックアップから始まり短期の復元および長期のアーカイブに至るエンドツーエンドのソリューションをコンプライアンス要件に適合させることができる。

10

#### 【 0 1 2 5 】

セルフプロビジョンド(self-provisioned)復元は、一般にエンドユーザがヘルプデスクまたは管理者の仲介無しにファイルを復元する場合に、少ない管理者によるデータ回復に関連する。これは、データが元のファイルシステムセキュリティを確保するEIDノード上に保存される際に可能である。一旦、即時利用性またはその他の技法を使用しある周知の位置に戻してボリュームをマップした後は、ユーザが既存の慣れたツールを使用してデータを検索し復元することができる(バックアップエクスプレス(backup express)GUIを使用して、管理者としてログインしなくてもデータを検索し復元することもできる)。EIDアーキテクチャ固有のプロパティによって、セルフプロビジョンド(self-provisioned)されたエンドユーザの復元は可能になり、このために、TCO(総所有コスト、Total Cost of Ownership)は大幅に削減される。

20

#### 【 0 1 2 6 】

(実施例)

図 3 は、ブロックレベルのインクリメンタル(incremental)バックアップとファイルレベルのインクリメンタル(incremental)復元の動作を、図 2 よりも詳細に、複数の上述の原理を説明する方法で示す図である。示される例には、以下のイベントおよび動作が関係する。

30

#### 【 0 1 2 7 】

午前 2 時：ベースバックアップが、早朝のバックアップウィンドウ時に第一次システム 3 0 0 で実施される。割当てられたブロック(3 0 1)のみが、バックアップされることに留意されたい。割当てられないブロック(3 2 0)は、第二次ストレージユニット 3 3 0 に転送されないの、経過時間を短縮し第二次ストレージ装置の要件を少なくする。第二次ストレージユニット上のスナップショット(3 4 1)は、午前 2 時に第一次システム上のすべてのデータ(ボリューム/ディレクトリ/ファイル)を表現する。

午前 1 0 時：これは、ベースバックアップ後のすべてのバックアップが自動的にインクリメンタル(incremental)であるため、インクリメンタル(incremental)バックアップである。ベースバックアップ以後に変更されたブロック(3 0 2)のみが転送されている点に留意されたい。第二次ストレージ装置上のスナップショット(3 4 2)は、午前 1 0 時における第一次システム上のすべてのデータ(ボリューム/ディレクトリ/ファイル)を表現する合成されたベースバックアップイメージである。

40

午前 1 1 時：午前 1 0 時のバックアップ以後に変更されたブロック(3 0 3)のみが転送されている。第二次ストレージ装置上のスナップショット(3 4 3)は、第一次システム上の午前 1 1 時のすべてのデータを表現する。

午前 1 2 時：午前 1 1 時のスナップショット(3 4 3)が、バックアップエクスプレス(backup express)復元スクリーン上に表示されるバックアップインスタンスから選択される。このバックアップインスタンスから、3 ファイル(3 5 1)が復元用に選択される。

50

#### 【 0 1 2 8 】

## (二重保護)

二重保護は最初の画像バックアップをテープまたはディスクにバックアップし、それらのライフサイクルを管理し、最初のバックアップの期限がきれるかまたはディスクストレージが利用不可能になったときにテープから直接復元することによって、インテリジェントディスクストレージに、最初の画像バックアップを保護する。

(ディスクに対するAPM(APM2D)：)

## 【0129】

(最初のバックアップ：)

1. ファイルシステムのイメージは、アプリケーション特有のメタデータ(APPS)と一緒にディスクにバックアップされる。このデータは、即時利用性および即時仮想化またはいずれか一方を可能にする形態で存在する。

10

2. 画像バックアップが支援されないファイルシステム/OSは、ファイルとしてディスクにバックアップされ、ソースファイルシステムのポイントインタイム(point-in-time)コピーとして移動先ディレクトリ下に存在する。

## 【0130】

(二重保護の説明：)

二重保護は、最低1つ(望む数)の最初のバックアップのバーチャルコピーをディスクまたはテープに作成する。ここで重要な点は、続くバックアップは、同一の変換されないコピーであることである。最初のバックアップは、フリーズされたポイントインタイム(point-in-time)の画像であるので、コピーを将来いつでも作成し、ファイルシステムの元の状態を依然として捕捉することができる。元のバックアップの多くのコピーを迅速にまたはポリシー決定時にいつでも作成することができるので、ツイニング(twinning)はそれ以上必要ない。支援されるアプリケーションに関しては、アプリケーションコンシステント(application consistent)なスナップショットが、テープバックアップが元の最初のバックアップ時に行われたかのように、テープに保存される。

20

## 【0131】

(プレゼンテーション/スケジューリング：)

GUIは、最初のバックアップジョブのリストの二重保護スクリーンに存在する。尚、これらのジョブは、二重保護用の候補である。これは、左のペインがバックアップジョブであるという事実以外は、従来の画像またはNDMPバックアップスクリーンに似ている。(装置選択は当初に、デフォルトクラスターとノードグループを含んだメディアプールを暗黙のうちを選択することにより回避される)二重保護ジョブは、最初のバックアップジョブネームと一緒にNDMPジョブとして、または、定義の一部としての最初のバックアップjobidとして保存される。スケジュールは単純であり、APM2Dとまったく同様のバックアップスケジュールであり、ベースインクリメンタル(incremental)またはディファレンシャル(differential)の設定はしない。最初のバックアップジョブの選択された特定のインスタンス(すなわちjobid)を有する二重保護ジョブは、関連スケジュールをもたず、ジョブは起動後に削除される。ジョブハンドラが、JOB\_STARTを受理し、これが二重保護ジョブであることを決定すると、CREATE\_DP\_JOBをジョブ名またはジョブIDを引数として指定するデータベースに発行する。データベースは、jobidを与えられると(snapidを参照することによって)ジョブ用のバックアップドキュメントを取得することができる。ジョブ名を与えられると、最新のバックアップジョブIDは、ジョブ用のバックアップドキュメントを検索するために使用される。バックアップドキュメントは、テープにNDMPジョブを元のAPM2Dジョブに一致するように構成することができることが要求される最初のバックアップの完全な状態を含む。元のジョブ1対1のマッピングは、ソース文の同等のセットを生じるDP\_JOBに作成される。

30

40

## 【0132】

例えば、タスクC、D、APPSを有するAPM2Dは3つのタスクに翻訳される。

## 【0133】

/vol/vol1/.snapshot/snapname/qt tree1、

50



/vol/vol1/.snapshot/snapname/qt tree2、  
/vol/vol1/.snapshot/snapname/APPS-qt tree

