

異種データベースの仮想化技術 (NoSQL データベース”Cassandra”の仮想化)

杉山 暢彦 澤本 潤 杉野 栄二 瀬川 典久*

概要. 本研究では, Apache のオープンソースプロジェクトである NoSQL データベース Cassandra の仮想化を行った. 近年, 大量のデータを高速に処理できるデータベースとして NoSQL が注目され, Web サービスを提供している企業等では実際に導入や開発が行われている. 大量のデータを扱う環境として, 環境モニタリング等のユビキタスセンサネットワーク環境がある. 同環境において今後大量のデータを高速に処理できる NoSQL が採用されると考えた. ユビキタスセンサネットワーク環境で収集されたデータは様々なデータベースによって管理されている. データベースは製品によってそれぞれ操作方法やデータモデル等が異なり, 複数のデータベースを扱う場合学習の手間コスト高くなってしまふ. しかし, 複数のデータベースを一つの操作方法に統一する事で, ユーザの学習コストを減らしデータベース利用の効率化が可能となる.

1 はじめに

今日では, センサネットワークは安否確認や環境モニタリング等様々な場面で利用されている. このユビキタスセンサネットワーク環境で収集されたデータはデータマイニング技術により傾向や知識の発見に活用されている. しかし, センサーから収集されるデータ量は膨大なため, 従来のリレーショナルデータベース (RDB) では処理に時間がかかってしまうといった問題がある. そのため, 膨大な量のデータを処理できるデータベースの導入が必要と考えられる.

ユビキタスセンサネットワーク環境では様々なデータベースが扱われている. データベースは製品によってデータモデルや操作方法が異なる. そのため複数のデータベースを扱うユーザは各データベースの操作方法やデータモデルの理解が必要となり学習コストが高くなってしまふ.

データベース利用の効率化の先行研究として, 渡辺らによる異種データベースの仮想化技術の先行研究 [1] がある. これは, データモデルや操作方法の違いをユーザに意識させないために, XMLDB を基準にデータモデルや操作方法を統一したものである. しかし, 現在のユビキタスセンサネットワーク環境で収集される大量のデータを処理できるデータベースは対応していなかった.

そこで本研究では, ユビキタスセンサネットワーク環境におけるデータベース利用の効率化のために, NoSQL データベース Cassandra を XQuery で操作

するシステムの提案と試作を行った.

2 関連研究

ユビキタスセンサネットワーク環境におけるデータベース利用の効率化の研究として, 渡辺らによる異種データベース仮想化技術 [1] がある. この研究では, RDB と XMLDB を対象にデータモデルや操作方法の差異を XMLDB に統一する手法を提案している. データベースにより異なるデータモデルや操作方法として XMLDB のクエリである XQuery を採用し, データモデルを XML Schema とした. これによりユーザは XMLDB のデータモデルと操作方法を覚えるだけで RDB と XMLDB の 2 つのデータベースの操作が可能となった. しかし, 現在のユビキタスセンサネットワーク環境で収集されるデータは膨大なため, RDB や XMLDB では処理しきれないという問題がある. 本研究では, 大量のデータを高速に処理できる NoSQL データベース Cassandra を XMLDB 形式に統一する事で, この問題を解決する.

3 NoSQL データベース Cassandra

NoSQL とは, Web 上での大量の書き込みを高速に処理できるデータベースとして注目されているデータベースである. RDB に比べスケラビリティが高く, 固定されたスキーマを持たない等の特徴を持つ. Twitter や Facebook などの Web サービスを提供している企業で多く利用されている.

Cassandra とは, Apache のオープンソースプロジェクトの 1 つである. NoSQL の列指向データベースに分類される. 列指向データベースは列の検索に強く大量の行を効率的に検索できるという特徴を持

Copyright is held by the author(s).

