

ストレージ階層仮想化機能における即応性の高いデータ移動方式 Agile Data Migration for Virtual Storage Tiering Function

山本 貴大[†]
Takahiro Yamamoto

坏 弘明[†]
Akutsu Hiroaki

揚妻 匡邦[†]
Masakuni Agetsuma

川口 智大[†]
Tomohiro Kawaguchi

1. はじめに

様々な業務分野の情報がデジタル化され、データ量は年々増加している[1]。その一方で、企業のIT投資額は、増加が見込めない状況にある[2]。このため、企業は、ITシステムの性能要件とコスト要件を満たすため、データのアクセス頻度に応じて、適切なストレージ階層（高価だが高速なSolid State Drive (SSD) や安価だが低速なHard Disk Drive (HDD)）にデータを格納することで、データ保持コストを削減している。しかし、アプリケーションからの時々刻々と変化するアクセスの頻度や範囲といった入出力特性に追従して、データを格納するストレージ階層を変更することは困難である。この解決手段として、サーバからの入出力に同期して、データへのアクセス頻度に応じたストレージ階層にデータを移動するストレージ階層仮想化機能がある[3]。

本稿では、ストレージ階層仮想化機能において、入出力特性の変化に即応してストレージ階層間でデータを移動する方式について述べる。

2. 目的

ストレージ階層仮想化機能は、データへのアクセス頻度を監視し、サーバからの入出力に同期して、高アクセス頻度のデータをHDDからSSDに移動することで、ストレージ装置の入出力速度を高める。ストレージ階層仮想化機能の処理の流れを図1に示す。

- (1) ストレージ装置は、サーバからストレージ装置内のデータへの入出力の頻度を監視し、アクセス頻度管理テーブルに記録する。(アクセス頻度の監視)
- (2) ストレージ装置は、サーバからの入出力要求に同期して、監視結果からデータを格納すべきストレージ階層を判定する。(階層判定)
- (3) ストレージ装置は、(2)で階層判定したデータを判定したストレージ階層に移動する。(データ移動)

データ移動において、HDDからSSDへの移動をプロモーション、SSDからHDDへの移動をデモーションと呼ぶ。ストレージ階層仮想化機能では、入出力特性に変化が生じてから、SSDへのアクセス率を最大化するまでに要する時間（以降、追従時間）が長くなると、例えば、サーバで稼働するデータベースが処理できるトランザクション総数が減少してしまう。そこで、短い追従時間を実現するデータ移動技術の確立を目的とした。

3. 課題と目標

ストレージ階層仮想化機能は、SSDに空き容量がある場合、入出力特性の変化後、即座にプロモーションを実行できる。一方、SSDに空き容量が無い場合、プロモーションを実行する前にSSDの空き容量を確保するため、デモーションを実行する必要がある。従って、プロモーション速度が低下し、追従時間が長くなる問題があった。そこで、

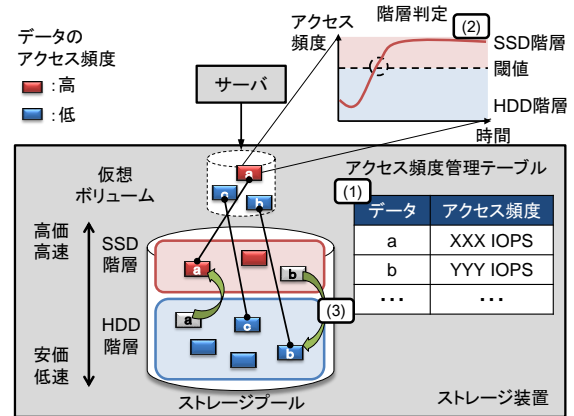


図1 ストレージ階層仮想化機能の概要

SSDに空き容量が無い場合でも、SSDに空き容量がある場合と同程度の追従時間を達成することを目標とした。

上述の問題を解決するため、SSD容量の一部をバッファ容量とし、入出力特性の変化前に、バッファ容量分のデータを予めデモーションしておくアプローチをとる。しかし、本稿が対象とするストレージ階層仮想化機能では、サーバの入出力速度の低下を防ぐため、データ移動の最大同時実行数を一定に制限し、且つプロモーションとデモーションでデータ移動に使用するリソースを共有するという制約があった。このため、無制限にデモーションを実行できず、以下が課題となった。

