

バックアップの高速化 における一手法

4B-7

藤武三夫, 太田貞夫
富士通(株)

1. はじめに

近年のデータ量の急速な増加によりDASD装置台数が増大し、以下のような問題が発生してきている。

- バックアップ時間の増加
- ストレージ管理コストの増加
- 設置スペースの不足

特にバックアップ時間は、翌日のオンライン業務に支障をきたす程増加してきている。

本稿では、バックアップ時間の削減を目的とし、バックアップ高速化の一手法として考案した差分バックアップ手法を紹介する。

2. 見解を角解決のアプローチ

従来のバックアップ処理ソフトウェア(例:富士通のARCS, TDUMP)は、ハードウェア性能をフルに活かして処理しており、ソフトウェアの改善で処理時間を削減するのは困難である。したがって、バックアップ処理するデータ量を削減することによって時間を削減する方法を考えてみる。

現状のバックアップ運用は、バックアップ処理の都度DASD装置上の全てのデータを処理している。あるバックアップ処理より以降に更新したデータも、更新しなかったデータも一様に次のバックアップ処理で処理している。未更新データは、実質的に1回バックアップすれば十分であり、何回もバックアップする必要はない。よって、未更新データをデータ量削減対象とする。

バックアップ処理後の未更新データは処理せず、更新データのみを処理するためには、どのデータが更新データかを認識できるような機構がDASD装置に必要となる。更新データの認識ができれば、バックアップ処理ソフトウェアで該当データのみを処理すればよいということになる。

3. 見解を角解決の手段

a) ハードウェア

ハードウェアは、更新されたトラック番号(更新位置情報:以降、更新ビットマップという)を記憶するメモリをファイル制御装置内(FCU)に置く。更新ビット

マップは、FCU内のNVS(不揮発性メモリ)で構成し、FCUファームウェアで管理/制御する。以下に、問題解決のためにハードウェアに必要な機能を示す。

- 更新ビットマップサポートの認識
- 更新ビットマップの初期化
- 更新ビットマップの自動更新(ソフトウェアは関与しない)
- 更新ビットマップの読み込み
- 離散トラックの高速読み込み

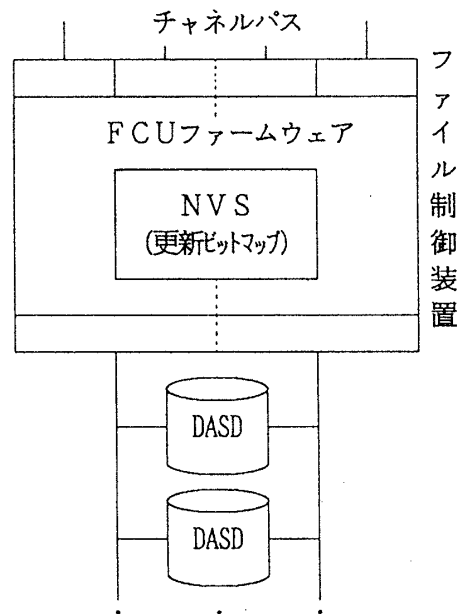


図1. ファイル制御装置内の更新ビットマップ

b) ソフトウェア

ソフトウェアは、全データのバックアップ(以降、全バックアップという)と以降の更新データのバックアップ(以降、差分バックアップという)を組み合わせて運用する。

全バックアップ処理は、差分機能を有するDASD装置を認識すると、更新ビットマップを初期化し、データの存在するトラックを全てバックアップする。全バックアップ処理を行うことにより、以降の差分バックアップ処理が有効となる。

差分バックアップ処理は、次に示す処理方式のうちいずれかが選択できる。

- 累積型
- 非累積型

累積型処理は、差分機能を有するDASD装置を認識すると、更新ビットマップを読み込み、ビットに対応するトラックをバックアップする。したがって、累積型は毎回の差分バックアップで全バックアップ以降の全ての更新データを処理する。累積型の特徴は、差分バックアップを行うにつれて、データ量と処理時間が次第に増加していくが、バックアップ媒体の管理やリカバリ処理は容易にできることである。

