

Oracle Big Data Spatial and Graph

ORACLE®
BIG DATA



Oracle Big Data Spatial and Graph は、Apache Hadoop と NoSQL データベース・テクノロジーのビッグ・データ・ワークロードをサポートする分析サービスとデータ・モデルのセットを提供します。オラクルは 10 年以上にわたって、Oracle Database 用の先進の空間およびグラフ分析テクノロジーを提供してきました。現在はこの専門性に基づいて、ソーシャル・ネットワーク・データとの連携やビッグ・データ・アーキテクチャの有効利用を進めています。

Oracle Big Data Spatial and Graph は 3 つのコンポーネントで構成されています。1 つめは、35 を超えるハイパフォーマンスの並列のインメモリ分析関数を使用した分散プロパティ・グラフです。2 つめは、物体同士がどの程度近い（または遠い）、物体が境界内（または領域内）にあるかどうかに基づいてデータを評価するため、または地球空間的マップ・データと画像を処理して視覚化するための幅広い空間分析関数およびサービスです。3 つめは、顔認識など、Apache Hadoop 内でビデオおよび画像データを処理するためのマルチメディアのフレームワークです。

分析者はグラフ機能を使用して、顧客、組織、資産の関係やつながりを発見できます。ユーザーは空間機能を使用して、位置情報を持つデータを取得し、補強し、データを調和させるために使用できます。空間サービスの適用により地球空間的データセットをクレンジング、フィルタリング、正規化、処理するだけでなく、空間の関係に基づいて結果をグループ化することもできます。

ビジネス上のおもな利点

- Hadoop 開発の複雑さの軽減と実装時間の削減
- 商用レベルの空間およびグラフ・アルゴリズムによりビッグ・データのワークロードでの深い洞察を実現
- ソーシャル・ネットワーク・データを使用し、顧客とプロスペクトの間の関係やパターンの発見に新しい次元を追加

グラフ・データの管理と分析

最近生成されるビッグ・データの多くには、収集されたデータ・オブジェクト間に固有の関係があります。たとえば、Facebook のソーシャル・ネットワーク・データや、Spotify などのオンライン音楽サービスのリスナーの音楽の好みや、eBay でのオンライン購買者の行動や、ブロガーとフォロワーやブロガー同士の関係には重要な関係とパターンが見られます。これらの関係は、グラフ内で相互接続されたオブジェクトのセットとして容易に構造化できます。グラフでは、頂点およびエッジと呼ばれるデータ構造と、関係をモデル化するための関連するプロパティや属性が使用されます。グラフによっては、道路網、電気通信、水、その他のユーティリティなど、固有の空間関係のあるものもあります。グラフごとに、モノのインターネット内でのエンティティ間のコネクション、生物学的経路やソーシャル・ネットワーク内のエンティティ間のコネクションなど、異なる関係を設定することもできます。

1.2 の新しいグラフの機能

- プロパティ・グラフ・データとパターン・マッチングの問合せのための SQL タイプの宣言型言語
- インメモリ分析のための分散機能
- Pagerank SALSA および K-Core のための追加の組込み分析

グラフのおもな機能

- 単一のエンタープライズクラスのビッグ・データ・プラットフォームでのグラフ、空間、ラスタ、およびマルチメディア・データの処理
- もっとも多く使用される 35 以上のグラフ分析関数
- 組込み分析として、グラフ・トラバースル、推奨事項、コミュニティとインフルエンサの検出、パターン・マッチングなど
- 高いパフォーマンスのための並列処理
- 自動化、カスタマイズ、分散が可能なテキスト索引付け
- グラフ・データにアクセスし、グラフの操作を行うための Java API (TinkerPop、Hadoop、NoSQL、HBase、Lucene、SolarCloud)
- Kerberos を使用する HBase 上の、また組込みセキュリティを使用する Oracle NoSQL 上のセキュアなグラフ・データベース
- Oracle Big Data Appliance で設計、テスト済み

以下の図 1 に示すように、グラフは表現しやすく、容易に直観的に視覚化できます。それだけでなく、多様なマシン・ドリブンの分析プロセスを下位の関係の発見に適用し、ビッグ・データから重要な洞察を得ることができます。