【 0 1 3 4 】

CREATE\_DP\_JOBは、ジョブハンドラによって一旦取得された定義がNDMPジョブの進行を許可する一時ジョブ名をもどす。一旦このジョブがテープに対するコピーを作成すると、そのコピーは、テープに対するバックアップがディスクバックアップの元の時間に実行されたものに見える。

【 0 1 3 5 】

最初のバックアップのjobidおよびtaskidは、最初のバックアップに対して二重保護タスクを関連するために必要である。CREATE\_DP\_JOBの一部として、データベースは、カタログエントリを作成する二重保護ジョブを予めカタログに登録することができ、実際のTASK\_CATALOGが送られてくる場合に、バリデーションされる。

【 0 1 3 6 】

必要条件がトリガされる(スナップショットその他が尽きる)ときに、CREATE\_DP\_JOBをSVHによって呼出すことができる。次に、SVHは、バックアップ後にまたはバックアップ前にさえJOB\_STARTなどを介してこのジョブを実行することができる。

【 0 1 3 7 】

ディスクおよびテープとライフサイクル管理を取込む包括的スケジューリングは、このプロジェクトの範囲外であるため、後の段階で考慮する。

【 0 1 3 8 】

(二重保護ジョブの実行：)

二重保護ジョブは、EIDソフトウェアまたは外部NDMPデータサーバ(所有権のあるNASバックアップ方法を含む)を介して仲介されるAPMバックアップである。最初のバックアップは、画像ファイルまたは複製されたディレクトリであり得る。EIDソフトウェアは、それらをバックアップするときに、二重保護バックアップがなされていると認知して、画像の場合または論理バックアップとしてはそれらが複製されたディレクトリである場合には、元のフォーマットを保存しながらそれらをバックアップする。外部エージェントは、それらのネイティブフォーマット(dumpまたはtar)で、画像または複製されたディレクトリをバックアップする。

【 0 1 3 9 】

二重保護バックアップがテープに送られる場合には、レガシージョブハンドラパスが使用される。第三次ディスク(第二次～第三次の複製)に向けられる二重保護バックアップは、SSSVHによってまたはある種の外部エージェント(単純なスクリプトを含み、その後カタログ登録ユーティリティステップが続く場合がある)によって処理される。

【 0 1 4 0 】

最初のバックアップ用の同一のファイル履歴によって冗長になるので、すべてのケースでファイル履歴は生成または捕捉されない。

【 0 1 4 1 】

すべての復元は、発信元のフォーマットに関係なく、ノードソフトウェアを介して実施される(これは、必要に応じて外部のdumpまたはtarフォーマットを理解することを意味する)。

【 0 1 4 2 】

(アーカイブフォーマット/コンプライアンス：)

長期アーカイブまたは規制の必要上、二重保護バックアップはtar、cpio、paxなどのある種のポータブルフォーマットの論理バックアップに画像バックアップを変換する場合がある。このようなバックアップをWORMテープまたはWORMドライブに送信して、コンプライアンス条件を満たすことができる。データは、最初のバックアップ時に保存されるファイル履歴を使用してこのアーカイブから復元可能である。直接アクセス復元(direct access restore (DAR))は、このように二重の保護プロセス時にファイル履歴生成を要求する関連fh\_infoと共にファル履歴を再保存することを要求する。

10

20

30

40

50

## 【 0 1 4 3 】

一般に、tarのような利用可能なユーティリティを、バックアップエクスプレス(backup express)から独立しているアーカイブフォーマットからファイルを復元するために使用することができる。現在の設計では、自由に異なるアーカイブフォーマットを作成および発行またはいずれか一方ができる。

## 【 0 1 4 4 】

(カタログ登録：)

各二重保護ジョブは、「sscat」の最初のバックアップと同じ多さのタスクをカタログ登録する。元のタスクおよびジョブID用のsscatの新しいフィールドは、元のジョブに対する参照を追跡するために追加される(この一部として、これが最初のバックアップジョブに対する高レベルで重要な構成であるので、sscatの一部としてファイルされるsnapidを追加することもできる)。二重保護ジョブは、第二次ディスクの位置を反映するパス名と共にsscatに独自の同等のディスクエントリを有する。

10

## 【 0 1 4 5 】

(sscatの例(一覧の一部)：)

JOBID	TASKID	ORIGINAL JOBID*	JOBNAME	DISK
1000055	1	0	First backup	C:
1000055	2	0	First backup	D:
1000055	3	0	First backup	APPS:
1000100	1	1000055	Double P	/vol/vol1/qtreesC
1000100	2	1000055	Double P	/vol/vol1/qtreesD
1000100	3	1000055	Double P	/vol/vol1/qtreesAPPS

20

## 【 0 1 4 6 】

(カタログ圧縮およびジョブ期限)

最初のバックアップとその後の二重保護バックアップは、別のジョブとして処理されるので、各々が独自の保持期間を有する。最初のバックアップの期限が切れると、二重保護バックアップがポリシーに依存して存在することを保証するためにチェックが行われる。保護されていない最初のバックアップが存在する決定がなされる場合、警告が出されるか、または、二重保護ジョブがこの時点でトリガされる場合がある。

30

## 【 0 1 4 7 】

第一次ジョブの圧縮時に、最初のバックアップの用のカタログエントリが保持されて、ファイル履歴を確保するために削除されない。アプリケーション復元に必要なため、バックアップドキュメントも保持される。これが復元ブラウズの一部に反映される必要があるものなので、元のジョブIDは進行されるジョブの一部として常に保持される。複数の二重保護ジョブが所定の最初のバックアップ用に存在する場合、それらのジョブは全て元のジョブIDを含み、そのジョブIDは元のssfileを示す。

## 【 0 1 4 8 】

カタログテーブルを介する単一パスが圧縮時に要求されるすべてであるので、このプロセスは比較的単純であるべきである。

40

## 【 0 1 4 9 】

(復元定義の生成：)

復元ブラウズは、復元表現用に元のジョブインスタンスから\$NDMPDATAを戻す。RJIプロセスも、適切な復元明細を作成するために元のssfileからファイル履歴を含むように強化される。プロセスは、元のssfileから復元パス名と一緒に二重保護バックアップに関するテーブルウィンドウを生成することを伴う。二重保護バックアップ用のssfileのルートディレクトリ(カタログ化される唯一のもの)は、無視される。

## 【 0 1 5 0 】

(復元：フォールトトレラント/位置独立)

