



クラウド時代のバックアップ手法

～事業継続とIT統制のためのデータ保護～

ここに記載された内容は更新される可能性があります。この文書に記載されている内容はこの文書の発行時点におけるエムアイシー・アソシエーツ株式会社の見解を述べたものです。エムアイシー・アソシエーツ株式会社が、この文書に記載された内容の実現に関して確約するものではありません。また発行日以降については、この文書に記載された内容の正確さは保証しません。

この文書は情報の提供のみを目的としており、明示的または黙示的に関わらず、この文書の内容についてエムアイシー・アソシエーツ株式会社はいかなる保証をするものでもありません。

エムアイシー・アソシエーツ株式会社は、本書に記載してあるすべて、または、一部の記載内容に関し、許可なく転載、または、引用することを禁じます。

バージョン	作成日付	変更点	総ページ数
1.0	2011/01/31	-	8
2.0	2011/02/04	5頁目に参照URLを追加	8

本書作成、編集、管理
エムアイシー・アソシエーツ株式会社
〒103-0004 東京都中央区東日本橋3-12-12
櫻正宗東日本橋ビル9F
Tel. 03-5614-3757 Fax. 03-5614

1. はじめに	4
2. バックアップの課題 - Disaster RecoveryとSOX法	4
3. Snapshotバックアップ	5
4. そして、CDP	5
5. CDPの進化ー圧縮と重複排除	6
5.1 データの圧縮	6
5.2 重複排除(Deduplication)	6
1) ファイルレベルの重複排除	6
2) ブロックレベルの重複排除	6
6. CDPシステム	6
6.1 ネットワークでのデータ保護	7
6.2 CDPバックアップデータのリプリケーション	7
6.3 バックアップデータからのデータクリエーション	8
7. まとめ	8

1. はじめに

ストレージのバックアップリカバリは古くからのテーマであり、ご存知の通り、今日まで新しい技術によって進化してきました。殊に、紙ベースの業務からITベースのワークフローを構築してゆく中で、データ保護という課題はアプリケーションが複雑になり、データそのものが大きく、リッチな物となったことで、ビジネスの継続ということと密接に関連するようになりました。そして、9.11の事件を機に、デザスターリカバリ(DR)の方法が真剣に検討されるようになってきました。

その後、Snapshotという新しいバックアップの手法が開発されたことで、データのバックアップの方法は大きく変化し、数年前からはDRによる事業継続に加え、過去のいつの時点のデータファイルでも復元でき、万一の場合には瞬時にシステムを復旧できる技術が紹介されるようになりました。その技術は Continuous Data Protectionの略語でCDPと呼ばれ、日本語としては『継続的データ保護』と訳されています。

今回、本書におきましてCDPの技術的な背景や、弊社が検証してきましたCDP技術に関しご説明させていただきます。

2. バックアップの課題 - Disaster RecoveryとSOX法

従来、ストレージバックアップはオペレーションが停止している限られた時間にテープ媒体やネットワークドライブに、ソースデータの全体や更新差分のデータをバックアップとしてコピーし、万一の場合は、バックアップデータからリストアしてデータ復旧を行なっていました。

ユーザは万一の場合に前回のバックアップに戻れるという安心感があり、バックアップ時間や、リストア時間というオペレーション上の大きな負担や、テープの保管スペース、更には、あるファイルのある時点の内容を探し出す手間等は安全安心との引き換えに、やむを得ないコストとして考えられてきました。2000年に入り、テロや災害が発生してもシステムを速やかに復旧させ、サービスを長期間停止をすることなく、事業継続を可能とするために Disaster Recovery (DR)を前提にしたシステムを構築することが求められるようになりました。プライマリーのシステムに障害が発生した場合でも、バックアップシステムで事業が継続することができるDRに対応したシステムでは、データは常にプライマリーのシステムとバックアップシステム間で複製されます。しかし、このようなシステムにおいては従来のテープベースのバックアップや、1日に1回のデータリпликаーションではバックアップウィンドウや、リカバリーの時間が長時間かかり、システム的に大きな負荷がかかることとなります。

また、会計監査の信頼性と内部統制の強化を目指したSOX法が2002年に制定(JSOX法は2006年制定)されると、この法に準拠したIT統制を実現するITベースのワークフローを構築し、より信頼性の高いシステムとするために、過去データの保全と容易なアクセス、そして以前より柔軟なバックアップシステムが求められるようになってきました。

