

自律分散的世界メッシュ統計基盤のアーキテクチャ設計

佐藤 彰洋[†]

横浜市立大学大学院データサイエンス研究科データサイエンス専攻[†]

1 はじめに

メッシュ統計[1]は我が国において 1960 年代に総理府統計局、国土地理院で研究が開始され、1976 年に日本工業規格（現在の日本産業規格）地域メッシュコード（JISX0410）を集計単位とする緯度経度に基づく各段に細かい大量統計として、公的統計、産業用統計において広く作成され利用がされている。

メッシュ統計は通常の統計と異なり、集計単位が各段に細かいため、人間では把握、利用が困難なほどに大量の集計単位に対して値が作成されるため、その取扱いには、専用のソフトウェアとデータインフラストラクチャが必須であるという課題がある。一方で、メッシュ統計の領域拡大と詳細化により大規模なメッシュ統計を生成、検索、再集計、可視化、計算したいという要求が近年高まっている。

異なる組織が保有する位置情報付きデータをメッシュ統計化することにより、時空間データ連結が可能となることから、地域メッシュコードはデータベースにおけるツリー構造を有する空間に埋め込みされたインデックスとして、高い検索性と分散処理の実装が可能である。例えば、Cassandra[2]など並列分散データベース上の並列計算は、MapReduce[3]アルゴリズムに基づく線形分解可能な並列計算処理に対するスケラビリティを示すが、メッシュ統計の演算処理に対しても、同様の方法が適用できる。

2 自律分散的アーキテクチャ

ビッグデータ参照アーキテクチャ[4]はクラウドコンピューティングのアーキテクチャを基に国際標準となっており、ビッグデータ利活用において、極めて有効な方式を提供している。著者は、全世界規模でメッシュ統計を利用することができる自律分散的な世界メッシュ統計基盤のアーキテクチャ設計とその実装系である MESHSTATS[5]の開発を JST 未来社会創造事業「自律分散的世界メッシュ統計基盤アーキテクチャの設計と実証」の研究として行っている。

全世界規模で生成される位置情報付きデータをメッシュ統計化し相互に接続、流通し、利用、

Web データアプリケーション開発により個別ニーズに即した配置ができる基盤技術となり得る、経済的にスケラブルなアーキテクチャを実現することにより、世界規模で組織を越えてメッシュ統計をビジネスモデルと連結して取り扱うえるシステム構成を目指して開発をすすめている。

メッシュ統計を作成するために利用可能な位置情報付きデータは大量に存在する。集計後のメッシュ統計はデータ量が元データより縮小するため、巨大な元データを保有する組織間で直接元データを交換しあうことをしなくても、相互に必要なとする計算を、メッシュ統計を介して近似的に実行することが可能である。

更に、メッシュ統計を作成する位置情報を含むデータは通常個人や組織に関する機密性の高い PII を含むため、元データを直接交換することは通常個人情報保護に関する法律のため極めて困難である。他方、メッシュ統計は位置情報に対してメッシュと呼ばれる矩形領域内で集計するため、個人を特定することができない秘匿性を有する。

結果、メッシュ統計であれば、組織横断的にデータの交換や結合分析を実行することが可能である。また、位置情報を含む元データの保有者が管理する計算資源上に元データを匿名化して配置し、データの近傍でメッシュ統計化の計算処理を行い、メッシュ統計化された処理後のデータをインターネット経由で相互に融通するというモデルにより、個人情報の保護とセキュリティ、通信量、計算量を配慮したデータ集計や、条件検索などが可能である。

複数のメッシュ統計を結合分析し、再集計することで、Web データアプリケーションを構成することで、利用者が Web 経由でデータ処理結果から直接意思決定を行うことができるアプリを大量に配置することが有効である。このためには、ユーザーごとに細かいアクセスレベル設定ができる必要がある。