50

通常のNDMPバックアップである二重保護テープバックアップは通常のNDMP復元により出現し、適合するファイルシステムに直接復元するために利用できる。元の第二次ディスクの位置が破壊されるか、破損される状況では、APM2Dの位置を再作成するか、または、段階的に復元するか仮想化を遂行するために、このようなバックアップを元の位置に復元することができる。アプリケーションが関与しない場合、これらの復元を正常のNDMP復元としてジョブハンドラーによって処理することができ、かつ、完全なソリューションの一部になり得る。

【0151】

第二次ディスクノードの災害回復またはフルノードバックアップは、個別のバックアップとして処理されて、災害の場合には第二次ディスクノードを復元するために独立して使用される場合もある。

10

【0152】

二重保護バックアップが最初のバックアップ用に存在する場合、それらのバックアップが表示されなという事実を除き、APM2D復元ビューは不変である。期限切れのバックアップについては、二重保護バックアップが存在する場合、ニアライン化されたバックアップとしてそのバックアップが出現して提示される。復元ブラウズプロセスは、APM2DバックアップとしてNDMPバックアップインスタンスを戻すために増補される必要がある。復元の選択は、現在のものとしてSSSVHに渡される。(ジョブハンドラが復元側のAPPH処理を実施する場合、アプリケーション復元用のNDMP復元ジョブを作成することは可能であるが、これはフォールトトレランスを処理する観点から十分に制限してもよい)。

20

【0153】

APPHがアプリケーション復元のためにコンタクトを取られ、復元ファイルリストが決定された後、プロトコルディレクタは、復元の選択を満たすために利用できるディスク移動先を介して回帰を試みる。これが失敗する(最初のバックアップの期限が切れているか、またはディスク移動先に到達できない)場合、テープからのNDMP復元ジョブはJOB\_STARTを介して構成され実行される(恐らくはジョブハンドラによって実行される)。一旦これが正常に完了すると、APPHは再びコンタクトが取られ、復元が完了される。

【0154】

(「遅延ミラーリング」)

第一次ボリュームを以下の手順に従って二次ボリューム上にミラー化してもよい。

30

【0155】

第一次ボリュームを組込む。

第二次ボリュームを組込む。

第一次ボリュームから第二次ボリュームにコピーされるブロックのリストを作成する。

リストが到着した後、第一次および第二次ボリュームに新規ブロックを書込む。

ブロックが書込まれる際に、先ほどのブロックのリストからブロックを除去する。

前述のリストをトラバースし、バンド幅が利用できて便利であるときは常に、第一次ボリュームから第二次ボリュームまでそのようなトラバースの結果として生じるブロックをコピーする。

40

前述のリスト上のすべてのブロックがコピーされるまで続ける。

【0156】

前述の最終結果は、第二次ボリュームが第一次ボリュームと同期するということである。この技術は第一次ボリューム上で処理を止める必要がなく、どれくらいの時間がコピープロセスを完了するためにかかり得るかに対して一切の制約を課さない。

【0157】

「即座に利用できる」仮想装置が、たとえば、第一次装置の故障後などに、利用された後に、「遅延ミラーリング」技法を、たとえば、物理的な一次装置を復元するために用いてもよい。仮想装置上のデータがそのスナップショットのポイントインタイム(point-in-time)現在、完全な状態であるという点で、仮想装置が一時的に使用される。しかし、仮

50

想装置は暫定ソリューションにすぎないこともあり、できる限り早く業務に代替第一次装置を復元する必要がある。「遅延ミラーリング」は、処理を連続的に続けて、実際のコピーが独自のペースで進行しながら、他のシステムコンポーネントでの負担を最低限にすることができる方法で、この能力を提供する。

【0158】

「遅延ミラーリング」技法は、クラッシュしたか、同期から外れたミラーを「再ミラー化する」ために有利に用いられる場合もあるが、第一次ミラーは生成時に残る。

【0159】

さらに、この技法は、停止することなくボリュームをコピーし、時間節約のために非常手段に関わることなくコピーすることが望ましい場合、「遅延ミラー」技法をどこでも使用することができる。

10

【0160】

(バックアップおよびファイルシステムの冗長性の解消)

複数のシステムがバックアップ動作でバックアップされている場合、マシンにはバックアップに関係する他のマシン上のブロックと同一である多数のブロックがあることは珍しくない。複数のマシンに同じオペレーティングシステムファイル、同じアプリケーションまたはデータファイルをそのマシン上にインストールさせたときに、これが起こる場合がある。同じ内容をもつブロックを何度も格納することは冗長である。冗長性は、ストレージの余計な利用だけではなく、複製ブロックを転送および格納するバンド幅の冗長な使用にも関わる。

20

【0161】

さらに、一つのファイルシステムにおいてさえも、ファイルの複製の結果として複製ブロックを有することも珍しくない。これも冗長性を意味する。

【0162】

そのような冗長性は、バックアップデータセットに書込まれるブロックごとのダイジェストを取得することによって、また、ダイジェストデータをリストまたはデータベースに入れることによって、バックアップコンテキストで解消してもよい。

【0163】

ブロックダイジェストの比較は、サーバー側で実行されるのが好ましい。

【0164】

バックアップすべきノードが変更されて、バックアップされる必要がある多数のブロックがある場合、そのノードはバックアップサーバにそのダイジェストを用いてそのようなブロックのリストを送る(それらのダイジェストがリストを作成するために個別のステップに関与しなくてもよいように、ノードがある他の目的(それ自身のブロックのどれかが変更したかを決定するなど)のために、前もってブロックダイジェストのリストを作成してしまうケースなどもある)。

30

【0165】

次に、サーバはブロックダイジェストを比較して、バックアップ用にそれらのブロックを要求し、既に有していないものを決定する(リストまたはブロックのデータベースは、迅速な参照を容易にするために、キーとしてダイジェストを使用するような方法で記憶される)。バックアップカタログの一部として、リモートノードによって送られるブロックの完全なリストが記憶される(サーバがそれがすでに有していると決定したたものに加えて送信されるものを含む)。

40

【0166】

好ましくは、バックアップされているノードに少数の変更ブロックのみがある場合、ノードは単にその状況でそれらのブロックを送って、冗長検査をスキップする。

【0167】

類似した技術は、単一のファイルシステムで冗長性を除くために使用される。ファイルシステムに書込まれる各ブロックは要約されて、すでに記憶されたブロックのダイジェストと比較される(ここでは再び、ブロックのリストまたはデータベースは、迅速な参照を