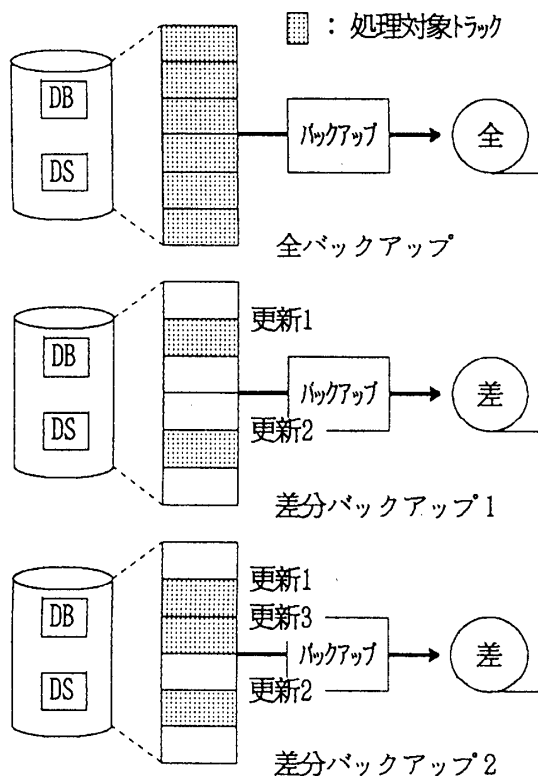


図2. 累積型差分バックアップ

非累積型処理は、差分機能を有するDASD装置を認識すると、更新ビットマップを読み込み、ここで更新ビットマップを初期化しなおして、ビットに対応するトラックをバックアップする。したがって、非累積型は毎回の差分バックアップで直前の差分バックアップ以降の更新データだけを処理する。非累積型の特徴は、最近の更新データのみを処理するのでデータ量や処理時間が増加することはないが、差分バックアップを行うにつれてバックアップ媒体が次第に増加するので、媒体管理やリカバリ処理が複雑になることである。

累積型と非累積型のどちらの処理方式を選択するかは、処理の対象となるデータの性質によって決定する。つまり、一定時間で区切った場合のデータの更新状況を調べ更新の重複度が高い場合は、累積型の方が有利であり、重複度が低い場合は、非累積型の方が有利である。

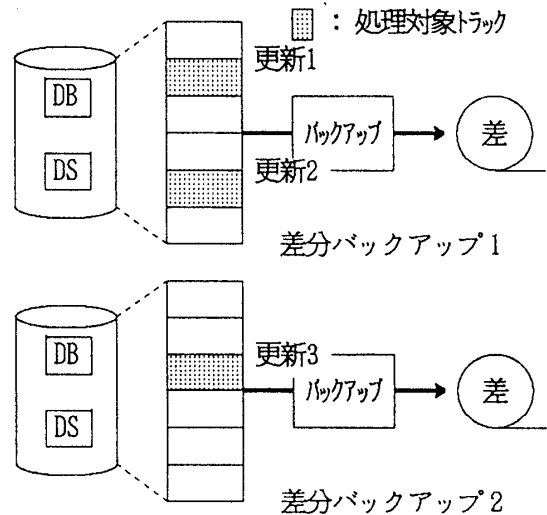


図3. 非累積型差分バックアップ

4. 効果

差分バックアップは、処理するデータ量を削減するので、処理時間も削減できる。

全バックアップでは、例えば、DASD中の1.2GBのデータを容量200MBのCMTにバックアップした場合、処理時間10分、CMT6本を必要とする。従来は、毎回のバックアップでこれだけの処理時間とバックアップ媒体を必要としていたことになる。

通常、時間間隔を一週間とした場合のDASDの更新率は、10~20%である。よって、非累積型の毎回の差分バックアップでは、時間1~2分、CMT1~2本で処理できる。累積型の差分バックアップでは、初回は非累積型と同じ値で処理でき、次回から時間、CMT本数ともに次第に増加していく。ただし、増加する伸び率はデータ更新の重複度に反比例する。

5. おわりに

本稿では、バックアップ処理を高速化する差分バックアップ手法を紹介した。今後もハードウェアとソフトウェアの連携によるさらなる高速化の方法を模索していきたい。

なお、累積型の差分バックアップ手法は、GSM/DARCV (Global Storage Management / Data ReCovery services) で実現している。