そこで、MESHSTATS ではビッグデータ参照モデルに従い、計算資源層—ミドルウェア層—API・データ層—アプリケーション層の 4 層からなる

データ処理ノードを共通管理用ノードで認証し相互接続している。

データ通信はデータ処理ノードがピアとして振る舞い、相互に WebAPI を通じてデータ転送を行う P2P アーキテクチャを採用した。メッシュ統計データの格納と、検索、データ結合を分散的に実現するために、地域メッシュコード (JISX0410) を独自拡張した世界メッシュコード体系 [6] を利用している。分散格納されるメッシュ統計に位置情報から一意に生成される世界メッシュコードを全てのメッシュ統計に共通キーとして付与することで、異なるサーバーに格納されるメッシュ統計をそれぞれ住所やポリゴンなどの位置情報から検索し、世界メッシュコード [6] をキーとして結合することで並列分散処理を実現する。この処理により、分散的に蓄積されるデータの分散検索を実装し、データ量にたいして拡張性を有するシステムを構成している。

3 試作と評価

この実装系として、横浜市立大学、一般社団法人世界メッシュ研究所、総務省が保有する異なるクラウドサービスの仮想マシンにデータ処理ノードを構築し、管理ノードにより相互接続が実現した環境を構築した。MESHSTATS 上に認証機構を通じて安全な通信経路を確立し WebAPI を通じて SQL クエリを相互に並列的に送受信できる仕組みを構築し、データベースクライアントを WebAPI 経由で操作できる並列データベースを構成した。26 フィールドからなる 5,718,535 行からなる表形式データを 5 つに重複なく分割し、それぞれの MESHSTATS に格納して、並列的にクエリを実行して計算結果を統合したときの実行時間を測定した。表 1 は 10 試行を 3 つの異なる MESHSTATS で 2 種類のタスク (task1: 集計, task2: 検索) に対して計測した平均値を示す。

表 1 MESHSTATS 上での並列計算処理平均時間

	MESHSTATS1	MESHSTATS2	MESHSTATS3
Task1	1.17 s	1.21 s	1.34 s
Task2	1.47 s	1.36 s	1.63 s

更に、HTML, JavaScript, CSS を用いて作成したクライアントサイドサーバースクリプティングによるアプリケーションから API 上に格納したデータを呼び出し、メッシュ統計データをインタラクティブ分析できる多数のアプリケーションを DevOps で提供できる仕組みを構成した。

検索や統計処理、データ加工の機能を有する各種演算機能を有するプログラマブル WebAPI を

利用することにより、ユーザーは個別にアクセス制御された状況で、Web ブラウザ上からアプリケーションの開発と配置、利用を行える。データ作成とアプリケーション開発をアジャイルで実施しながらサービス提供を行うことができる。

4 まとめと今後の課題

本研究では、自律分散的な全世界規模でのメッシュ統計利活用を可能とする MESHSTATS のアーキテクチャの設計とその試作物について紹介した。MESHSTATS 上に現在存在するメッシュ統計の種類はすでに 1 万種類を超えており、100 程度のアプリケーションが試作され、ユーザーがアクセス制御された状況で利用ができるようになっている。今後の課題として、ユーザーレベルでのデータ利用量を計測することで分散的な課金を行う機構の構築と、メッシュ統計を動的に生成し利用できる仕組みの構築が挙げられる。

謝辞

本研究は科学技術振興機構 (JST) 未来社会創造事業「自律分散的世界メッシュ統計基盤アーキテクチャの設計と実証」(研究代表者: 佐藤彰洋; 2020 JPMJMI20B6; 研究期間: 2020 年 11 月～) の財政的支援を受けています。

参考文献

- [1] A. Lakshman and P. Malik, Cassandra – A Decentralized Structured Storage System, ACM SIGOPS Operating Systems Review, Vol. 44, No. 2, pp. 35–40 (2010)
- [2] MapReduce: simplified data processing on large clusters, Communications of the ACM, Vol. 51, pp. 107–113 (2008)
- [3] 佐藤彰洋, メッシュ統計, 共立出版 (2019)
- [4] ISO/IEC 20547-3:2020, Information technology – Big data reference architecture – Part 3: Reference architecture
- [5] MESHSTATS, <https://www.meshstats.xyz/meshstats/>, 最終アクセス日 2022 年 12 月 28 日
- [6] 世界メッシュコード, https://www.fttsus.org/worldgrids/ja/our_library/, 最終アクセス日 2022 年 12 月 28 日

Architecture Design of Distributed Autonomous World Grid Statistics Platform

†Aki-Hiro Sato, Department of Data Science, Graduate School of Data Science, Yokohama City University