

論文の内容の要旨

論文題目 テープマイグレーション機能を有する三次記憶装置におけるファイル管理手法とその衛星画像データベースシステムへの適用に関する研究

氏 名 根本 利弘

近年のプロセッサや通信回線の速度の向上にともない、計算機システムにおいて扱われるデータの容量は急速に増大している。二次記憶装置は急激に容量を増し、価格を下げつつあるものの、膨大なデータをアーカイブするには未だ充分ではなく、現状では磁気テープドライブ装置とライブラリ装置を組み合わせた三次記憶装置が用いられている。しかしながら、現在の商用の三次記憶装置では、高速なアクセスを実現するためにデータ転送速度の高いテープドライブ装置、高速動作可能なロボットアームが用いられており、非常に高価なシステムとなっている。さらに、三次記憶システムに記録される大規模データのアクセス特性を考慮していないため、高速なデバイスを用いながらも、その能力を十分に活かしているとは言いがたい。三次記憶システムの応答性能を下げる要因となりうる大規模データのアクセス特性としては、ファイル全体ではなくそのファイルの一部分のみが参照される傾向があるという部分参照性、および、多数のファイルの中で、各ファイルが均等にアクセスされるのではなく、各ファイルのアクセス頻度に偏りが存在するという参照局所性の2点が挙げられる。ファイル内の一部分のみが参照される部分参照性については、今日の三次記憶システムではデータはファイル単位で

管理されるため、ファイル内の一部分のみが必要である場合においてもファイル全体をテープからディスクへマイグレートし、不要な部分のマイグレートに長い時間を要してしまう。また、各ファイルのアクセス頻度に偏りが存在するという参照局所性に関しては、アクセス頻度の高いファイルと低いファイルが1本のテープ内に混在し、高アクセス頻度ファイルがテープ内に散在している場合、それらのファイル間のシークを繰り返すこととなり応答性能が悪化する。また、複数のアーカイバが存在する環境下においては、テープ間のアクセス頻度の偏りも問題となる。高アクセス頻度テープが多く存在するアーカイバのドライブのみが使用され、テープドライブ装置やロボットアームなどの資源の利用率が低下する。

本研究は、これらの三次記憶システムにおける問題点を考慮し、スケーラブルテープアーカイバと名付けたアーカイブシステムを提案するとともに、大規模三次記憶システム上のファイルに対するアクセス特性を考慮した3つの高性能化手法を提案し、その基本性能、および衛星画像データベースシステムへ適用した際の性能について評価、考察を行うものである。

大規模アーカイブシステムの価格の問題に対し、エレメントアーカイバと名付けた小規模アーカイバを、エレメントアーカイバ間で物理的にテープの移送を可能とする移送装置により複数台接続することで構成するスケーラブルテープアーカイバを提案し、安価に大規模アーカイブシステムを構築する方法を示した。コモディティ化された小規模アーカイバを用いることで大規模なアーカイブシステムを安価に構築でき、また、エレメントアーカイバ数を変更することにより、任意の規模のアーカイブシステムの構成を可能とする。後述するホットデクラスタリング手法を適用することにより、64台までエレメントアーカイバを接続しても応答性能が損なわれることはない。

三次記憶システム上の大規模ファイルに対するアクセス特性を考慮した高速化手法として、第一に、ファイルの一部分のみが参照される部分参照性の問題に対し、ファイルをブロックに分割して管理するとともに、必要な部分のみをマイグレートすることによりマイグレートに要する時間を短縮する部分マイグレーション機能を提案した。部分マイグレーション機能を有するファイルシステムの構築法を示すとともに、部分マイグレーション機能に加えてプリフェッチ機能も有する試作システムを構築した。I/O処理に対して計算処理の負荷が高い放射量・幾何補正処理、計算処理に対してI/O処理の負

荷が高い植生指数生成という2つの実際の衛星画像処理において用いられる実アプリケーションを試作システム上において実行し、放射量補正においてはディスク上にデータが存在する場合とほぼ同じ処理時間で、植生指数生成においては必要なサイズのデータのテープからの読み込み時間と同等の時間で処理が完了し、従来のファイルシステムに比べて大幅に処理時間を短縮することが可能であることを明らかにした。