3. Snapshotバックアップ

2000年に入り、バックアップ技術の中にSnapshotを利用してバックアップを取るという技術が紹介されるようになりました。2003年にはこのSnapshot技術をシステムサービスとして実装したMicrosoft Windows 2003 がリリースされました。このシステムサービスをVSS(Volume Shadow Copy Service)といい、Snapshot技術の一つであるCopy-on-Writeという方式です。データの書き込み時にその直前のデータブロックのコピーを取り、別の場所にそのブロック情報を保存します。このVSS Snapshotによりデータの更新が履歴情報(Point in Time)として保存することができます。(詳しくは、<http://www.micassoc.co.jp/techdocs/SnapshotandVSS.html> を参照してください。) 多くのアプリケーションはこのVSSに対応したインタフェースを用意し、データの更新時にディスクにデータを書き出し、Snapshotでその更新内容を別の場所に保存することができるようになっています。

4. そして、CDP

ここまでお話ししたDRや、SOX法準拠のIT統制への対応の他、非構造化データと呼ばれるコンテンツ系のデータが急激に増大するようになり、アプリケーションデータのみならず、メールサーバやファイルサーバのバックアップが大きな課題になってきました。それに呼応して、CDPというバックアップ手法が2005年ごろから紹介されるようになりました。ストレージ業界団体であるSNIAではCDPについて以下のように定義しています。

(参照:

http://www.snia.org/forums/dmf/knowledge/white_papers_and_reports/CDP_Solving_recovery_20080808.pdf)

『Continuous Data Protection(CDP)はデータが変更される都度、データを取り込み、その変更をトラック可能とし、プライマリーデータの個々の変更を保存し、過去のどの時点からでもデータをリカバリーすることができる手法のことである。CDPにはブロック、ファイル、アプリケーションベースのシステムがあり、保護対象のシステムやデータを過去の細分された時点へ遡って復元し、アクセス可能にします。』

以上の定義から、以下の3つの基本的な仕様を満たしている必要があります。

1. データの変更は継続的保存され、その変更を辿れること。
2. 総ての変更データはプライマリストレージとは別のストレージに保存されること。
3. Recovery point objectives(データ変更の継続的な保存目標) は任意に変更可能で、固定されていないこと。

CDPは従来のバックアップソフトのように、オペレーションが停止している時間(バックアップウンド) にデータをバックアップするという手法と異なり、Snapshotベースでシステムのオペレーション中でもデータ保護を行うバックアップのシステムです。ストレージにデータ更新が発生する都度、Snapshotデータを継続的に取り続け、元のデータを別の場所に保存します。この更新履歴を持つSnapshotデータにより、いつでもある時点のファイルオブジェクトまで遡ることができるようになります。

5. CDPの進化—圧縮と重複排除

5.1 データの圧縮

CDPの技術は業界で紹介されて以来改良が加えられました。まず、圧縮技術を使用して、バックアップデータをソースデータに対して数十パーセントにしてしまうことで、バックアップストレージの容量削減が可能になります。6TBのソースデータを2/3に圧縮し、毎日5%のデータが更新されると仮定すると、1ヶ月のバックアップデータは10TBになります。本来15TB程度のストレージ容量が必要なバックアップデータ容量から大幅に削減することができます。

5.2 重複排除(Deduplication)

1) ファイルレベルの重複排除

圧縮技術はバックアップ用ストレージ容量の削減の効果がありますが、更に高頻度な更新が行なわれるシステムでのスナップショットデータの継続的な保存では、大容量のストレージが必要になります。また、メールに添付されたPDFファイルを多くの人に配信した場合、メールサーバに保存されたデータはオリジナルの何倍にもなってしまいます。こうした無駄を排除するために、ファイルレベルでの重複排除の技術が取り入れられています。別名、Single Instance Storeと呼ばれる重複排除技術ですが、ファイル全体をチェックして重複が確認できれば、バックアップを取らないという方法です。この方法により、リードオンリーのデータでは容量削減に大きな効果があります。

