ジオタグ付きデータをクラウドでスケーラブルに処理するジオフェンシングシステム

Martin Bauer・Dan Dobre
Nuno Santos・Mischa Schmidt

要　旨

モバイルインターネットの爆発的成長により、予測可能な応答時間内で大量のモバイルデータの処理を可能にする、スケーラブルなインフラが必要とされています。私たちは、クラウド上のジオタグ付きデータストリームに対する継続的なジオクエリをサポートする、スケーラブルなシステムを開発しました。実験の結果、開発したシステムは、サポートするジオクエリの数及びスループットの両方においてスケーラブルであることが確認できました。

キーワード

●ジオフェンシング　●M2M　●空間インデックス　●ストリーム処理

1. はじめに

急成長するモバイルインターネット、そして増え続けるM2M（Machine-to-Machine）技術の採用により、位置情報対応モバイル端末から発生する位置データの量は、かつてないほどの速さで増大しています。実際スマートシティにおいて、自動化されたM2Mサービスは、ますます増大する追跡対象オブジェクトを取り扱わなければならず、しかも、それらオブジェクトの位置情報は高い頻度で発生するかもしれません（**図1**）。

したがって、位置情報のタイムリーな取得・処理・対応には、スケーラブルなインフラサービスが必要となります。この問題に対処するため、クラウド上で位置情報ストリームに対し、継続的なクエリを行うためのジオフェンシングシステムを開発しました。これはスマートシティをプログラミングするための布石となるものです。

インフラの利用率を高めるためにM2Mインフラを共用することは、業界では一般的にみられるトレンドです。ジオフェンシングのようなイネーブラー機能は、さまざまな用途や応用分野で共用されているインフラへの実装に適していますが、それだけスループットとスケーラビリティに対する需要は更に大きくなってしまいます。

ジオフェンシングは、NECのM2Mプラットフォーム「CONNEXIVE」にとって有望な構成要素になり得るものであり、競争力を高め新しい利用領域への道を開くものです。私たちの研究は、空間データベース、大規模システム及び分散コンピューティングの知識と技術を組み合わせることで、予測可能なスケーラブル性能、適時性（タイミング）、高い可用性を実現します。私たちの知る限り、これらの特長を同時に達成しているシステムは、これまでありませんでした。

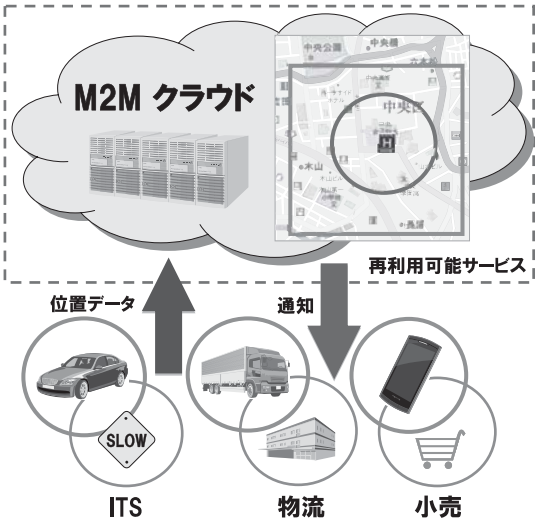

図1　ジオフェンシングの応用範囲

2. 目的

私たちは、下記の目標達成に取り組んでいます。

(1) スケーラビリティ

主な目標の1つは、ユーザーの需要に応じたスケーリングを行うことです。ジオフェンシングにおいて、スケーラビリティは、(a) ユーザーによって設定されたジオフェンスの数、(b) 単位時間当たりの位置情報更新回数として測定される負荷、の2つの次元において必要です。また、スケーラブルなシステムは、(a) ジオフェンスの数が増大しても低い一定待ち時間、(b) マシン数に比例したスループット、の両方を提供します。