(課題1) SSDに空き容量を確保するために、高頻度にデモーションを実行すると、デモーションがリソースを占有してしまい、プロモーション速度が低下する。

(課題2) 多量のデータを即座にプロモーションするために、バッファ容量を大きく設定すると、永続的に利用可能なSSD容量が減少する。

4. 解決方式

(課題1)を解決するためには、プロモーション速度への影響が小さい機会に、デモーションを実行する必要がある。デモーションを実行するにあたり、プロモーション速度への影響が小さい機会として、次の2つが考えられる。

- (機会1) プロモーションを実行していない時
プロモーションは、入出力特性が安定すれば、発生しない。プロモーションを実行していない時であれば、デモーションを実行してもプロモーションに影響を与えない。
- (機会2) プロモーションを実行しているが、空きリソースが存在する時

[†] (株)日立製作所 研究開発グループ 情報通信イノベーションセンタ, Center for Technology Innovation, Research & Development Group, Hitachi, Ltd.

Oracleは、Oracle Corporation及びその子会社、関連会社の米国及びその他の国における登録商標または商標です。

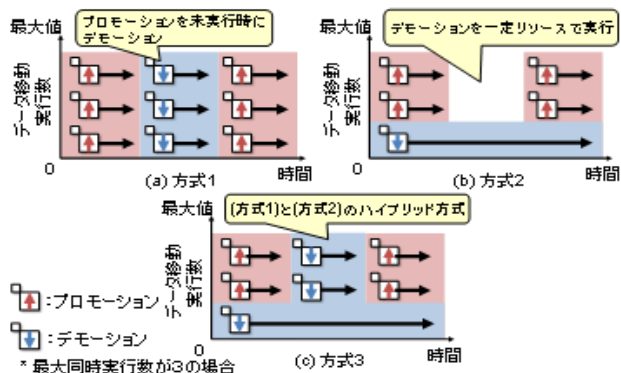


図 2 解決方式の概要

プロモーションは、実行中でも空きリソースが存在する時がある。その空きリソースでデモーションを実行するのであれば、プロモーションに影響を与えない。

次に、各機会を活用した方式について説明する。まず、(機会 1) を活用した (方式 1) を説明する (図 2 (a))。 (方式 1) は、プロモーションを実行中でなければ、リソースを全て使用し、デモーションを実行する。階層判定により、プロモーションが必要となった時、またはバッファ容量を確保し終えた時、デモーションを停止する。(方式 1) は、プロモーション未実行時、デモーション速度が速い利点がある。その一方で、プロモーションを常時実行しているとバッファ容量を確保できない欠点がある。

次に (機会 2) を活用した (方式 2) を説明する (図 2 (b))。 (方式 2) は、一定の限られたリソースで定期的にデモーションを実行する。(方式 2) は、プロモーションを実行中でもデモーションを実行できる利点がある。その一方で、デモーションを実行するリソース分だけプロモーション速度が低下する欠点がある。

最後に (機会 1) (機会 2) を活用した (方式 3) を説明する (図 2 (c))。 (方式 3) は、一定の限られたリソースで定期的にデモーションを実行しつつ、プロモーションの未実行時には、全てのリソースでデモーションを実行する。(方式 3) は、(方式 1) (方式 2) の利点がある。その一方で、(方式 2) 同様にデモーションを実行するリソース

表 1 評価条件

項目	説明
ストレージ構成	SSD x 4, HDD x 16
アプリケーションが高頻度にアクセスするデータサイズ	100 GB
単位時間あたりのプロモーション実行相対時間	0.8

