

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実用化技術の研究開発

研究開発機関名：

株式会社日立製作所

研究開発責任者

松並 直人

# 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

### 【当該年度計画】

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、日立製作所に於いては、東京大学との産学連携の下、当該データベースエンジンの基礎設計（実用システムへの適用を意識した実現方式検討）を実施する。

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、日立製作所に於いては、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東京大学との産学連携の下、実用化のための基礎設計の検討を進める。

### 【当該年度目標】

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。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実用化のための基礎設計の検討を完了する。

## 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

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採用する「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（実用システムへの適用を意識した実現方式検討）を完了した。非順序型実行原理を採用する日立製商用データベースエンジンである Hitachi Advanced Data Binder(HADB)をベースに、複数ノードで処理可能なように拡張する。また、仮想化環境を活用し処理ノード数を変更する。

エラスティシティに必要とされる、仮想化環境における処理ノード数増加による性能スケーラビリティ検証を目的とした予備実験を行った。その結果、10 処理ノードまで性能がスケールすることを確認した。本予備実験環境におけるストレージアクセス性能は最大でおよそ毎秒 60 万回であり、ストレージ性能増強により毎秒 100 万回程度のストレージアクセス性能は達成可能との見通しを得た。（詳細は 2-2 成果に記載）

開発技術のヘルスケア分野への適用を想定し、SS-MIX2 規格の大規模診療データ利活用、レセプトデータ利活用、診療画像データベース等、を題材に、この分野においてビッグデータ解析プラットフォームの実用化に必要とされる要件を調査した。その結果を基に、基礎設計の検討を完了した。調査結果として挙がる大きな課題は以下の通りである。i) データ変換：データフォーマットが統一されていても分析に適する形への変換は簡単ではない。ii) システム構成の最適化：分析ソフトウェア(R 等)を含めた最適化が必要である。iii) 性能管理：複数ユーザがデータベースを共有する等の多くの性能変動要因があるが、ユーザは一定の性能を期待する。

## 2-2 成果

エラスティシティに必要とされる、仮想化環境における処理ノード数増加による性能スケーラビリティの検証を目的とした予備実験を行った。実験環境を図1に示す。Query実行に際し処理ノード数として2以上が指定された場合、DBエンジン(master)は与えられたQueryを指定処理ノード数分のSub queryに分割し、自身で実行する分を除きDBエンジン(slave)に処理を要求する。処理ノード数を1から10まで変化させた場合の結果を図2に示す。図の縦軸は処理ノード数が1のときの性能を基準とした相対性能を示し、図の横軸はDBエンジンの処理ノード数を示す。処理として、1表全走査処理(Table scan)、2表ハッシュ結合処理(Hash join)、2表Nested-loops結合処理(Nested-loops join)の3種類を実行した。非順序実行原理の効果が最も高いNested-loops結合処理でスケーラビリティが一番良く、10ノードまで処理ノード数比に近い相対性能を発揮することを確認した。他の処理においても、処理ノード数に対して良いスケーラビリティを有することを確認した。なお、本実験環境ではおよそ毎秒60万回のストレージアクセス性能を有するストレージを利用している。処理ノード数を更に増やして限界性能に達した場合、ストレージアクセス性能が律速要因となることを確認した。ストレージのアクセス性能を強化することにより、毎秒100万回程度のストレージアクセス性能を達成可能との見通しを得た。

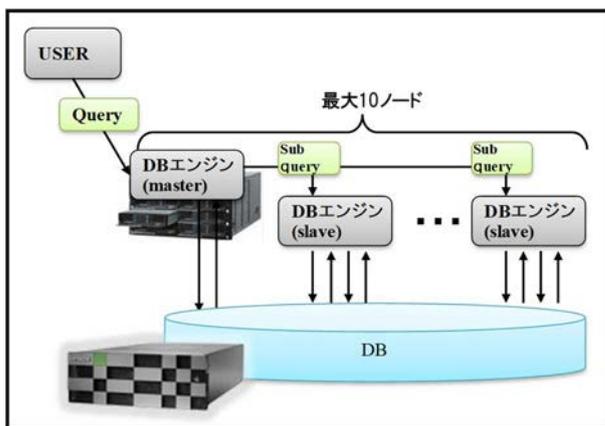


図1 実験環境

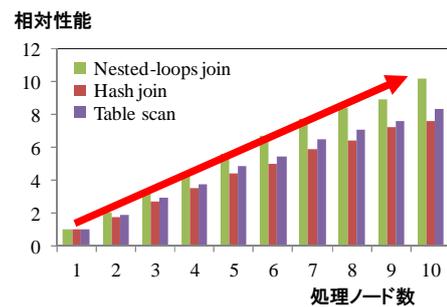


図2 処理ノード数と相対性能

## 2-3 新たな課題など

実用化に向けた要件調査の結果、ビッグデータ解析プラットフォームの実用化においては、データベースの単一処理性能もさることながら、複数ユーザが同時に処理する際の性能維持、等が重要であることを確認した。平成29年度以降、複数ノードへのエラスティシティをこの観点で活用する方法に関しても検討する。

## 3. アウトリーチ活動報告

なし。