(2) 予測可能な性能

予測可能な性能は、どのようなジオフェンシングシステムにとっても重要な特性であると考えます。つまり、分布が偏っている場合にも、バランスの取れたシステムが要求されるのです。特定のサーバだけに処理が集中するようなことになってはなりません。

(3) 適時性（タイミング）

位置情報は、それらについてのインデックスを作る必要なしに、即座に処理されなければなりません。ただし、位置情報の迅速な処理のためには、ジオフェンスに適切なインデックスが必要になります。

(4) 高い可用性

システムは、ノードが故障した場合でも利用可能でなければなりません。システムの総合能力は、利用不能なマシンが生じても、その割合に比例した低下をみるにとどまるべきです。

3. システムモデル

ここでのシステムモデルは、位置データを地理座標の形でクラウドバックエンドへ繰り返し送り続けるモバイルオブジェクトの集合と、そのデータの利用者として行動するクライアントの集合によって構成されます。クラウドバックエンドはこのデータを処理し、利用者に対して情報を提供します。

クライアントは、ジオフェンスを登録することでこのサービスの利用者になり、関心を持っている地理的領域内でのモバイルオブジェクト（例えば車両）の位置を追跡することができます。クラウドバックエンドは、位置データのストリームを登録されたジオフェンスと照合し、関心を持っているオブジェクトがあらかじめ指定された地理的領域に出入りするたびに、各利用者に継続的に通知を行います。

4. 既存のアプローチ

空間データの取り扱いや処理は既に目新しい話題ではなく、大規模なストリーミング・データの取り扱いも、クラウド研究ではよく知られたテーマです。しかし、私たちの知る限りでは、空間データの取り扱いや処理と、スケーラブルな分散アーキテクチャとの間に橋を架けようという試みはほとんど行われていません。これが私たちの研究の焦点です。以下では、関連する研究について簡潔に概観し、それらの研究にみられる不足点と、それらの研究がそれだけでは私たちが持つ目的には適していない理由について説明します。

4.1 空間インデックス

空間インデックスとは、空間データへのアクセスの効率化に特化したインデックス構造です。ここでは、デカルト座標（X, Y）またはGPS座標（緯度, 経度）上に1対の点で表される「点データ」と、ペア（1対）の集合によって表される任意の多角形である「領域データ」とを区別しています。

四分木、八分木、KD木などのバイナリ空間分割木構造、及び最近提案されているジオハッシュ構造は、点データには適していますが、領域データではうまく機能しません。これに対し、R木[1)]及びその派生構造は点データと領域データの両方をサポートしています。

4.2 空間データベース

空間データをサポートするデータベースはいくつか存在し、それらは通常は上記の空間構造のいずれかを持っています。

一般に普及した例としては、MongoDBとPostgreSQLがあります。MongoDBは、点データへ効率的にアクセスするためにジオハッシュを利用したNoSQLストレージシステムですが、領域データのネイティブなサポートは提供していません。一方、PostgreSQLは、領域データはサポートしていますが、1台のノードの能力を超えたスケーリングを行うためのメカニズムは提供していません。

それぞれに長所があるにもかかわらず、既存の成熟した空間データベースはどれも、スケーラブルなジオフェンシングシステムの具体的なニーズには適応していません。なお、私たちの研究と最も関連が深いものに、CANオーバレイの上に空間インデックスを構築する方法を説明した最近の論文があります[3]。

4.3 ストリーミングシステム

データ生成及びクエリ頻度の驚異的な伸びによって、通常は永続的で比較的変化しないデータを対象とする従来のデータベース技術に、多くの欠陥が露呈することになりました。これに対し、ストリーム処理システムは、極めて動的なデータを扱う大規模でリアルタイムなアプリケーションのニーズに応えるもので、永続性に関してはある程度の妥協がなされています。

例としては、Twitterの「Storm」とYahoo！の「S4」があり[4]、これらは共に、データストリームのリアルタイム処理を行うスケーラブルなフレームワークを提供しています。