* Nobuhiko Sugiyama, 岩手県立大学 ソフトウェア情報学部 ソフトウェア情報学科 基盤ソフトウェア学講座, Jun Sawamoto and Eji Sugino and Norihisa Segawa, 岩手県立大学 ソフトウェア情報学部 基盤ソフトウェア学講座

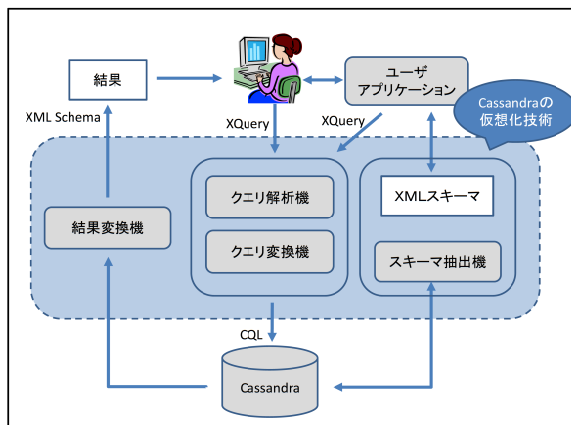


図 1. Cassandra の仮想化技術

つ．そのため，行の検索に強く列の検索に弱いRDBの相性補完となる．また，データベースを分散して利用する際に，管理ノードが存在せずノードが独立しているためシステムを安定稼働させる事ができる．これらの特徴により，Cassandra はユビキタセンサネットワーク環境に適していると考えた．

4 Cassandra の仮想化技術

4.1 システムについて

本システムの概要を図1に示す．Cassandra の仮想化技術とは，Cassandra のデータモデルや操作方法をXMLDBの形式に統一し操作を可能とするものである．本システムでは，Cassandra のクエリであるCassandra Query Language(CQL)をXMLDBのクエリであるXQueryに変換し操作する．また，データモデルの把握にはXMLDBのスキーマ構造であるXML Schemaに変換してユーザに提示する．これにより膨大なデータを扱うようなユビキタセンサネットワーク環境においても，XMLDBの知識のみでCassandraを扱える．

4.2 システム構成

本システムは，スキーマ抽出部，クエリ解析部，結果変換部から構成される．まずスキーマ抽出部では，Cassandra のデータ構造を上位の階層から取得．取得したデータをXMLスキーマ形式に整形する．ユーザは生成された共通スキーマを見ることで仮想化データベースの構造を把握する際に使われる．次にクエリ解析部では，まず受け取ったクエリに対して構文解析を行いCQLに変換する．生成されたクエリは実データベースへの問い合わせに用いられる．最後に結果変換部では，Cassandra から受け取った検索結果をXML形式に変換する．これらの機能により，ユーザはXMLDBの知識のみでCassandraを利用することができる．

5 ユーザインタフェース

ユーザが仮想化データベースを利用して検索を行う場合，まずアプリケーションにより共通スキーマを取得してデータ構造を把握する．次に，アプリケーションまたはJavaプログラミングから検索を行える．

5.1 Javaプログラミングによる検索

ユーザは仮想化データベースライブラリを用いてJavaプログラミングすることにより検索することができる．

5.2 アプリケーションによる検索

ユーザはアプリケーションを通じて仮想化データベースのスキーマを取得ができる．また，XQueryを入力することでデータを検索することができる．

6 まとめと今後の課題

本研究では，データベース利用の効率化のためにNoSQLデータベースCassandraをXMLDBのクエリであるXQueryで操作するシステムの提案・試作を行った．これによりユーザはXQueryでCassandraへのアクセスが可能となる．現在は，先行研究である異種データベースの仮想化技術との連携をさせていないため，本システムで検索できるデータベースはCassandraのみとなっている．今後は先行研究との連携が行えるよう改善していく予定である．

参考文献

- [1] 渡辺裕太, 菖蒲佳右, 和田雄次, 澤本潤, 加藤貴司, 異種データベースの仮想化技術．第8回FIT講演論文集, pp, 211-214, 2009.
- [2] XQuery <http://www.w3.org/TR/xquery/>.
- [3] Cassandra <http://cassandra.apache.org/>.