

発表概要

ストリーム処理系を活用した リアルタイム・ビッグデータに対するSQLクエリーの高速化

今井 晴基^{1,a)} 川人 基弘¹ 古関 聡¹ 小松 秀昭¹

2014年7月30日発表

各種センサなどから送信される大量のデータのリアルタイム解析が求められているが、既存のデータベース技術ではデータのスキャン、演算に時間がかかり、クエリー結果をリアルタイムに取得するのは困難である。本発表では、そのようなリアルタイム・ビッグデータに対するSQLクエリーの高速化手法の提案を行う。本手法は、入力SQLの構文解析とテーブル属性からストリームデータ処理を行うサブクエリーを決定し、それをストリーム処理コードに変換・コンパイル後、実行モジュールをストリーム処理実行系にデプロイする。そのストリーム処理の結果と残りのクエリー部分を組み合わせ、SQL全体のクエリー結果を得る。ストリームデータ処理のサブクエリーを実際のクエリー発行前に実施することにより、高速化が実現される。これにより、SQLを利用してリアルタイムのストリームデータと既存データベースを透過的に処理することができるようになる。本発表においては、リアルタイム・マーケティングの事例として、現在の車両状況と運転者の嗜好情報データを利用し、商品割引クーポンの発行対象車両を決定するSQLクエリーを用いて、本手法の評価を行った。3,000台の車両データを利用した場合、インメモリーデータベースへのSQLクエリーに比べて、約7,400倍の応答性能を実現可能であることを確認した。

Accelerating SQL Queries for Real-time Big Data Using Stream Processing

HARUKI IMAI^{1,a)} MOTOHIRO KAWAHITO¹ AKIRA KOSEKI¹ HIDEAKI KOMATSU¹

Presented: July 30, 2014

We need real-time analysis for massive data sources such as various sensor units, but traditional database technologies cannot process such data in real time because scanning and calculating the data is too slow. We have devised a faster method for SQL queries against the real-time big data. Our method extracts sub-queries for the streaming data depending on structure analysis and table attributes, transforms the sub-queries into compiled stream processing code, and then deploys the executable to a stream processing engine. The results of the stream processing are then integrated with the rest of the original query to produce the complete results. Our approach accelerates response times by executing the sub-query for the streaming data before issuing the actual query. By using our method, we can access streaming data and existing database transparently via SQL. In this presentation, we evaluate this method using SQL queries for a real-time marketing application that selects drivers to be targeted with discount coupons. The targeting is based on each vehicle's current situation and the driver's preferences. With data from 3,000 vehicles, we improved the response times to 1/7,400 of the times required for SQL queries with an in-memory database.

¹ 日本アイ・ビー・エム株式会社東京基礎研究所
IBM Research, Tokyo, Koto, Tokyo 135-8511, Japan

^{a)} imaihal@jp.ibm.com