50

容易にするために、キーとしてダイジェストを使用するような方法で記憶される)。同一のコンテンツブロックがすでにファイルシステムに存在する場合、既存のディレクトリポイントが使用され、複製のブロックは書込まれない。ファイルが削除されるとき、そのブロックはそのファイルから割当てを解除される。他のファイルが同じブロックを使用する場合、それらの割当ては実質的に変わらない(ファイルがブロックを参照しなくなるまで、ブロックは「解放されない」)。

【0168】

(実施例：高速アプリケーション回復)

以下は、本発明で提供される高速アプリケーション回復を示す一連の例である。

【0169】

(実施例への導入：)

以下の例は、ファイラ(NASファイラなど)上でバックアップ画像から作成される仮想ストレージに帰属させることによって待機中または元サーバ上で迅速にアプリケーションをオンラインに導入するために、本発明で提供される能力を具体的に示す。

【0170】

一般的にNASファイラ上で、ソースノードから得られるコンシスタント(consistent)なボリューム画像は、バックアップとして関連するアプリケーションコンシスタント(consistent)な状態でニアライン化される。ユーザーがアプリケーション論理オブジェクトに対処する一方、バックアップエクスプレス(backup express)エージェントはアプリケーションから成る物理的実体の最新のベースバックアップを作成する。すべてのデータベースバックアップが最大限度であるという事実を犠牲にすることなく、永久インクリメンタル(incremental)画像は最後のバックアップの後で変更されたブロックだけがファイラにコピーされることを確実にする。アプリケーションデータおよび状態がニアライン化されるので、復元は、アプリケーションファイルのポイントインタイム(point-in-time)のコピーを回復することによって、非常に迅速に影響をされて、アプリケーションをオンラインに導入して、少数のやり直しログ記録を適用する。FARがバックアップの時点で存在するときは、FARはストレージを再作成し、アプリケーションが論理的に予想する物理的関係を確立してから、完全に機能的インスタンスにアプリケーションを回復する。

【0171】

(メカニズムの説明)

アプリケーション復元は、概ね以下の2つのプロセスである：回復されるデータファイルを復元してから、アプリケーションを復元する(前進復帰回復として知られる))))))。ユーザーは、災害、ユーザーによるエラーの種類または他の業務上の必要性により、バックアップのインスタンスまたはPIT画像(通常は最新のもの)を選択する。第一のステップは、ユーザが選択したPITボリューム画像からファイラ上でオンザフライ(on the fly)でアドレス指定可能な仮想ストレージ(LUN)を作成することによって達成される。これらのLUNは、次に問題のターゲットノードに対して視覚化される。次に、LUNは、ファイラへのiSCSIログインを介して復元ターゲット上のローカルディスクとして帰属される。実際のデータの移動が関与しないので、このプロセスはほとんど瞬間的である。一旦アプリケーションデータファイルがターゲットノードのローカル名スペースで見えた後は、アプリケーションを適切なアプリケーション特有のAPIを使用してプログラマ的に回復する。これには、ファイラバックアップ位置から必要に応じて取得される付加的なログファイルのアプリケーションを必要とする場合がある。これは、アプリケーションインスタンスをバックアップのポイントインタイム(point-in-time)までに送る。現在のログが利用できる場合、故障発生時点までのロールフォワードは可能である。バックアップがスナップショットバックアップであるので、アプリケーションは非常に短期間ホットバックアップモードであり、少数のトランザクションのみがデータベースをコンシスタント(consistent)な状態にするために適用されなければならない。相対的に単純なこれらのステップの迅速さにより、アプリケーションは、FARプロセスが始動された後、わずか数分で立上ることができ

従来との復元と比較して、FARは日単位または時間単位から分単位までより高速にアプ

10

20

30

40

50

リケーションダウンタイムを桁違いに短縮しつつある。FARは、データセットのサイズとは関係なく、高速になる。

【0172】

(復元後)

FARで全てが終わるのではない。FARがブロック変更を完了した後、トラッキングを使用することができるし、必要であれば、ローカルスライスアタッチメント(local slice attachment)を行ってもよい。これはバックグラウンドでの復元を進行可能にする一方、アプリケーションの起動および実行も可能にする。変更ブロックのトラッキングが可能であるので、インクリメンタル(incremental)バックアップは復元のポイントインタイム(point-in-time)から開始してもよい。このアプリケーションは、最終的にすべての最近の変更(復元以後)を維持して、最小限のダウンタイムで元のまたは別のノードに戻り、フェールバックする場合がある。

10

【0173】

(要件：)

ソースノードおよびターゲットノードは、実行中であり、APMに関して認可されている必要がある(アプリケーションがもしあれば、個別に認可される必要がある場合がある)。

NAS装置または第二次ストレージユニットは、iSCSIに関して認可される必要がある。

ターゲットノードには、iSCSIイニシエータソフトウェアをインストールする必要がある(iSCSI HBAも支援される)。

待機中ノードは最小限のアプリケーションのインストールと一緒に予め構成される必要がある。

20

プラットフォーム/アプリケーションのサポートには、Windows XP/Windows2000/Windows 2003、SQL Serve 2000(SP2以上)、Exchange 2000(SP2以上)/Exchange 2003、SQL Server 2005、Oracle、Linuxを含む。

【0174】

急速なアプリケーション復元と復元後のデータのライフサイクルに関する様々なシナリオおよびアプリケーションは、以下のセクションで探求する。

(実施例1。 APPsに対するIV(即時検証、instant varification)

【0175】

必要性：バックアップは実際には検証されないもので、復元は常に当て推量である。テープは信頼できない。一般に、検証は、バックアップ画像の内部一貫性を確認することを意味する。アプリケーションの一貫性と回復性は、偶然の問題である

30

【0176】

。

アプローチ：APPs用のIVは、可能な場合には、代替検証ノードまたは元のノードに復元する(FAR)によって、ほぼ瞬間的にアプリケーションバックアップを検証する。次に、アプリケーションが復元されてプロセスが完了する。あらゆるバックアップがインテグリティに関して常にチェックされ、他に災害時の訓練を災害シナリオを再度作成再現するために実施しなくてもよいように、これを予定することができる。

【0177】

40

使用されるPIT画像：通常は最新のものであるが、検証を一括する場合には、過去の画像であってもよい。

【0178】

実施場所：一般に、最低限のアプリケーションのインストールを予め済ませた代替ノード上で実施される。バックアップ用ソースと同じノードは、アプリケーションがサポートする場合には、使用することができる(例、検証されているデータベースに再度名称をつけるオプションを有する回復ストレージグループ(Recovery Storage Group)またはSQL Serverを用いて構成されるExchange 2003)。一般に、元のノードの検証は、余分なストレスをアプリケーションサーバにかけるので、推奨されない。

