



空間ビッグデータプラットフォーム: 購入前に 知っておくべきこと

空間ビッグデータ



最適なデータベースを選択することは、成功へのキーとなります。

最初から正しい決定を下すことは非常に重要であり、製品構築後の変更は高コストとなり、時には変更自体が不可能なこともあります。

ビッグデータには、次の3つの大きな課題があります。

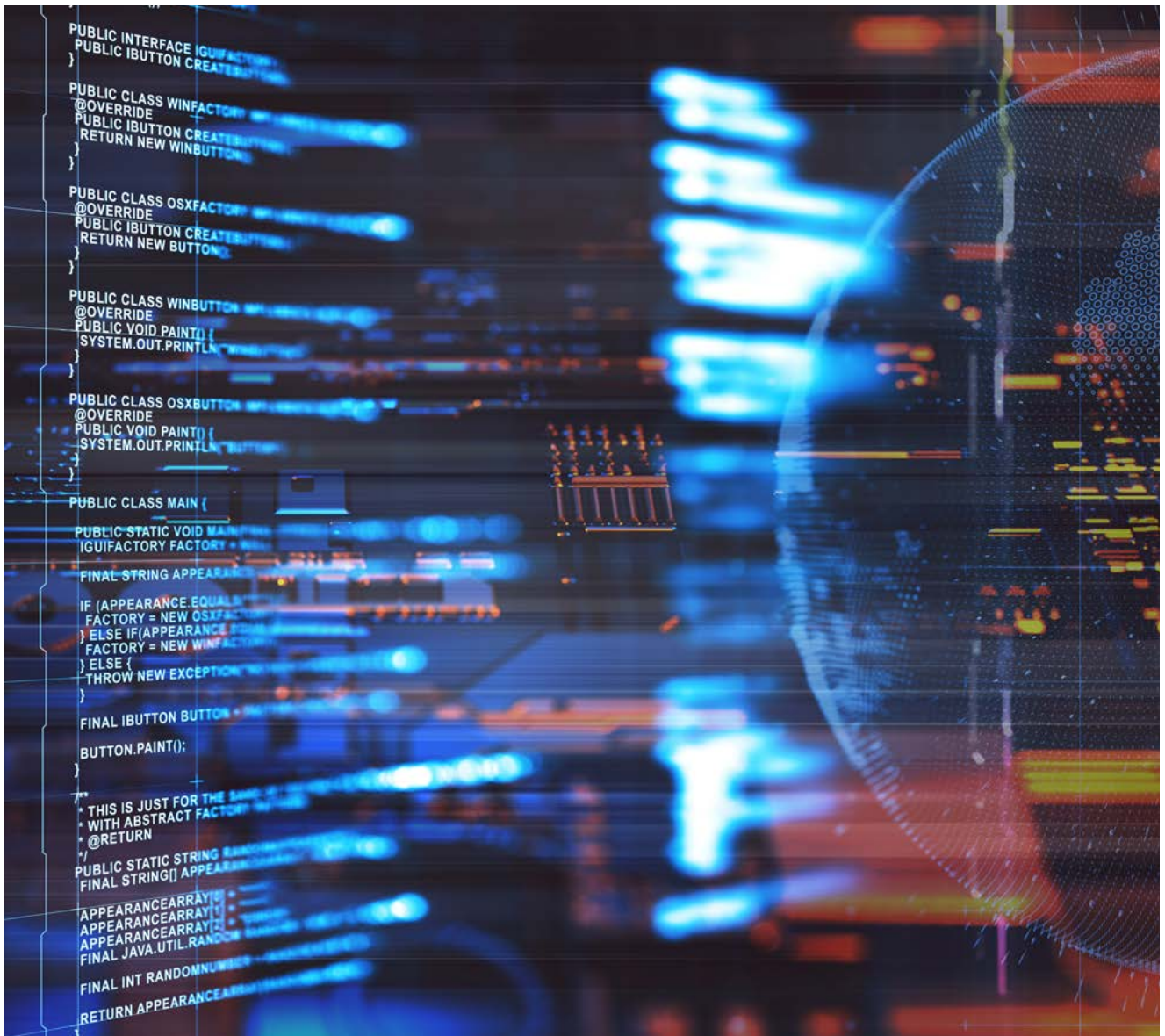
スケール: システムに継続的に、より多くのデータを追加し続ける方法

スループット: 急速に変化するデータの取り扱い方法

応答性: システムをリアルタイムで実行させる方法

現在のデータベースシステムは、過去数十年の問題を解決するために作られたものです。そのため、問題解決のためにパフォーマンスを最適化した一方で、同じくらい重要な課題には取り組まれていませんでした。