2) ブロックレベルの重複排除

ファイルレベルの重複排除の技術には前述でご説明した長所がありますが、欠点もあります。データベース等の場合は1レコードが更新されただけでもチェックの結果、データベースファイル全体をバックアップすることが必要になり、バックアップストレージ内部で大きなスペースを浪費することになります。この問題を回避する方法としては、バックアップされるボリュームをブロックレベルでチェックをし、重複するブロックは以前にバックアップされたものとしてコピーをしないことです。一方、変更されているブロックはコピーをします。以上の方法でファイルレベルの重複排除より、遥かに大幅なストレージスペースの削減を実現します。このようにデータの圧縮、ブロックレベルの重複排除の技術を合わせることにより、バックアップデータは通常の数十から数百分の1の圧縮率で保存されます。

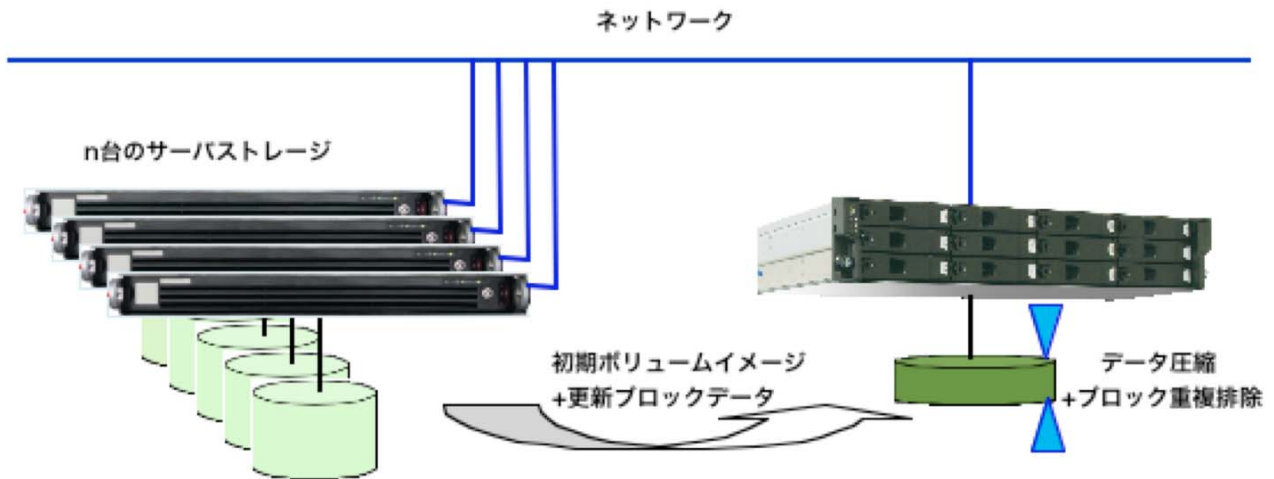
6. CDPシステム

クラウド、仮想化サーバ、仮想化ストレージシステムと新しい技術が続々と現実のITインフラに取込まれつつあります。また一方で、新たに作成されるデータは年率150%を超える勢いで増大しています。このような環境においてデータ変更のトラックが可能な切れ目の無いデータの保護、どの時点のデータでも瞬時に復元できるリカバリーといった事業継続可能なCDPシステムに対するニーズが増えつつあります。

今回、本書で紹介してきましたように、CDPの連続性と即時性、省ストレージスペースに加え、CDPではバックアップソース、ターゲットストレージ、バックアップサーバ等のデータ保護に関するリソースが企業内ドメインのみならず、世界のどの場所でも配置できるようになり、データの複製を幾つでも作成することができるようになってきています。

6.1 ネットワークでのデータ保護

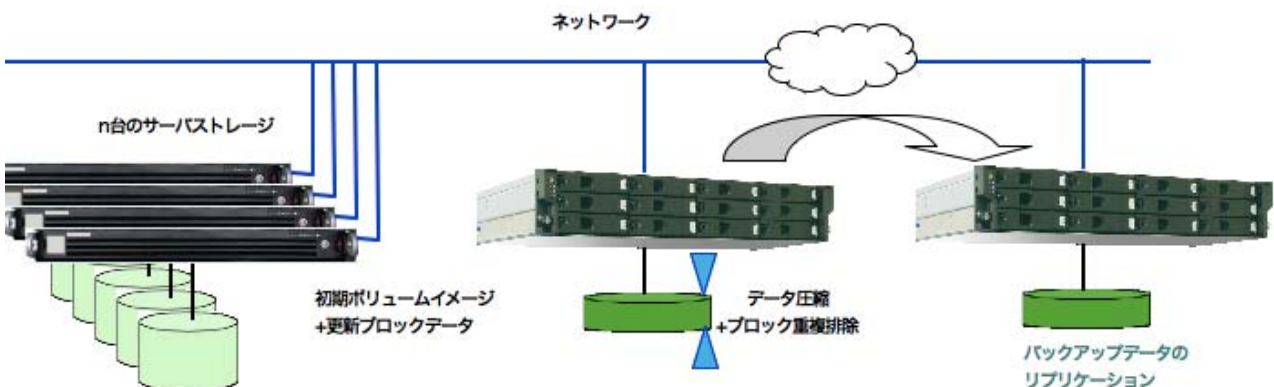
CDPシステムの一例として、まず、プロテクトされるサーバボリュームとバックアップターゲットとしてのシステム環境を下図に示します。



