

# ストリームデータに対する空間的 OLAP システムの設計と開発

那須 勇弥<sup>†</sup> 塩川 浩昭<sup>‡</sup> 天笠 俊之<sup>‡</sup> 北川 博之<sup>‡</sup>  
 筑波大学情報学群情報科学類<sup>†</sup> 筑波大学計算科学研究センター<sup>‡</sup>

## 1. 序論

センサや IoT デバイス, ソーシャルストリームなどの技術の急激な普及に伴い, ストリームデータに対するリアルタイムな分析要求が増加している. 我々は, これまでストリームデータに対し多次元的な OLAP 分析を行う StreamOLAP [2] システムを開発してきた. しかし, ストリームデータには, GPS や移動体の位置等などの空間情報が含まれる場合が多く, 空間情報を分析次元のひとつとして扱うことが可能な空間的 StreamOLAP の必要性が高まっている.

本研究では既存の StreamOLAP の枠組みを拡張し, 空間情報を含むストリームデータに対する StreamOLAP 分析を実現するためのシステムアーキテクチャを提案する. プロトタイプシステムによる性能評価実験を通じ, 処理速度に関するボトルネックを明らかにするとともに, キャッシュを利用した効率的な OLAP 分析手法を提案する.

## 2. 関連研究

OLAP [1]は蓄積された大量のデータからデータウェアハウスを構成し, 複雑な集約分析を素早く実行する分析手法である. OLAP ではデータウェアハウス内のデータに対して多次元的なデータ分析を行うために, スキーマ間の関係を表すスタースキーマからデータキューブと呼ばれる多次元データモデルを構築する. データキューブは分析対象となる数値の集合であるメジャーとそれに関連する複数の軸で構成される. このデータキューブに対して, ユーザは集約結果の粒度を変化させる処理を対話的に行い, メジャーの集約結果を得ることができる.

StreamOLAP [2]は, ストリームデータに対してリアルタイムに OLAP 分析を行うことのできる

OLAP 分析システムである. 既存の OLAP システムは事前に蓄積したデータからデータウェアハウスを構築するため, 時々刻々と変化するデータに対してリアルタイムに OLAP 分析する機能を持たない. この問題に対し, StreamOLAP はストリームデータに対する問合せ処理を行うことのできる SPE (Stream Processing Engine)を用いたシステムアーキテクチャを提案している. StreamOLAP では, SPE がストリームデータを集約する役割を担い, SPE が集約したデータを蓄積するバッファに対して OLAP Engine が OLAP 処理を実行する. しかし, StreamOLAP は空間データのような複雑なデータ型を扱うことはできない.

Spatial OLAP[3]は, 位置情報を持つデータに対して OLAP 分析を可能とするシステムである. Spatial OLAP は分析の軸として幾何学データを含む空間次元を持ち, 幾何学データの領域内に含まれる位置情報を持つデータを空間検索することで空間内の集約処理結果を求めることができる. しかし, Spatial OLAP はストリームデータの集約を考慮していない.

## 3. 提案システム

### 3.1. システムアーキテクチャ

StreamOLAP に空間データベースを組み合わせたアーキテクチャを提案する. 図 1 にその概要を示す. まず, 入力されたストリームデータを SPE が受け取る. SPE はデータに含まれる位置情報から空間データベースへ問合せを行い, 位置情報を地名に変換する. SPE は集約処理を実行し結果をバッファに蓄積する.

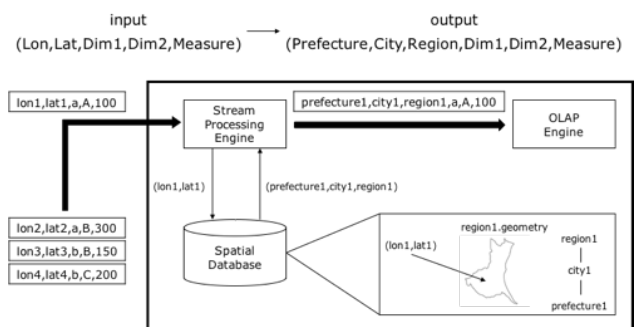


図 1 システムアーキテクチャ

Design and Development of Spatial Stream OLAP System  
 Yuya Nasu<sup>†</sup>, Hiroaki Shiokawa<sup>‡</sup>, Toshiyuki Amagasa<sup>‡</sup>,  
 Hiroyuki Kitagawa<sup>‡</sup>