【0179】

50

(モード：)

ライトウェイトな(lightweight)検証：アプリケーション(一般にはデータベース)が正確に起動/回復し、そのためにバックアップの正確さを検証する。

【0180】

包括性：必要ならば、さらに検証を実施し(よりリソースに集約的になる)、アプリケーション特有の技法を用いて、すべてのデータベースページがクリーンであること、および論理オブジェクトが適切に機能すること、またはそれらの一方を検証する(テーブルをスパン(span)するデータベースクエリを想像すること。結果はデータベースの健全性を明確に正当化するものである)。

【0181】

アプリケーションに特有の注記：

Exchange：ストアの組込みは、一般に重要なステップである。さらなる検証は、代替ノードの「eseutil」を用いて実施することができる。

SQL Server：データベースの組込みは、一般に重要な検証ステップである。さらなる検証は、「DBCC」を介してまたはSQLクエリを実行することによって、実施される。

【0182】

フォローアップ：

無し。検証は、一過性の動作であり、iSCSIログオフまたはリブートが、マシンの状態をクリアにする。APPs用のIVは、次の検証の実行により以前の検証マッピングをクリーンアップするように構成される。マップされたドライブと一緒にマシンの状態を確保する必要はない。このため、この代替ノードをさらにバックアップする必要はない。

【0183】

(例2. 事業継続性に関するAPPsのIA(即時利用性)

必要性：ダウンタイムが、分レベルまで最小限化される。大部分の最近のアプリケーションのバックアップ状態は、復元される(バックアップの頻度次第では、ほとんどデータは失われないこともある)。

【0184】

アプローチ：FARは、待機時のアプリケーションインスタンスまたは元のノードを瞬間的に戻し、ダウンタイムを最小限にする。バックアップのときのアプリケーションの状態が復元される。最後のバックアップが行われた後の変更は、アプリケーションログが利用可能でない限りは、失われる(元のノードから、またはある種の複製された位置のいずれかから、復旧される)。現在のアプリケーションログが利用可能であり、後に適用できる場合には、そのアプリケーションをデータを損失させずに故障時にロールフォワードすることができる。

【0185】

使用されるPIT画像：一般には、最新のものであるが、災害上の理由によっては(例えば、ウイルス攻撃)、事件の前の画像。

アプリケーション特有の注記：

【0186】

Exchange 2003：「ダイヤルトーン復元」のように、空のデータの作成を含み、RSG(回復ストレージグループ、Recovery Storage Group)に対する回復が実施されるときにデータベースを切換えると、再マージは必要なくなるが、その理由はFARが高速であり、アプリケーションの休止を手間をかけずに最低限にできるからであるという複雑なシナリオ。

【0187】

SQL Server：待機データベース、複製、およびログ SHIPPINGあるいはそれらのいずれかは、SQL Serverの利用可能性に関する高価で管理者集約的なオプションである。FARは、容易に配置ができて、管理コストが高速なバックアップと需要に応じた迅速な利用可能性を組合わせて低くなる。

【0188】

10

20

30

40

50

(実施例 2 a オンラインでの復元)

必要性：アプリケーションデータは、ローカルのまたはSAN帰属のストレージに最終的に復元して戻されなければならない。第二次ストレージからのストレージを用いることのみが、一時的なオプションとなり得る。

【0189】

実施場所：一般には、元のアプリケーションまたは最も近いノードであるが、災害またはセットアップの妨害の性質に依存する。

【0190】

フォローアップ(LAR)：アプリケーションはオンラインであり、ユーザは数分以内でアプリケーションの使用を始めることができる。復元はアプリケーションが立ち上がり実行する間に、ローカルディスクスライス(local disk slice)に対してバックグラウンドで継続する。すべてのデータが、ローカルスライス(local slice)に復元された後、アプリケーションは、間もなく中止または停止され、iSCSIマッピングが除去される。次に、ローカルスライス(local slice)は、格上げされて単独のアプリケーションディスクとなる。このアプリケーションは、レジュームまたは再起動される。アプリケーションは(仮にでもあれば)、エンド(end)で短期間しか利用できない。

【0191】

BAR：新規の復元ボリュームキックイン時にアプリケーションを保護するための正規のバックアップスケジュール(このサイクルは、アプリケーションが将来復元される必要がある場合に繰り返す)。

【0192】

(実施例 2 b . オンライン復元のない場合)

必要性：バックグラウンド復元が必要ない理由は、待機ノードが一時的であり、低い性能で十分だからであるか(フェールバックは、元のサイトが再構成された後、起こる場合がある)、または、バックアップ画像を記憶するファイラがアプリケーションをホストするために十分強力だからである。

【0193】

冗長な移動先：ハイエンドファイラ(可能であればリモートサイト)は、元のバックアップ移動先に記憶されたバックアップ画像を(第三次ストレージなどに)ミラー化する。この構成は、ハイエンドファイラに再指定される復元に向いているが、元のファイラには向いていない。ローカルスライス(local slice)に対するバックグラウンドでの復元は、この場合には必要ではない。その理由は、ファイラストレージが、ハイエンドで永久的だからである。

【0194】

復元されたストレージの品質：

A . 低品質 第二次ストレージへのiSCSIの組込み：アプリケーションは、特に、一時的な状況であり、高品質のストレージおよびノードが修復されるかまたは独立して復元された後で、フェールバックが短時間で期待される場合には、iSCSI上でストレージを適度実施しながら耐久することができる場合がある。

B . 高品質 第二次～第三次への複製またはバックアップ後の元のファイルからのコピーによって確立される高性能ストレージへのiSCSIの組込み：アプリケーションは十分に機能し、これが永久的なソリューションになり得る。ただし、これは、フェールバックを含まない。

【0195】

フォローアップ：必要な場合には、復元後バックアップ(BAR)をターゲットマシンから継続することができる。または、ストレージがNAS装置に効果的に移動してからNASブロックレベルのバックアップを開始することができる。ファイラ上に永久的なストレージが独自のライフサイクルで確立されているので、ファイラ上のLUNは、元のスナップショットからの依存性を破壊するようにクローン化してもよい。

【0196】

10

20

30

40

50



APPsに対するERF(最終高速フェールバック、Eventual Rapid Failback)：アプリケーションは、以下の方法で、元のノードまたは別の再作成されたノードに最終的にフェールバックしてもよい。