ジオフェンシングとは、やってくる位置データストリームに対して継続的に空間クエリを実行すること、と表現でき、そう見れば、これはストリーミングシステムにいくらか類似する点があります。

ただし、典型的なストリーミングシステムでは、データストリームに対し単純な演算子を適用していますが、ジオフェンシングでは、従来のデータベースと同様、効率的な検索を実行するため、クエリには永続性を持たせ、かつインデックスを付加する必要があります。更に、複製と分散処理の方式を明示的に制御することは、極めて望ましいものですが、ストリーミングシステムではこれらはネイティブにサポートされていません。このためジオフェンシングシステムでは、ストリーム処理システムと従来のデータベースシステムの両方の側面を組み合わせる必要があります。

5. 研究のアプローチ

分かりやすくするため、ここでは基本的アプローチと、スケールアウトに必要なメカニズムとを区別して説明します。以下では、まず単一ノードで実行する計算について、このアプローチを説明します。続いて、データセンター向けの構成でスケーリングを行うためのシステムを、どのように設計したかを説明します。

5.1 基本的アプローチ

単一ノードのレベルでは、ジオフェンスは領域データとして扱い、空間インデックス構造（例：R木構造）を通してアクセスできるようにしておきます。位置更新は、その領域データに対するポイントクエリとして扱われます。空間インデックスは、実際のジオフェンスのおおよその近似物である最小外接矩形（Minimum Bounding Rectangle：MBR）のみを保存します。これには多くの利点がありますが、その1つとして、サイズのコンパクトさが挙げられます。これにより、インデックスをメモリ内に保管して検索速度の向上を可能にします。永続的なキーバリューストア（Key Value Store：KVS）により、ジオフェンス・データへのアクセスを提供します。

個々のポイントクエリでは、空間インデックスへのアクセスが行われ、MBRにクエリ点を含む、候補となるジオフェンスのセットが構築されます。第2段階では、KVSから実際のジオフェンスが読み出され、多角形内外判定法によって正確な結果セットが決定されます。

単純なアプローチを取るならば、結果セット内の各ジオフェンスについて、対応する利用者は「内部」イベント通知を受け取ります。同様に、結果セットに含まれていない各ジオフェンスについては、対応する利用者は「外部」イベント通知を受け取ります。その後、各利用者は、実際にジオフェンスの境界を越えたオブジェクトに対応するイベントをローカルに判定します。利用者のジオフェンスが越えられていない場合でも、各利用者に通知が送られるという点からみて、このアプローチは明らかに単純すぎます。

利用者に対してオブジェクトのジオフェンスに対する現在の相対位置を通知するのではなく、システムは「進入」と「退出」という状態遷移のみを通知するべきです。また、この目的のためには、オブジェクトの1つ前の位置に関する知識が必要となります（**図2**）。

あるオブジェクトの1つ前と現在の位置に対応する2点のポイントクエリ点をpとp'、対応する結果セットをRとR'とすると、「進入」と「退出」の遷移EとLは下記のように単純な式で計算できます。