<sup>†</sup>College of Information Science, University of Tsukuba

<sup>‡</sup>Center for Computational Sciences, University of Tsukuba

図1では位置情報を含む5つの属性を持つデータを入力とし、変換後の地名を含む6つの属性を持つデータを出力する流れを示している。

### 3.2. キャッシュを用いた地名変換高速化

3.1節で示したアーキテクチャではストリームデータが到着する毎に空間データベースへの問合せが生じる。問合せコストは大きく、ストリームデータのデータレートが大きい場合にスループットが低下すると考えられる。実際に予備実験を行った結果、従来の StreamOLAP と比較して 10 倍以上の処理時間がオーバーヘッドとして発生することが明らかとなった。

頻繁に位置を更新する移動体ストリームデータはその移動距離が非常に小さい。ゆえに、移動体が位置情報を更新する際に空間データベースから取得する地名は、その移動体が前回更新した際に獲得した地名と一致する可能性が高い。

そこで本稿では、この性質を考慮し、位置情報更新時の移動距離を判定基準とした地名のキャッシング手法を構築する。まず、移動体の位置情報が到着した際に、その移動体の id と位置情報、空間データベースから得た地名をキャッシュに保存する。移動体の位置情報を更新する場合、キャッシュに保存した位置情報と更新時の位置情報から移動距離を計算する。この移動距離が事前に設定した閾値以下であればキャッシュから地名を返し、超える場合は空間データベースへ問合せを行い、キャッシュを更新する。

## 4. 評価実験

本実験ではデータセットとして「2008 年東京都市圏 人の流れデータセット」[4]を用いた。本実験は 1 万人の 10 回分の移動データ（合計 10 万件）を SPE に送信し、1 件目のデータ入力開始時から最後のデータが SPE から出力されるまでの時間を計測する。空間次元として日本の行政区画を用いる。本実験では、Intel Core i5 CPU, 16GB RAM を搭載した計算機を用いた。

図2より提案手法はどの閾値でもキャッシュなしと比較して処理時間を 60%以上短縮していることがわかる。特に、閾値が 1km を超える場合は処理時間を 80%以上短縮できている。これは閾値が大きくなるほど、問合せ回数が減少するためである。

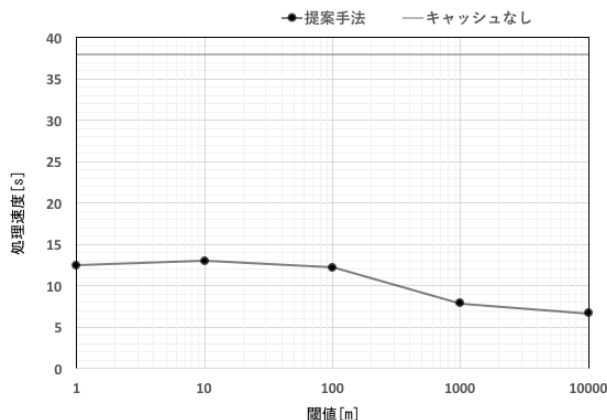


図2 処理時間の比較

## 5. 結論

本研究では空間情報を持つストリームデータの OLAP 分析を行うことのできる新たなシステムのアーキテクチャを示した。また、キャッシュを利用することによって処理速度が向上することを確認した。今後の課題として、より効率的なキャッシング手法の検討や、行政区画ではなく道路や線路を空間次元として扱う手法の検討に取り組む予定である。

### [謝辞]

本研究の一部は、理研「実社会ビッグデータ利活用のためのデータ統合・解析技術の研究開発」、ならびに、NICT 高度通信・放送研究開発委託研究「欧州との連携による公共ビッグデータの利活用基盤に関する研究開発」による。

### [参考文献]

[1] S. Chaudhuri et al, An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record, 26 (1), pages 65-74, March 1997.

[2] Kosuke Nakabasami, et al., An Architecture for Stream OLAP Exploiting SPE and OLAP Engine. In Proc. IEEE Big Data 2015, pp. 319-326, 2015

[3] J. Han, K. Koperski, and N. Stefanovic. GeoMiner: a system prototype for spatial data mining. In Proc. of the ACM SIGMOD Int. Conf. on Management of Data, pages 553-556, 1997.

[4] 人の流れプロジェクト, <http://pflow.csis.u-tokyo.ac.jp/>.