1. 現在実行中ノード上でアプリケーションを短期間シャットダウンする。
2. 第二次ストレージと代替ストレージの間にある関係が成立しており、元の第二次ストレージが最終移動先に近い位置にある場合には、複製ソースおよび移動先を逆転させ、現在のストレージから第二次ストレージを再同期化し、アップデートする。あるいは、ステップ3に移動する(このプロセスは、最新の共通スナップショットを除去し、それ以後のコピーを変更する。これは、フェールバックが故障発生時の直後に適切に始動されたと仮定して、迅速に完了すべきである)。
3. 所望のノードに対してFARを実施する。
4. アプリケーションインスタンスは、待機ノード上に位置し正規の動作をレジュームできる状態(最新の変更を有して)に戻る。

10

【0197】

(例3. アプリケーション全体のバックアップからのファイングレイン(fine grain)復元)

必要性：大部分のアプリケーションについては、ファイングレイン(fine grain)復元は、アプリケーション全体のバックアップからは、不可能である。粒度の小さいアプリケーションオブジェクトのバックアップは、信頼できず、極端にリソース集約的である。ファイングレイン(fine grain)アプリケーションオブジェクトの現代のバックアップ/復元ソリューションの最新の技術を考慮すると、代替インスタンスに対するアプリケーション用のFARを実行し(これは非常に迅速に完了する)、その後アプリケーション特有のツールを使用してファイングレイン(fine-grain)オブジェクトを回復することは、極端に魅力的な選択肢である。

20

【0198】

アプローチ：FAR後にアプリケーション特有のツールでアプリケーションオブジェクトをドリルダウンし試験する。次に、これらを元の移動先にマージするか、外部で使用するために抽出することができる。

【0199】

使用されるPIT画像：ファイングレイン(fine grain)オブジェクトが削除されたか、破損していないかた時点に依存する。

30

【0200】

実施場所：一般に、異なるノード上の代替インスタンスまたは元のノードに対して実施する(セットアップおよび必要性に依存する)。

【0201】

フォローアップ：一般に、行わない。一時的に必要であり、インスタンスはトーンダウン(tone down)され、iSSIマッピングは取り消されるためである。

(アプリケーション特有の注記：)

【0202】

Exchange 2000：バックアップの不利益を被らない単一メールボックスの復元は、FARを使用してから、EXMERGE.EXEまたはその他のツールを使用して可能となる。

40

【0203】

Exchange 2003：復元ストレージグループとFARの強力な組み合わせで、過去のある時点からファイングレイン(fine grain)復元のために、単一のメールボックスまたは下位メールボックスでさえ復元する。極端に速く、負担のないオプションである。

【0204】

SQL Server：テーブルレベルの復元であり、「bcp」またはそれ以外のツールを使用して代替FARされたインスタンスからテーブルを復元することができる。

【0205】

(例4. 分析、報告、およびデータウェアハウジング用のAppsのインスタントレプリカ)

必要性：分析または報告用に第二のコピーを一般に取得しているのは、広範なディスク

50

スペースで高価なスプリットミラー技術を実現してきた贅沢な余裕のある大企業である。FARを使用すれば、これははるかに低コストで実現可能になるだけではなく、ほぼ瞬間的に複数の移動先に実施される。企業は、自信をもってより分析的な可能性を探求し、競争力がつく。

【0206】

アプローチ：FARを望むだけ多く1つ以上のノードに対して用いる。

使用されるPIT画像：通常は最新のものであるが、過去のあるポイントインタイム(point-in-time)での分析的または業務上の理由に依存する(最近のクリスマスセールに関するデータということもありえる)。

【0207】

実施場所：代替ノードに対して実施。元のノードは依然として業務用アプリケーションラインを実行中である。

【0208】

次に何が起こるか(LAR)?：レプリカが独自のタイムラインまたは寿命を有する必要がある場合、バックアップしなければならない。バックアップは復元されたコピーからのインクリメンタル(incremental)な変更と共に継続する。

【0209】

(例5．長期保持のためのテープバックアップに関する代替ノード復元)

必要性：付加的な保護および長期保持またはいずれか一方は、テープのバックアップを必要とする。ニアライン画像はすぐに期限が切れるため、テープバックアップは必ずといってよいほど長期保持のためには必要である。

【0210】

アプローチ：iSCSIマップされたボリュームの画像バックアップをテープへ。その後、テープ画像は、粒度を問わず将来のポイントインタイム(point-in-time)にノードに復元できる。

【0211】

使用されるPIT画像：一般に、バックアップスケジュールからスタガされ(staggerd)、ニアラインに残る必要があるインスタンスの数によって決定される。

【0212】

実施場所：待機ノードに接続されたある種のテープ。これは、APPsノード用のIVであってもよい。

【0213】

フォローアップ：テープに対する画像バックアップは、FARボリュームから実施される(認可が必要である)。バックアップ成功後、iSCSIのマッピングは、削除され、ステージは次のサイクルにセットされる。

【0214】

(実施例6．ストレージ移動用のFAR)

必要性：直接取り付けられたまたは従来のSANストレージをブロック指向のNASファイラに、コスト、統合、性能、または管理上の理由から移動する必要がある場合がある。

【0215】

アプローチ：一旦ブロックレベルのバックアップがファイラに実施されれば、移動は既開始されている。バックアップ画像は、ハイエンドファイラにコピーまたはスナッピーミラーされて(snap-mirrored)、プロセスがさらに容易になる。FARは、効果的に移動プロセスを完了する。

【0216】

使用されるPIT画像：一般には最近のもの

【0217】

実施場所：ファイラ上に作成されたLUNに帰属する新規のアプリケーションノード。

【0218】

フォローアップ：さらに、LUNは、(バックグラウンドで)クローン化されるが、アプリ

10

20

30

40

50

ケーションは立上り、LUNを含むスナップショットに拘束されずに自由に稼動する。その後、スナップショットは、リサイクルされて、スペースを再利用する。復元後のバックアップ(BAR)は、LUNによってバック(back)されるボリュームまたはLUNを含むファイラボリュームまたはクォータツリー(quota-tree)をレジュームする。

【0219】

(例7. FAR4C -- コンプライアンス用のFAR)

必要性：法的理由。一般に、コンプライアンスは、所有権のあるハードウェアに関連する高価なソリューションを伴う。第二次WORMストレージに対するバックアップエクスプレス(backup express)画像のバックアップは、過去のある時点のマシン状態を即時に正確に再作成することができる手ごろなソリューションを提供する。

