

論文 / 著書情報
Article / Book Information

題目(和文)	高い鮮度で一貫性保証と更新集約適用を行う HTAPシステム高性能化に関する研究
Title(English)	A Study on Performance Enhancement of Highly Data-Fresh HTAP Systems Guaranteeing the Consistency and Aggregating Updates
著者(和文)	塩井隆円
Author(English)	Takamitsu Shioi
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12816号, 授与年月日:2024年6月30日, 学位の種別:課程博士, 審査員:宮崎 純,吉瀬 謙二,小林 隆志,渡部 卓雄,金子 晴彦
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12816号, Conferred date:2024/6/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	塩井 隆円		
論文審査 審査員		氏名	職名		氏名	職名
	主査	宮崎 純	教授		金子 晴彦	准教授
	審査員	渡部 卓雄	教授	審査員		
		吉瀬 謙二	教授			
		小林 隆志	教授			

論文審査の要旨 (2000 字程度)

本論文は、「A Study on Performance Enhancement of Highly Data-Fresh HTAP Systems Guaranteeing the Consistency and Aggregating Updates (高い鮮度で一貫性保証と更新集約適用を行う HTAP システム高性能化に関する研究)」と題し、データベースの更新処理を行う OLTP (On Line Transaction Processing) と、蓄積されたデータの解析処理を行う OLAP (On Line Analytical Processing)をつないだ HTAP (Hybrid Transactional/Analytical Processing)において、データの高い鮮度と一貫性を保証しつつ、処理の性能を向上させる手法について論じるもので、英文7章からなっている。

第1章「Introduction」では、OLTP と OLAP の処理の違いと、HTAP の構成方法、および鮮度を保ちつつ一貫性保証と性能向上を両立させることの困難さを述べ、本論文の貢献についてまとめている。

第2章「Related Work」では、既存研究における HTAP システムのアプローチと、それらにおける課題についてまとめている。

第3章「Preliminaries」では、本論文で提案する手法の前提として、マルチバージョンに基づくデータベースの一貫性保証の枠組み、HTAP における要求事項の理論的解説、さらに現状の HTAP システムにおいて性能的な理由から対応がされていない不具合である ROA (Read Only Anomaly) の発生について述べている。

第4章「Serializability of HTAP systems」では、本論文における高性能化手法のベースとして提案する RSS (Read-Safe Snapshot) の概念と、理論的枠組み、およびその実現方法について述べている。RSS では、一貫性を保証した状態の OLTP 側の履歴 (ログ) を用いて、トランザクション間の依存情報を抽出し、読み出し処理のみを行う OLAP 側で ROA が発生しないように、利用する OLTP でのトランザクションをスナップショットとして抽出する。これによる OLAP 側での処理は OLTP 側へ影響を与えず、一貫性を保ちつつ高い性能を実現することが可能となっている。この効果を確かめるために、OSS (Open Source Software) で、一般に広く用いられている PostgreSQL を OLTP と OLAP で使用し、RSS を実現した HTAP システムの性能評価を行っている。HTAP の評価において多用される CHbenCHmark を用いて、シングルノード構成とマルチノード構成における既存システムと提案システムの比較を行い、一貫性保証をしたシングルノードでの性能が大きく向上し、既存システムでは一貫性保証ができないマルチノードにおいても性能に遜色がないことを示している。また、OLTP で更新したデータが OLAP で扱えるまでの遅延を大幅に短縮できることを示している。

第5章「Data Conversion from row to column」では、HTAP システムにおける性能向上のためのもう一つの課題である行形式から列形式に変換する処理の効率化手法について述べている。OLTP における更新は一般に行単位で行われるために行形式で格納する方が効率が良い。一方、解析処理を行う OLAP では、少ない列に絞って列全体を読み出して解析することが多く、列形式で格納する方が効率が良い。このため、HTAP システムでは OLTP の行単位の更新を、OLAP の列単位のデータに適用することが求められる。特に、第4章で提案している RSS は行単位でのバージョン管理を前提とするため、行形式から列形式への効率的な変換が求められる。提案手法では、列形式のデータをブロックで管理し、更新する行の ID (TID) と列ブロックの ID (LBID) の対応をハッシュテーブルで管理し、ブロック単位で更新集約することで効率化を行う。変換に要する時間を評価するために、行単位の更新に対する標準的なベンチマークである TPC-C を用いて、従来の行列変換を行うデータベースである MariaDB Column Store と、提案手法を MariaDB ColumnStore に適用した場合の比較を行った結果、提案手法により約2倍の性能向上が得られたことを示している。

第6章「Discussion」では、大規模なデータ解析で使われる Spark を OLAP 側に想定した場合等の可能性に関して議論している。

第7章「Conclusion」では、本論文で提案する内容とその効果をまとめ、展開の可能性について論じ

ている。

以上を要するに、本論文は今後益々重要となるデータベースにおいて、データの更新の鮮度を保ちつつ一貫性を保証した上で性能向上を図る手法を提案し、ベンチマークを用いた評価でその効果を示したもので、その適用性、有用性、発展性は高く、その成果は工学上貢献するところが大きい。よって我々は、本論文が博士（工学）の学位論文として十分価値があるものと認める。

注意：「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。