

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 29 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（ただし、予算制約から検証は毎秒300万回程度の水準を目指すこととしている）、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。上述の目標を達成するために、平成29年度は、東京大学に於いて、毎秒100万回程度（ただし、東京大学に於いては、研究開発を強化することにより、毎秒200万回程度とする）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、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機能設計ならびに評価を実施し、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、基礎実験ならびにセキュア実験システムの部分構築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29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（ただし、東京大学に於いては、研究開発を強化することにより、毎秒200万回程度とする）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機能設計ならびに評価を実施した。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、前年度に設計を行った新たな実行方式を改良し、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200万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。この際、東京大学に於いては、医療経済研究機構との連携の下、自治体等への解析サービス提供実験のためのセキュアシステムの部分構築を実施した。

加えて、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3回、技術検討会を14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。

## 2-2 成果

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、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（スケールファクタ1,000規模）を組み込み、各種設定パラメータを変化させて、複数の試験問合せの実行性能を測定した。この結果、80インスタンスを用いた場合に、最大毎秒210万回程度のストレージアクセス性能を確認し、この際、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66倍の性能向上が得られることを確認した。即ち、年度目標であった基本設計に基づく毎秒200万回程度のストレージアクセス性能の達成の目的について、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関しては、従前の全国を対象とする解析規模が1年分に留まっていたのに対し、新たにデータセットが利用可能となり解析規模を拡大し、6年分の全国データセットの解析を可能とする基盤を構築した他、ヘルスセキュリティプロジェクトと議論を重ね新たに複数の解析応用を開発して当該プロジェクトに提供し、ヘルスセキュリティ分野の新たな知見獲得に繋げた。また、都道府県を対象とする解析についても解析規模を5年程度分に拡大した。更に、自治体等への解析サービス提供実験のためのセキュアシステムの構築を進めた。ヘルスセキュリティプロジェクトと共同で、従前に開発した解析サービスから提供サービスを選定し、また、必要な改良を検討し、サービス提供実験のための試験を行ったほか、潜在ユーザへのヒアリング等を並行して進める等、構築は順調に進展した。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し、自治体等への解析サービス提供実験のためのセキュアシステムの部分構築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、(1)「本格版超高速動的スケーラブルデータベースエンジン」に於けるパブリッククラウドを対象とする実証試験の性能目標を毎秒600万回程度のストレージアクセスに向上する、(2)自治体等への解析サービス提供実験の規模を拡大する、(3)ヘルスセキュリティプロジェクトで進められているマクロシミュレーションを強化するために基盤を構築する、(4)超高速匿名加工処理技術に関する実証試験を実施するの4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、平成29年度途中で概ねお認め頂くことができた。平成30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

研究成果の広汎な展開を目指した体制の構築を進めている。また、ヘルスセキュリティプロジェクトと共同で行った地方自治体を対象とする健康・医療ビッグデータ解析に関する報道発表には、反響が高かった。