グラフとは

- 頂点とエッジ(その他、オプションで属性)のセット
- グラフとは単にリンクされたデータ

なぜ注意が必要か

- グラフはどこでも使える
 - ソーシャル・ネットワーク/ソーシャルWeb (Facebook, LinkedIn, Twitter, Baidu, Google+, ...)
 - サイバー・ネットワーク、パワー・グリッド、プロテイン・インタラクション・グラフ
 - ナレッジ・グラフ (IBM Watson, Apple SIRI, Google Knowledge Graph)
- グラフは直観的で柔軟性が高い
 - ナビゲートが容易、パスの形成が容易、視覚化が自然
 - 定義済みのスキーマが不要

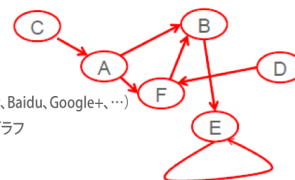


図1：シンプルなグラフのデータ・モデル

Oracle Big Data Spatial and Graph のグラフ機能

Oracle Big Data Spatial and Graph は最適なデータ・ストレージ、プロパティ・グラフの問合せと分析、共通のグラフ・モデルを備えており、データ・アクセス・レイヤーとインメモリ分析が含まれています。データベース (Oracle NoSQL および Apache HBase) の選択により、分散型の、スケーラブルでセキュアなグラフ管理が行えます。

Oracle Big Data Spatial and Graph は強力で効率的なインメモリ分析エンジンを備えています。従来、グラフ分析はグラフ内の大部分のノードに定期的に順不同で (ランダムに) アクセスするため、時間のかかる作業でした。インメモリ分析ではこのパフォーマンスの課題に対処するために、グラフ分析をメモリ内で行い、現代のシステム・アーキテクチャに固有の並列処理を適用し、必要に応じてクラスタ内の異なるノード上で動作するインメモリ分析の複数のインスタンス間で特定の分析を分散させています。

自分でコーディングするための少数のアルゴリズムとプリミティブを備えた他の製品とは異なり、インメモリ分析は通常のグラフ分析のニーズに対応する 40 個近い豊富な組込みソーシャル・ネットワーク分析 (SNA) アルゴリズムがあらかじめ統合されています。たとえば、グラフ横断、推奨事項、コミュニティとインフルエンサの検出、その他のパターン・マッチングなどがあります。メモリ内のグラフもまた、宣言型言語によりパターンの問合せをすることができます。

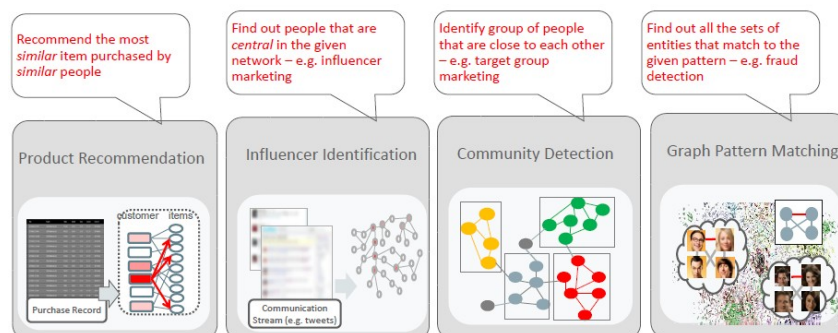


図2：共通のグラフ分析のユースケース

1.2 の新しい空間の機能

- 空間分析と処理のための Hive のサポート：開発者は SQL を使用して HDFS でのデータの分析および処理が可能
- ベクター処理のための空間結合
- 空間のクラスタリングとビンニング
- ラスター・ローダー：マルチバンド・イメージのサポート

空間のおもな機能

- 単一のエンタープライズクラスのビッグ・データ・プラットフォームでの空間およびラスター・データの処理
- 標準の空間拡充サービス
- 近接性問合せ、距離計算、バッファ生成を含む、フィルタリングと分類のための空間分析
- 管理インタフェースとマップ視覚化ツール
- 大規模な空間データ処理のサポート（ベクターとラスター）