第二に、スケーラブルテープアーカイバにおけるテープ間のアクセス頻度の偏りによる参照局所性の問題に対し、ホットデクラスタリングと名付けた、熱と温度の概念を利用した負荷分散手法を提案した。ホットデクラスタリングは、スケーラブルテープアーカイバにおけるアクセス頻度の抽象である熱の概念を用いた負荷分散手法である。使用可能なドライブ装置を持たないエレメントアーカイバ内のファイルに対して新たなリクエストが生じた際に、そのファイルが存在するメディアを他のエレメントアーカイバ内の空きドライブ装置まで移動させてサービスを行うフォアグラウンドマイグレーション、および各エレメントアーカイバの過去一定期間のアクセス数（熱）を均衡化することで負荷の分散を図るとともに、フォアグラウンドマイグレーションの実行を円滑化するために各エレメントアーカイバ間の空きスロット数を平衡化するバックグラウンドマイグレーションの2つのマイグレーションによってホットデクラスタリングは構成される。フォアグラウンドマイグレーションは応答性能を向上させるのに対し、バックグラウンドマイグレーションはスケーラブルテープアーカイバの負荷均衡状態への収束時間を短縮する。また、ホットデクラスタリングはテープドライブの故障に対する信頼性も向上させる。故障により使用可能なテープドライブを有しないエレメントアーカイバが存在する場合においても、スケーラブルテープアーカイバはリクエストに応えることが可能である。また、ホットデクラスタリングは、従来よりテープシステムの高速度手法として用いられているファイルスライピング手法とは独立に適用可能であり、スケーラブルテープアーカイバにおいてファイルスライピングを用いた場合においても、ホットデクラスタリングは応答時間を短縮する。

第三に、同一テープ上のファイル間のアクセス頻度の偏り、およびテープ間のアクセス頻度の偏りによる問題に対し、ホットレプリケーションと名付けた手法を提案した。ホットレプリケーションは、テープを巻き戻すことなくロード/イジェクトが可能なテープドライブ装置を使用するという条件の下、高アクセス頻度データ（ホットデー

タ)の複製を予め確保しておいた空き領域に作成する手法である。ホットデータの複製をクラスタリングすることにより、ホットデータが連続してアクセスされる場合のシーク長を短縮し、応答時間を短縮する。また、アクセス要求されたデータが複製を持たず、そのファイルが記録されているテープが他のリクエストにより使用されている場合にはそのリクエストを直ちにサービスすることはできないが、異なるテープ上に複製を作成することによりオリジナルデータにアクセスできない場合においても複製をアクセスすることを可能とし、応答時間を短縮する。

東京大学生産技術研究所においてインターネットを通じて公開している衛星データのブラウジング用縮小画像に対するアクセス履歴を用いてシミュレーションを行い、ホットデクラスタリング、ホットレプリケーションの有効性を示した。この衛星画像に対するアクセスは、多数の利用者からの最新画像に対するアクセス、および一人の利用者からの過去一定期間内の画像への一括したアクセスが多いという特徴を有する。これらの特徴は一般的な衛星画像データベースのアクセスと共通すると考えられる。シミュレーションの結果によると、ホットデクラスタリングは十分な量を持つディスクによるキャッシュ以上に応答時間を短縮可能であることを示した。さらに、ディスクによるキャッシュと相補的に働くため、両者を併用することにより一層高速化が図れる。また、ホットレプリケーションは、十分な容量を持つキャッシュが存在する場合、存在しない場合のいずれにおいても最大で30~40%程度応答時間を短縮するとともに、ホットデクラスタリングと併用することによりさらに応答性能を向上させる。特に、ホットデクラスタリング、ホットレプリケーションは、キャッシュの効果が小さい一人の利用者からの過去一定期間内の画像への一括したアクセスに対して有効である。このようなアクセスは、リクエストされたデータは直前にアクセスされていないためキャッシュは有効ではなく、また、少数のテープに多数のリクエストが集中し、応答時間の劣化の原因となるが、ホットデクラスタリングはリクエストされたテープを移動してスケーラブルテープアーカイバに分散させることにより、また、ホットデクラスタリングは異なるテープ上の複製にアクセスすることにより応答時間を短縮する。