10

【0220】

アプローチ：FARから待機ノードに対して、アプリケーション状態またはマシン全体の状態を再度作成。

【0221】

使用されるPIT画像：これは、アニュアルレポートまたは需要による(セキュリティ上の理由によっては過去のポイントインタイム(point-in-time)のいずれであってもよい)。

【0222】

実施場所：待機ノードのいずれか。

【0223】

フォローアップ：一般には、一過性であり、レギュレータが満足した後は、トーンダウン(torn down)する。マシン全体の状態は、オフライン試験またはポータブルコンプライアンスのために必要であればWORMテープにシナリオ5を介してアーカイブされる。

20

【0224】

(追加事例)

図4(AおよびB)は、即時利用性を利用して、復元プロセス時に仮想的に業務への割込みを削除する即時利用性および回復のシナリオである。

午前11時00分：第一次ノード300上でのディスク故障前のNAS107上の最新ルーチンバックアップを示す。

午後12時00分：ボリュームD406が故障する。午後12時05分、数分以内に、第二次ストレージユニット上の論理ユニット番号(LUN)411を介して午前11時00分のバックアップインスタンスが、iSCSI(412)を介してマッピングされ、文字Dを駆動する。業務は継続する。第二次ストレージユニット107へのiSCSIの接続は、ユーザに見えるようになっている。データの変更は、第二次ストレージユニット上の「ライブデータエリア(live data area)」414に記憶される(四角で囲んだ白いブロック)。午前11時のバックアップインスタンス413自体は、読み専用であり、変わらない。

30

午後12時05分から午後1時00分。故障したディスク406を新規のディスク421に交換する。正常な業務の使用は、第二次ストレージユニット107への生きたiSCSI接続を介して続行する。

午後1時00分～1時45分。午前11時のバックアップインスタンスは、第一次装置300とその新規のディスク421に転送される(451)。生きたiSCSI接続を介した業務は、システムが午前2時45分に停止されるまで途切れることなく続く。

40

午前2時45分～午前3時。管理者は、データの再同期化(遅延ミラー)(452)を実施する。この期間、ユーザはシステムを利用できない。即時利用性によって、管理者は一晚のメンテナンス期間に柔軟に再同期(452)を実施する。

午前3時00分。復元が完了し、即時利用可能性接続は、ボリュームDを新規のディスク421に再マップすることによって、終了する。

【0225】

本文に記載された実施形態は、本発明に関して記載された目的を達成することは明らかである。目下の好ましい実施形態について詳細に説明してきたが、当業者であれば、本発明の原理がそれ以外の装置、システム、および方法によって、以下の請求の範囲に定義さ

50

れるように、本発明の適用範囲および精神から逸脱しないことは明らかであろう。

【図面の簡単な説明】

【0226】

【図1】本発明の一実施例の典型的な企業内配置を示すハイレベルシステムブロック図である。

【図2】ブロックレベルのバックアップデータ転送およびファイルレベルの復元を示すブロック図である。

【図3】ブロックレベルのインクリメンタル(incremental)バックアップの一部である動作後に模範的なファイルレベルの復元が続く時系列を示すブロック図である。

【図4A】ブロックレベルのインクリメンタル(incremental)バックアップ、即時利用性復元、および「遅延ミラー」複製に関わる災害回復シナリオ例を時系列で示すブロック図である。

【図4B】ブロックレベルのインクリメンタル(incremental)バックアップ、即時利用性復元、および「遅延ミラー」複製に関わる災害回復シナリオ例を時系列で示すブロック図である。

