

プログラム名：社会リスクを低減する超ビッグデータプラットフォーム

PM名：原田 博司

プロジェクト名：超ビッグデータ処理エンジン

委 託 研 究 開 発

実 施 状 況 報 告 書 (成 果)

平成 28 年度

研究開発課題名：

超高速動的スケーラブルデータベースエンジンの基盤技術の研究開発

研究開発機関名：

国立大学法人東京大学

研究開発責任者

喜連川 優

I 当該年度における計画と成果

1. 当該年度の担当研究開発課題の目標と計画

担当研究開発課題は、非順序型実行原理を基として、新たな非連続性を産み出すべく、複数ノードへのエラスティシティ（伸縮可能性）を備えた超高速動的スケーラブルデータ処理技術を確立することにより、毎秒 1,000 万回程度のストレージアクセス性能を備えた新たな「超高速動的スケーラブルデータベースエンジン」の実現を目指すと共に、当該データベースエンジンを核として、ImPACT 研究開発プログラム傘下の他のプロジェクト等との連携により、先進的なビッグデータの利活用を可能とするための解析プラットフォームの構築のための検討を進めるものである。東京大学に於いては、平成 28 年度末迄に、「限定版超高速動的スケーラブルデータベースエンジン」の基礎設計を完成させ、毎秒 100 万回程度のストレージアクセス性能の達成の目途を得るとともに、また、先進的なビッグデータの利活用を可能とするための解析プラットフォームの基礎設計を検討し、有力な先進的ビッグデータ応用を対象として、基礎実験を完了することを目指し、研究開発を実施してきた。

上述の目標を達成するために、平成 28 年度は、東京大学に於いて、日立製作所との産学連携の下、「限定版超高速動的スケーラブルデータベースエンジン」の基礎設計（基本方式検討、基礎実験）を実施し、また、先進的なビッグデータの利活用を可能とするための解析プラットフォームの基礎設計ならびに基礎実験を実施することを計画していた。また、プロジェクト全体の統括と総合的な推進を実施することを計画していた。

2. 当該年度の担当研究開発課題の進捗状況と成果

2-1 進捗状況

平成 28 年度に於いては、東京大学に於いて、非順序型実行原理を基として、複数ノードへのエラスティシティ（伸縮可能性）を備えた超高速動的スケーラブルデータ処理技術を確立することにより、毎秒 100 万回程度のストレージアクセス性能を備えた「限定版超高速動的スケーラブルデータベースエンジン」の実現を目指し、東京大学に於いては、日立製作所との産学連携の下、当該データベースエンジンの基礎設計（基本方式検討、基礎実験）を実施した。即ち、従前の非順序型実行原理に基づく超高速データベースエンジンは原則的に単一ノードでの動作に限られていたのに対して、データベースオペレーションを複数ノードに於いて動的に実行可能とするべく、新たな実行方式を設計し、試作ソフトウェアとして実装し、パブリッククラウドに於いて性能試験を行った。この結果、後述する通り、毎秒 100 万回程度のストレージアクセス性能の達成の目途を得た。

また、超高速動的スケーラブルデータベースエンジンを核として、先進的なビッグデータの利活用を可能とするための解析プラットフォームの設計と部分構築を目指し、東京大学に於いては、ImPACT 研究開発プログラム傘下の他のプロジェクトの参画機関等との連携ならびに日立製作所との産学連携の下、基礎設計の検討を進め、有力な先進的ビッグデータ応用を取り上げて、基礎実験を実施した。即ち、ヘルスセキュリティプロジェクトから提供された全国規模ならびに都道府県規模の医療保険レセプトビッグデータの利活用を具体的な課題として取り上げ、基礎実験を実施するべく、複数の解析応用を開発し、ヘルスセキュリティプロジェクトに提供し、ヘルスセキュリティ分野に於ける新たな知見の獲得を可能とした。

加えて、超高速動的スケーラブルデータベースエンジンの研究開発に掛かるプロジェクト全体の連携を密とし円滑に運営していくため、東京大学に於いて、ステアリング委員会を9回、技術検討会を20回開催する等により参画各機関の連携と調整にあたった。特に、プロジェクト全体の進捗状況を確認しつつ計画の合理化を検討し、必要に応じて国内外に研究者を派遣して調査を行う等により、プロジェクト全体の統括と総合的な推進を行った。

以上、このように研究開発の進捗は良好と言える。

2-2 成果

上述の通り、超高速動的スケーラブルデータベースエンジンの試作ソフトウェアに関して、パブリッククラウドに於いて性能試験を行った。この際、キーバリュ型共有ストレージ上にデータベースを構築し、標準ベンチマークである TPC-H に基づくデータセットを組み込み、サーバインスタンス数を変化させて、試験問合せの実行性能を測定した。サーバインスタンス数を増加させ、80 インスタンス程度まで良好な性能向上を確認することができた。80 インスタンスを用いた場合には、最大で毎秒 137 万回程度のストレージアクセス性能を確認し、単一インスタンスに於いて従来型の非順序型データベースエンジンを実行した場合と比較して 37 倍程度の性能向上が得られた。即ち、年度目標であった基本設計に基づく毎秒 100 万回程度のストレージアクセス性能の達成の目途については、これを上回る成果を達成することができた。一方、先進的なビッグデータの利活用を可能とするための解析プラットフォームの構築については、上述の通り、ヘルスセキュリティビッグデータに基づく複数の解析応用を開発し、当該応用から、高齢者に於ける疾患の重症化過程等、ヘルスセキュリティ分野に於いて従前は未知であった事象の解明に糸口を得ている。即ち、有力な先進的ビッグデータ応用を対象として基礎実験を完了すると年度目標を超えて、成果を得ることができた。

以上、このように目標を上回る成果が得られているところである。今後更なるデータベースエンジンの性能向上とビッグデータの新たな利活用の開拓に向けて研究を強力に推進する。

2-3 新たな課題など

既に述べた通り、当初の計画を上回るペースで研究が進捗していることから、平成 28 年度途中に(1)「限定版超高速動的スケーラブルデータベースエンジン」の目標性能を毎秒 200 万回程度のストレージアクセスに向上し、パブリッククラウドに於ける適用と実証を実施する、(2) 地方自治体等へのヘルスセキュリティ解析サービス提供実験による実証実験の実施の 2 柱からなる研究開発の強化案を作成し、追加予算の措置をお願いし、概ねお認め頂くことができたことから、平成 29 年度より研究開発を一層強化して推し進める予定である。

3. アウトリーチ活動報告

技術的優位性の確保を優先して推進し、研究成果の広汎な展開を目指した体制の構築に向けて、先行的な検討を進めている。また、ヘルスセキュリティプロジェクトとの連携を核として、本研究成果の適用が特に期待される健康・医療分野の多様な応用について、検討を進めているところである。