(1) $E = R' - (R \cap R')$

(2) $L = R - (R \cap R')$

したがって、集合E（もしくはL）に対応する利用者のみが

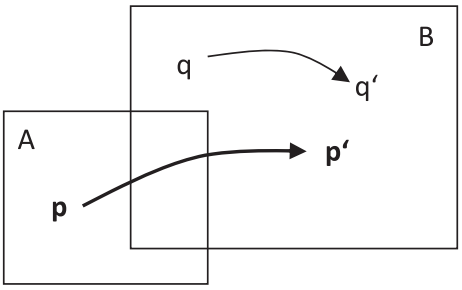

図2　利用者AとBは、p→p'の遷移のみについて通知を受ける

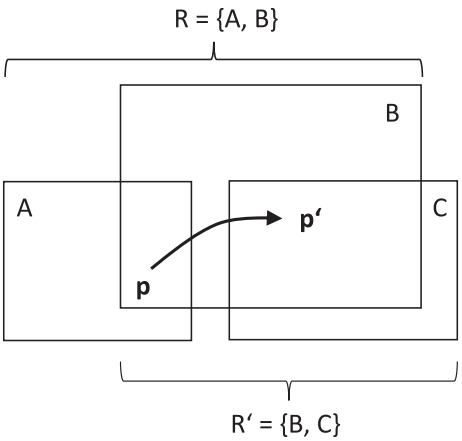

図3　式（1）及び（2）により、E＝{C}及びL＝{A}となり、利用者Bは不要な通知は受けない

「進入」（もしくは「退出」）イベントの通知を受け取ります（**図3**）。

5.2 クラウド上でのスケーリング

以降では、クラウド環境においてスケーリングを行うシステムの設計について説明します。

前述したように、ジオフェンシングシステムは、（a）ジオフェンスの数、（b）単位時間当たりで処理される位置の数として表されるスループット、の2つの次元でスケーリングを行う必要があります。

この両方を可能にするため、私たちのアプローチでは、論理的なグリッドの中にマシンを配置しています。この論理的グリッドは、読み書き記憶装置用のグリッドクォーラムシステムについての研究[2]に着想を得ています。簡潔に言うと、クォーラムシステムとは、ノードの集まりを要素として持つ集まりで、どの要素も他の要素と重複するノードを含んでいます。この重複があるおかげで、読み出し処理は、書き込みによって更新されたノードのうちの少なくとも1台とは確実に重複します。グリッドクォーラムシステムは特殊なクォーラムシステムであり、ノードの集まりは行と列を成すよう配置されています（**図4**）。データは、ある行全体に複製して配置しておき、読み出しはある列全体から行えば、読み出しは、書き込みによって更新されたノードのうちの1台からの応答を1つは受け取ることになり、よって、すべての読み出しが書き込みに気づくことが保証されます。

グリッドクォーラムシステムの利点は、行や列の数を増加することでスループットのスケーリングが容易に行えることです。

これを私たちの問題に置き換えてみると、ジオフェンスを設定した利用者は書き込み動作に対応し、オブジェクトの位置は読み出し動作に対応することになります。より具体的に説明すると、あるジオフェンスGが設定されると、Gはある行全体に広まるということです（図4）。1つの行は本質的にはレプリカ（複製されたデータ）の集合になるため、これらの行はクラスタと呼ばれます。クラスタ内の各ノードはGを記憶し、前節で説明した方法で、そのMBRをインデックスに付加します。位置情報の更新は、各クラスタ内の少なくとも1台のノードに広められます。個々のノードはセットL及びEをローカルに計算し、対応する利用者に通知を行います。Gが位置情報の更新に

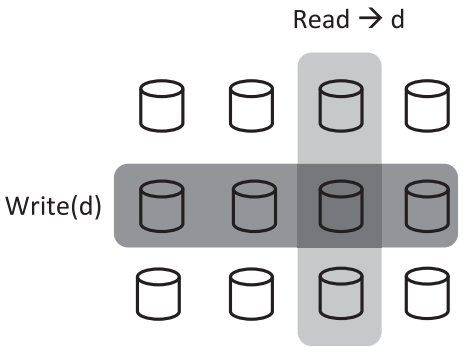

図4　グリッドクォーラムシステムでは、各データ項目が1つの行に書き込まれ、読み出しは1つの列から行われる

対してチェックされることは、クォーラムに重複があることによって保証されます（図4の色の濃い部分）。

私たちは、ジオフェンシングの作業負荷は主に読み出しによって占められると予想しています。つまり、ジオフェンスに対する操作ではなく、位置情報の更新に関連するスループットのスケーリングを行うべきだということです。また、システムは突然の負荷の上昇を吸収できなければなりません。これにはクラスタへの新規レプリカの迅速な追加が必要です。そのため、クラスタをグループメンバーシップ（Group Membership：GM）によって維持し、GMミドルウェアが提供するjoin( )及びstate_transfer( )関数を経由して新規レプリカを付加することで、能力の向上を可能にします。GMサブシステムが標準的に提供しているその他の必要な機能の例には、故障検出と複製の取り除き（不要データの削除）があります。