スケーラビリティとパフォーマンスのため、インメモリ分析エンジンはデータ・アクセス・レイヤーに基づいて分散グラフ・データベースを効率的にスキャンしてフィルタリングし、解析用にメモリに読み込むノード、エッジ、プロパティを特定します。

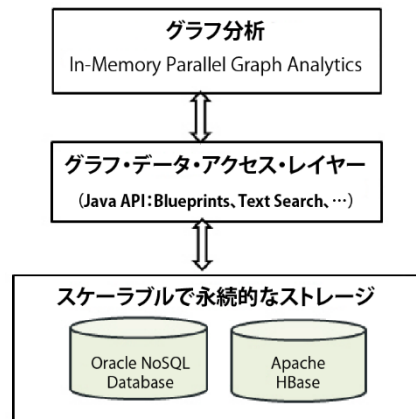


図3：オラクルのプロパティ・グラフ・アーキテクチャ

開発者は各種の Java API を使用してグラフ・データにアクセスし、グラフの操作を行うことができます。その他の機能には、パラレル・バルク・ロードの最適なサポート、プロパティ・グラフ・データのエクスポート、Apache Lucene および SolrCloud とのテキスト検索の統合があります。

空間分析およびサービス

空間データとは、地理または形状領域で検出されるものの場所、サイズ、形状に関する情報を表す（または含む）データです。空間データには公園、隣近所、住所、土地の一区画、国、都市、郡、郵便の境界、GPS 座標などが含まれます。ショッピング・モール内の店舗、オフィス・ビル内の仕事スペース、スタジアムやアリーナのレイアウトなどの屋内の場所も空間データの例です。このような情報はよくマップに表示されます。

空間データには、ベクターとラスターという 2 つの大きなカテゴリがあります。ベクター・データは 2 次元の点と線とポリゴン（それぞれ、住所、道路、境界などに対応）、または 3 次元の点とポリゴン・データによって表現され、表面、建物の輪郭、高さの次元を含むその他のオブジェクトを表わします。

ラスター・データはグリッドに格納され、各セルが 1 つの値を持つセル（またはピクセル）のセットで構成されます。これらの値では光の強さ、水分値、音の値、温度、センサーの測定結果を表現できます。ラスター製品の商的使用には、衛星画像、デジタル空中写真、農業の土壌測定、天候測定などが含まれます。

「Hadoop 環境の爆発的な拡大に伴い、空間に基づくワークロードの必要性はこの上なく高まっており、オラクルが Oracle Big Data Spatial and Graph を紹介するには最高のタイミングでした。このエキサイティングな新しいテクノロジーにより、空間処理の価値が高まり、Hadoop 環境で非常に大きなラスタ・ワークロードの処理が可能になります。もっとも厳しいデータ処理の要件にどのように対応できるのか、試すのが楽しみです」

BALL AEROSPACE
CHIEF ARCHITECT AND TECHNOLOGIST
KEITH BINGHAM

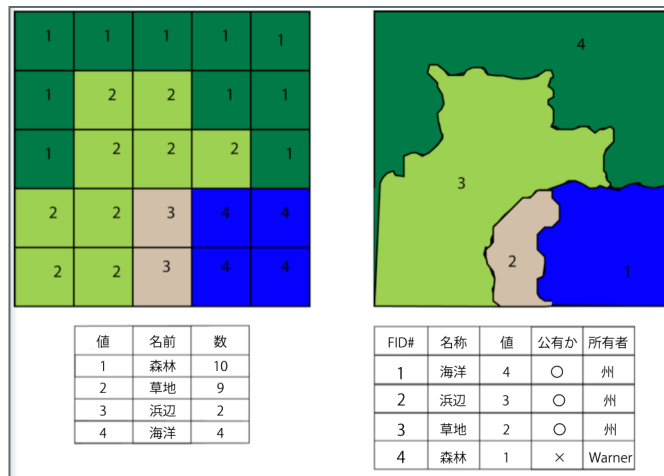


図4：ラスタ（左）データとベクター（右）データ

（コネチカット大学のコネチカット州データセンター提供）

位置情報は、テキストの住所であっても、都市や Twitter フィード内のランドマーク名であっても、または GPS の緯度/経度の座標であっても、ビッグ・データの共通要素です。企業は異なるデータセットを関連付けたりリンクしたりするための基本としてこのデータを使用できます。位置情報は、他の人、場所、オブジェクトへの近接性に基づいて、または特定領域内にそれらが存在するかどうかに基づいてエンティティを追跡し、分類するために使用することもできます。位置情報では、位置に基づく広告をサポートするジオフェンシングという技法が利用でき、近隣領域に入る顧客への販売促進が行えます。