10

【図1】

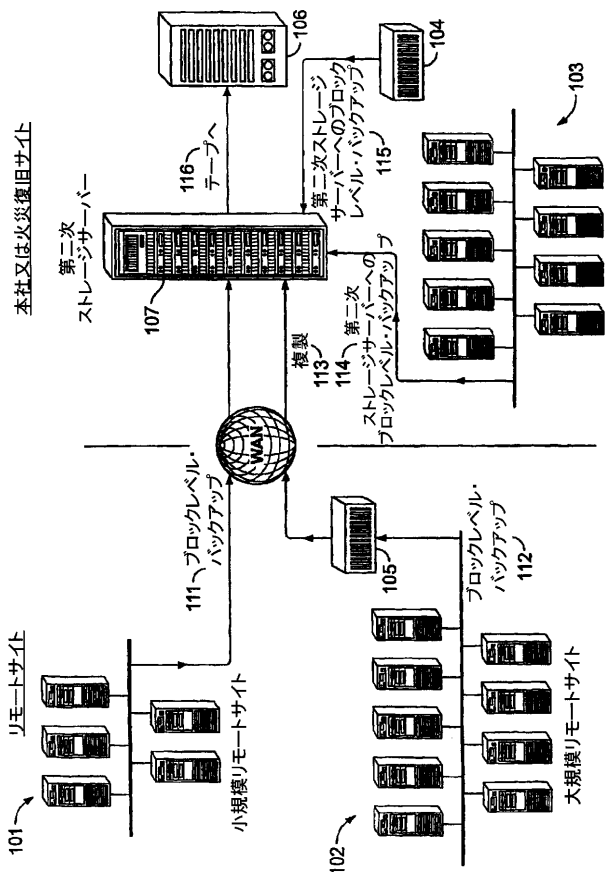


FIG. 1

【図2】

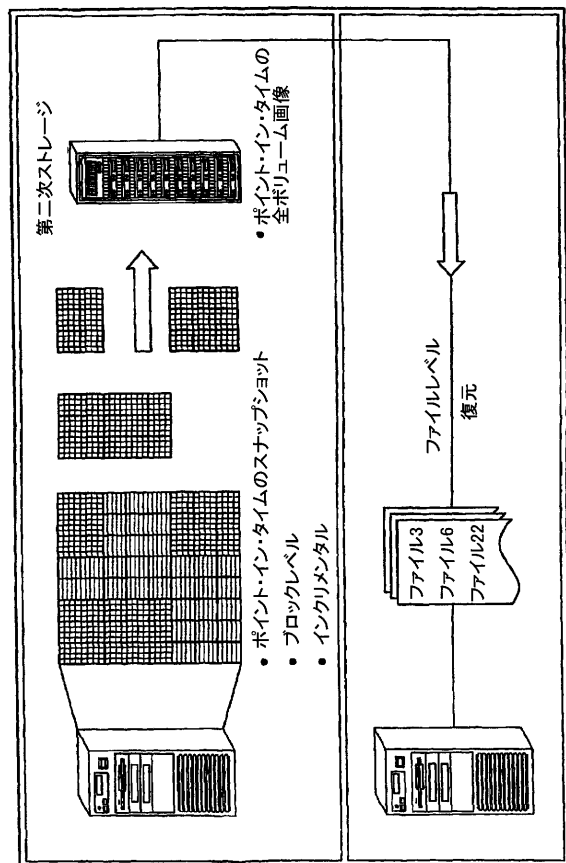


FIG. 2

【 図 3 】

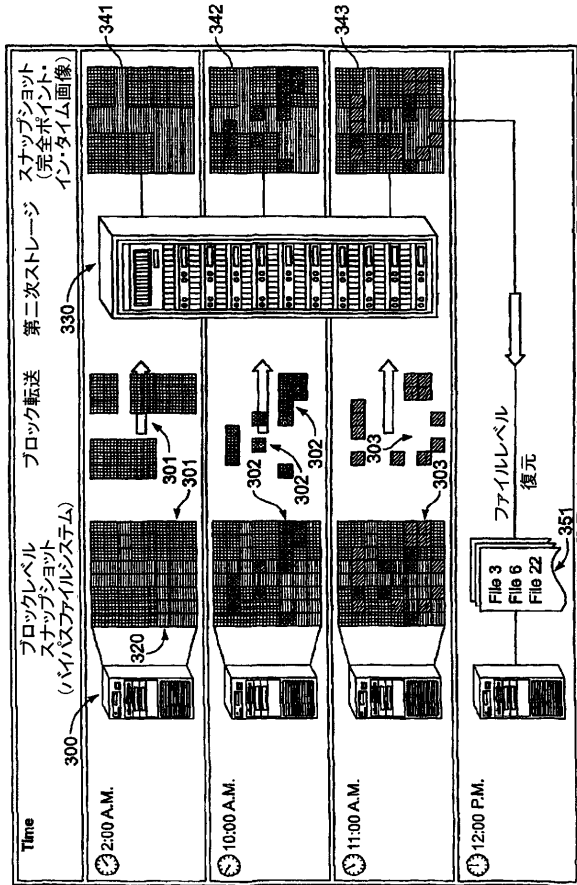


FIG. 3

【 図 4 A 】

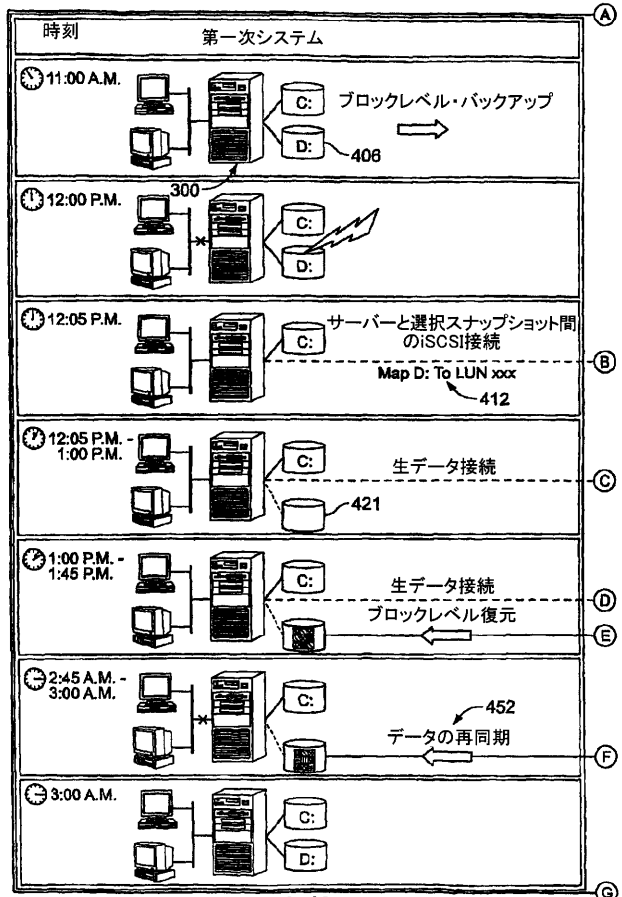


FIG. 4A

【 図 4 B 】

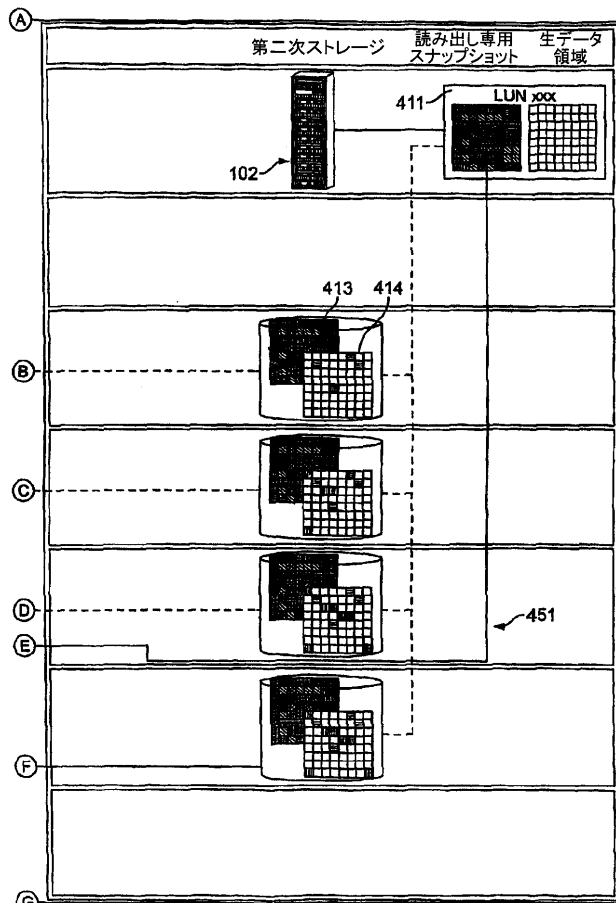


FIG 4B

## フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(特許庁注：以下のものは登録商標)

- 1 . W I N D O W S
- 2 . U N I X
- 3 . L i n u x

(72)発明者 リウ ピーター チ シウン  
アメリカ合衆国 ニュージャージー州 07652 パラマス ブラウン サークル 116

(72)発明者 アチャーリャ ソウビル  
アメリカ合衆国 ニューヨーク州 10570 プレザントヴィル スカイ トップ ドライヴ  
73

Fターム(参考) 5B065 BA01 CA11 EA35  
5B082 DE06 GA04

【要約の続き】

【選択図】図4A