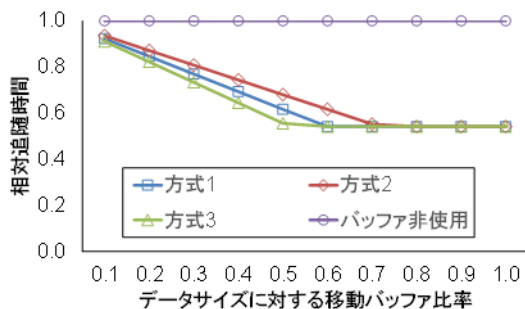


図 3 方式比較

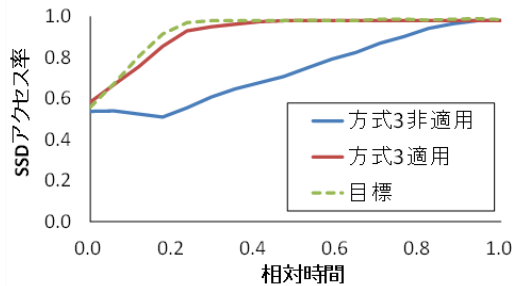


図 4 追従時間の評価結果

分だけプロモーション速度が低下する欠点がある。

次に、(課題 2) を解決するため、小バッファ容量で、最も追従時間を短縮できる方式を明らかにする。そこで、バッファ容量を変化させた時の追従時間を表 1 に示す条件でシミュレートした。図 3 は、横軸にアプリケーションが高頻度にアクセスするデータサイズに対するバッファ容量の比率、縦軸にバッファ非使用時を基準とした相対追従時間を示している。図 3 より、相対追従時間の最小値を達成するにあたり、(方式 3) が最小バッファ容量で達成できると分かった。このため、(方式 3) を採用した。

5. 評価

提案方式の効果を確かめるため、実機により、評価した。アプリケーションとして Online Transaction Processing (OLTP) を想定し、Oracle® データベース [4] と TPC™ Benchmark C (以降、TPC-C) [5] を使用した。TPC-C は、入出力特性の変化を模擬するため、最初に 30 分間実行した後、実行パラメータである倉庫数 (アクセス範囲) と接続セッション数 (アクセス負荷) を増加させた。評価環境は、表 1 のストレージ構成を用い、SSD と HDD の容量比率は 2:8 とし、バッファ容量はデータサイズの 50% とした。

評価結果を図 4 に示す。図 4 は、横軸に相対時間、縦軸に SSD へのアクセス率を示している。図 4 より、(方式 3) 適用時は、(方式 3) 非適用時と比較し、SSD アクセス率 1.0 到達までの時間を 46% 短縮でき、目標である SSD に空き容量がある (つまり、デモーションなしに即座にプロモーション可能な) 場合と同程度の追従時間となった。

6. おわりに

本稿は、ストレージ階層仮想化機能において、入出力特性の変化に即応してストレージ階層間でデータを移動する方式について述べた。評価の結果、提案方式の適用により、非適用時と比較し、追従時間を 46% 短縮でき、目標と同程度の追従時間となることを確認した。

参考文献

- [1] John Gantz and David Reinsel, "The Digital Universe In 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East - United States", IDC country brief (2013).
- [2] GVA connect, "Worldwide IT spending to grow 2.4% in 2015", Datasource, Issue 131 (2015).
- [3] 坏弘明他, "ストレージ自動階層配置機能におけるデータ再配置の最適化", 情報処理学会論文誌, Vol.54, No.4, pp.1592-1608 (2013).
- [4] Oracle Corp., "Data Sheet: Oracle Real Application Clusters (RAC)", Oracle Technology Network (2013).
- [5] Transaction Processing Performance Council (TPC), "TPC Benchmark™ C Standard Specification Revision 5.11", TPC Documentation (2010).