画像およびセンサー・データを分析することにより、さまざまなビジネスの利点をサポートできます。分散センサーにより、クレンジングと準備のために大規模なジオプロセッシングが必要な生データ形式の大量のラスタ画像データが生成されます。Hadoop 環境は、これらの大量のデータを高速で、MapReduce ノード間で並列処理により格納し、処理するのに最適です。

Oracle Big Data Spatial and Graph の空間機能

開発者や Hadoop のユーザーは、Oracle Big Data Spatial and Graph の多様な機能とサービスを利用して、空間データ分析のために Hadoop データ処理システムを使用できます。

空間機能には、位置情報のデータ拡充のサポート、距離と位置による分析に基づくフィルタリングと分類、デジタル・マップ、センサーで生成される情報、衛星および空中画像などのデータセットのためのベクターおよびラスタ処理が含まれます。この製品にはマップの視覚化のための豊富な API のセットも含まれます。

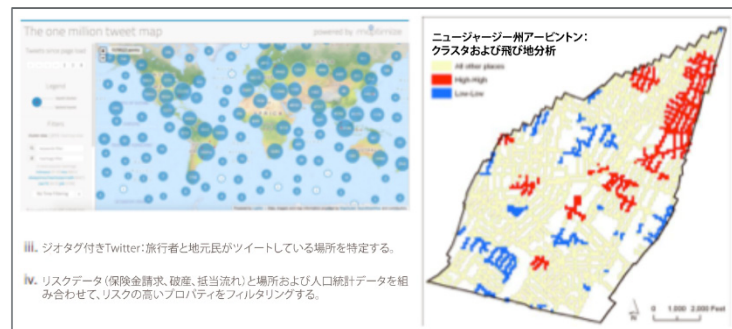


図5：位置に基づく分類、フィルタリング、集計

データ拡充、フィルタリングと分類、ベクター・データ処理のための特別なサービスには以下があります。

- 場所の名前（都市や州の名前など）や緯度/経度情報を持つデータおよびドキュメントを、実世界の場所定義や関連するデフォルトの管理境界と関連付ける能力。
- テキストベースの 2D および 3D 地球空間形式のサポート（GeoJSON ファイル、Shapefiles、GML、WKT など）。または、Geospatial Data Abstraction Library (GDAL) を使用して、Oracle SDO_Geometry、ST_Geometry、サポートされるその他の形式など、よく使用される地球空間エンコーディングを変換することもできます。
- さまざまな形式と座標系でデータを調査、分類、表示するための HTML5 ベースのマップ・クライアント API およびサンプル・コンソール。
- トポロジと距離の操作：Anyinteract、Contains（または Inside）、Distance calculation、Length calculation、Within Distance、Buffer、Point-in-polygon など。
- データの高速取得のための空間索引付け。
- 空間のクラスタリングとビニング分析。
- 2 つのデータセットのすべてのレコード間で空間の相互作用を検出するための空間結合の問合せ。
- 空間分析と処理のための Hive のサポート。開発者は SQL を使用して HDFS でのデータの分析および処理が可能です。

また、大量の空間ラスタ・データと連携するためのサービス・セットも備えています。

- GDAL によってサポートされる数十種類のイメージ・ファイル形式と HDFS に格納されたイメージ・ファイルのサポート。
- ラスタ・イメージを表示し、ラスタ・データ処理ワークフローを管理するためのサンプル Java コンソール。
- サブセット（ユーザー指定の領域に対応するカタログからイメージのセットを検索）、ジオリファレンシング（地理座標をイメージ・データに関連付け）、モザイク（ギャップとオーバーラップに対処するために入力イメージを実質的に結合）、型式変換を含むラスタ操作。