スループットのスケーラビリティに加えて、個々のノードに負荷を掛けすぎることなく、より大量のジオフェンスを収容する必要もあります。したがって、データセット全体を入れ替える必要なしに、新規クラスタの追加ができなければなりません。この課題に取り組むため、コンシステントハッシュ法を利用して、ジオフェンスのノードのクラスタへのマッピングを行っています。コンシステントハッシュ法は、新規クラスタが追加されると、（a）各クラスタにはほぼ同数のデータ項目が記憶されること、（b）新クラスタのみがデータ項目を受信すること、を保証します。例えば、C個のクラスタ（クラスタ1個の追加後の数）がある場合、データのうち1/Cを越えるデータ項目を新クラスタに移行する必要はありません。

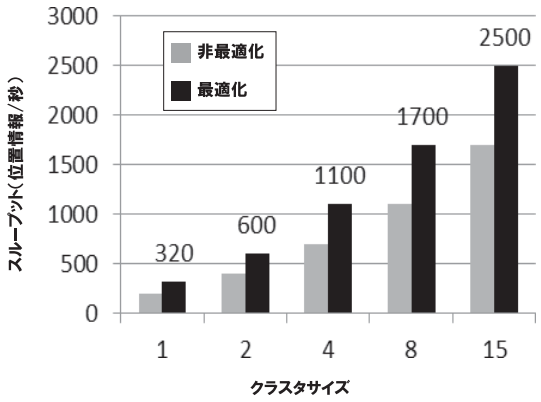

図5　スループットとクラスタサイズ（500万ジオフェンス）

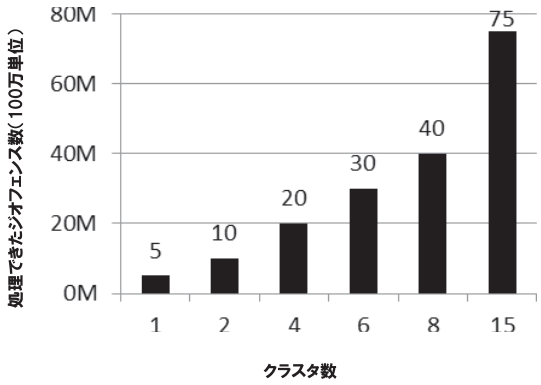

図6　短い一定待ち時間（約20ms）でのデータスケーラビリティ

5.3 評価

OpenStackの仮想化をベースにした小規模のクラウド環境で、私たちのアプローチを実験により評価しました。それぞれ8個のコアを持つデスクトップPC2台（総費用1,600ドル未満）で測定した結果、スケーラビリティが実証され、高い性能対価格比が得られました。多数のジオフェンスと照合される位置情報の数で計られるスループットは、クラスタサイズにほとんど比例して向上しており（**図5**）、一方待ち時間はジオフェンスの数が増大しても一定のままです（**図6**）。図5には、プロアクティブキャッシングを行った結果として得られたスケーラビリティ最適化も示してありますが、これについては別の機会に改めて説明する予定です。

6. むすび

クラウドでジオタグ付きデータの処理を行うためのスケーラブルなシステムを提示しました。このシステムの中核には、ノード数に比例してスケーラブルな完全疎結合アーキテクチャがあり、ボトルネックや単一障害点を回避しています。分布を常に均一に保つため、ジオフェンスは空間属性（非常に偏りのありうるデータ）に依存しない形でノードにマッピングされているので、分布状態が最悪の場合であってもシステムのバランスを保持することができます。しかしながら、このアプローチには課題もあります。それは、位置情報更新はデータセット全体に対して照合する必要があり、それがスループットに不利に働いていることです。偏りが存在する場