複数のサーバのバックアップを1台のバックアップサーバで取ることが可能になります。バックアップデータは初めにボリュームイメージデータを取り、それ以降、任意に設定したリカバリーポイント取得頻度でデータのスナップショットのデータをサーバに転送し、データを圧縮、ブロックレベルの重複の排除を実行します。結果として、数十、数百分の1のデータサイズとなり、バックアップコストの大幅な削減が可能になります。

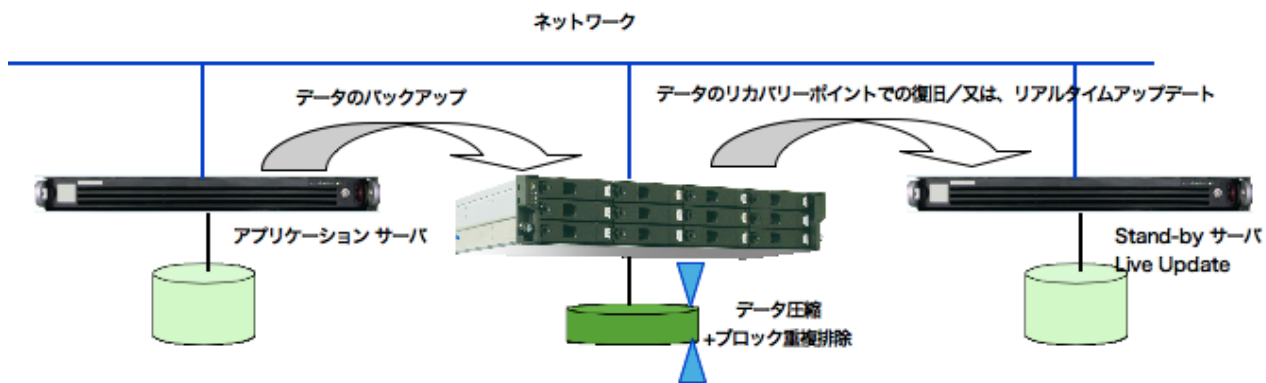
6.2 CDPバックアップデータのリプリケーション

多くのCDPアプリケーションはバックアップデータをネットワーク上の他の物理サーバや、仮想サーバへ持たせることができ、更に、その複製(リプリケーション)も持たせることが出来ます。このことで、On-siteでの障害に対し、Off-Siteでのリカバリー、事業継続が可能になります。下図はローカルドメインを超えてバックアップデータのリプリケーションサイトを構成した図です。



6.3 バックアップデータからのデータクリエーション

一旦ベースボリュームイメージとそれ以降のSnapshotデータを持つCDPサーバは、ネットワークのいかなる場所の物理サーバや仮想サーバへもそのバックアップデータを送り、ユーザデータをロールアップ(保存データ+更新データ)により最新のデータの状態にすることが可能です。また、システムディスクも最新の状態に維持し、プライマリーのサーバに障害が発生した場合でも直前の状態で業務を継続することを可能にします。(下図参照)



7. まとめ

以上のように、単にデータのバックアップと言うだけではなく、アーカイブデータへの高速データアクセス、ダウンタイムの無いサービス継続、クラウド環境でのデータ保全等、ITマネジメントにとっては多くの課題が目の前に山積しています。CDPはITインフラを構築する上で、データ保護と、過去のいかなる時点へも遡ってデータアクセスを可能にする不可欠なソリューションになりつつあります。

弊社は現在CDPソリューションとネットワークストレージに関し、種々検証を重ねています。今後、みなさまにこれらの成果をご紹介します。ご期待下さい。