マルチメディアの機能

- HDFS と HBase でビデオおよびイメージ・データを処理し、分析するための API
- Hadoop での並列処理によるスケーラブルな高速処理
- 組み込み顔認識
- カスタム・ビデオ/イメージ処理のプラグインのためのフレームワーク

関連製品

オラクルの関連製品は次のとおりです。

- Oracle Big Data Appliance
- Oracle NoSQL Database
- Oracle Big Data Connectors
- Oracle Exadata
- Oracle Big Data Discovery
- Oracle Spatial and Graph

参考資料

Oracle Big Data Spatial and Graph について詳しくは、以下にアクセスしてください。

Oracle Technology Network
ソフトウェアのダウンロード、ドキュメント、
チュートリアル、ホワイトペーパー：
www.oracle.com/technetwork/jp/database/availability/exadata-patching-2195339-ja.pdf

Oracle.com
製品の概要、ビデオ、プレス：
www.oracle.com/technetwork/jp/database/database-technologies/bigdata-spatialandgraph

ブログ
技術的なヒント、コード例：
<https://blogs.oracle.com/bigdataspatialgraph/>

マルチメディア分析

Oracle Big Data Spatial and Graph のマルチメディア分析機能は、Apache Hadoop でビデオおよびイメージ・データを処理するためのフレームワークを備えています。このフレームワークではビデオおよびイメージ・データの分散処理が可能です。フレームワークの機能には以下があります。

- Apache Hadoop でビデオおよびイメージ・データを処理し、分析するための API
- Apache Hadoop での並列処理によるスケーラブルな高速処理
- OpenCV を使用した組み込み顔認識
- フレームワークを使用して Apache Hadoop で実行するためにカスタム・ビデオ/イメージ処理（ライセンス・プレート認識など）をインストールし、実装する能力

Oracle Big Data Appliance とその他の Hadoop プラットフォームのサポート

Oracle Big Data Spatial and Graph は、Hadoop および NoSQL 処理のためのオープンで多目的のエンジニアード・システムである Oracle Big Data Appliance のほか、サポートされる他の Hadoop および NoSQL システムに展開できます。サポートされるプラットフォームについて詳しくは、<http://www.oracle.com/technetwork/jp/database/database-technologies/bigdata-spatialandgraph/>にアクセスしてください。

まとめ

オラクルは 20 年近くも空間およびグラフ・テクノロジーの開発に携わり、もっとも要求の厳しいワークロードに対応するよう設計された新しいビッグ・データ・プラットフォームを提供してきました。Oracle Big Data Spatial and Graph を導入することで、開発者やデータ・サイエンティストはもっとも要求の厳しいグラフ、空間、ラスタ・データの処理を 1 台のエンタープライズクラスのビッグ・データ・プラットフォームで管理できます。

Oracle Big Data Spatial and Graph はビッグ・データ・プラットフォームのための新しい機会を示します。空間拡充サービスおよび分析機能とよく使用されるグラフ分析機能が標準で提供されるため、分析者と開発者は商業レベルのアルゴリズムをビッグ・データのワークロードに適用することが可能になっています。



CONNECT WITH US

blogs.oracle.com/oraclefacebook.com/oracletwitter.com/oracleoracle.com

お問い合わせ

詳しい情報については oracle.com を参照するか、+1.800.ORACLE1 でオラクルの担当者にお問い合わせください。

Hardware and Software, Engineered to Work Together

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. 本文書は情報提供のみを目的として提供されており、記載内容は予告なく変更されることがあります。本文書は一切間違いがないことを保証するものではなく、さらに、口述による明示または法律による黙示を問わず、特定の目的に対する商品性もしくは適合性についての黙示的な保証を含み、いかなる他の保証や条件も提供するものではありません。オラクルは本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクルの書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。

Oracle および Java は Oracle およびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。

Intel および Intel Xeon は Intel Corporation の商標または登録商標です。すべての SPARC 商標はライセンスに基づいて使用される SPARC International, Inc. の商標または登録商標です。AMD、Opteron、AMD ロゴおよび AMD Opteron ロゴは、Advanced Micro Devices の商標または登録商標です。UNIX は、The Open Group の登録商標です。0115



Oracle is committed to developing practices and products that help protect the environment