合でもジオフェンスの分布を維持しつつ、空間分割を利用してスループットを改善することが、今後の研究課題です。

---

*MongoDBは、10gen, Inc.の登録商標または商標です。

*PostgreSQLは、PostgreSQL Global Development Groupの登録商標または商標です。

*Twitterは、Twitter, Inc.の登録商標です。

*Yahoo! は、米国Yahoo! Inc.の登録商標または商標です。

*OpenStack は、米国 OpenStack, LLC の登録商標です。

参考文献

1) A. Guttman：“R-trees：a dynamic index structure for spatial searching,”SIGMOD '84 Proceedings of the 1984 ACM SIGMOD International Conference on Management of data,pp47-57
2) S.Y. Cheung et al.：“The grid protocol：A high performance scheme for maintaining replicated data,”ICDE 1990,pp438-445
3) J. Wang et al.：“Indexing multi-dimensional data in a cloud sys - tem,”SIGMOD '10 Proceedings of the 2010 ACM SIGMOD Inter - national Conference on Management of data, pp591-602
4) L. Neumeyer et al.：“S4：Distributed Stream Computing Plat - form,”Data Mining Workshops (ICDMW),2010 IEEE International Conference on. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5691154

執筆者プロフィール

Martin Bauer
Senior Researcher
NEC Laboratories Europe
NEC Europe Ltd.

Dan Dobre
Research Scientist
NEC Laboratories Europe
NEC Europe Ltd.

Nuno Santos
Research Associate
NEC Laboratories Europe
NEC Europe Ltd.

Mischa Schmidt
Senior Researcher
NEC Laboratories Europe
NEC Europe Ltd.

NEC 技報のご案内

NEC技報の論文をご覧いただきありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報(日本語)　　NEC Technical Journal(英語)

Vol.65 No.2　ビッグデータ活用を支える基盤技術・ソリューション特集

ビッグデータ活用を支える基盤技術・ソリューション特集によせて
ビッグデータを価値に変えるNECのITインフラ

◇ 特集論文

データ管理 / 処理基盤

超高速データ分析プラットフォーム「InfoFrame DWH Appliance」
SDN技術で通信フローを制御する「UNIVERGE PFシリーズ」
大量データをリアルタイムに処理する「InfoFrame Table Access Method」
大量データを高速に処理する「InfoFrame DataBooster」
ビッグデータの活用に最適なスケールアウト型新データベース「InfoFrame Relational Store」
高い信頼性と拡張性を実現したExpress5800/スケーラブルHAサーバ
大規模データ処理に対するOSS Hadoopの活用
大容量・高信頼グリッドストレージiStorage HSシリーズ(HYDRAstor)

データ分析基盤

ファイルサーバのデータ整理・活用を支援する「Information Assessment System」
超大規模バイオメトリック認証システムとその実現
WebSAMの分析技術と応用例～インバリアント分析の特長と適用領域～

データ収集基盤

スマートな社会を実現するM2Mとビッグデータ
微小な振動を検知する超高感度振動センサ技術開発とその応用

ビッグデータ処理を支える先進技術

多次元範囲検索を可能とするキーバリューストア「MD-HBase」
高倍率・高精細を実現する事例ベースの学習型超解像方式
ビッグデータ活用のためのテキスト分析技術
ビッグデータ時代の最先端データマイニング
ジオタグ付きデータをクラウドでスケーラブルに処理するジオフェンシングシステム
柔軟性と高性能を備えたビッグデータ・ストリーム分析プラットフォーム「Blockmon」とその使用事例

◇ 普通論文

地デジTVを活用した「まちづくりコミュニティ形成支援システム」

◇ NEC Information

NEWS

スケールアウト型新データベース「InfoFrame Relational Store」が2つの賞を受賞

Vol.65 No.2
(2012年9月)

特集TOP