今日、マシンは現在の想像を超える速度とスケールでデータを生成しており、データベースは故障し始めています。今こそ、新しいタイプのデータベースが必要なのです。



GeoSpockのソリューション

当社のソリューションは、使いやすくデザインされたNoSQLデータベースであり、膨大なスケールにおいて一貫して高い性能を維持することが可能です。GeoSpockでは他の技術においても、キーバリューストア型と共にワイドカラム型データベースも使用しており、また、十分な柔軟性を持ってユーザーにドキュメントストレージシステムを公開できます。

このシステムは、高パフォーマンスを維持しながら膨大なスケールのデータセットを処理することを目的とした、効率的な多次元データベースです。インフォメーションストアにアクセスすれば、その比類ないデザインにより、2Dおよび3Dの地理空間情報をも超えてしまいます。さらに、K近傍探索など計算量の多い多次元クエリでない限り、非常に効率的に処理することが可能です。

当社の現在のソリューションは、データベースとしてのサービスをクラウドで提供しています。自動負荷分散と同様にデータの動的なシャーディングも滞りなく処理でき、使用量が多い場合でもシステムの応答性を維持し、容易に起動実行することができます。

GeoSpockは、データインセンティブアプリケーション、および、データテック、ロジスティクス、IoTやテレコミュニケーションアプリケーションなどのリアルタイムシステムに最適です。

最適化されたパフォーマンス

より多くのデータやマシンが既存のNoSQLソリューションに追加されるのに比例して、通常、クエリにも時間がかかるようになります。しかしGeoSpockでは、クラスタ内のマシン数が数千を超えていても一貫したリアルタイムの応答時間を維持できる新しい技術を開発しました。

- 「フォン・ノイマン・ボトルネック」: メモリバスがCPUの必要とするデータ量に対応できないために、パフォーマンスが限界に達し、非常に強力なCPUでさえも飢餓状態になることがよくあります。

当社は、データへのアクセスをより効率的にする分散インデックスメカニズムを使用して、情報へのアクセス方法を改善することでこれを解決しました。

- 他のNoSQLおよびSQLのソリューションでは、パフォーマンスのボトルネックが発生してしまいます。これは、信頼性の高い読み取りおよびレポートが可能になるまで、十分なデータを保持する読み取り/書き込みロックまたは反復可能読み取りロックがインデックス作成メカニズムに必要なためです。

当社のソリューションでは、これらの読み取り/書き込みロックを回避する分散インデックスメカニズムを使用して、



この問題を解決しています。これにより、非常に高い書き込みスループットを実現し、クエリの柔軟性を維持しながら隠れたコストを回避して、**読み取りと書き込み両方を同時に実行することができます**。これはIoT環境に不可欠と言えます。

- パフォーマンスと迅速なクエリ応答を求める際、多くのソリューションは、クライアントがデータを要求することを予想したクエリとレポートの事前集約結果の技術を活用します。データの预处理と集約は、スケールとクエリ応答の両方の予想からソリューションの欠点を克服するために必要であるため、システムの柔軟性とインサイトへの速度を低下させてしまいます。さらに、この種のシステムで実行される質問とクエリは、ビジネスニーズの変化によって唯一無二となるため、回答の集計とクエリの事前計算を絶えずやり直す必要があり、関連するコストとダウンサービスに影響を与えてしまいます。

GeoSpockではすべてのデータを元の状態で保存し、取り込んだすべてのデータにアクセスできるようにして、スケールに関係なく、業界リード標準の応答時間を維持します。**当社のソリューションの回答は事前計算を必要とせず、またシステムに保存する必要もありません**。データベースで要求されたとおりに回答され、必要に応じて変更することも可能です。

- 考慮すべきもう1つの重要な要素は、ソリューションをホストするプラットフォームのコストです。従来のデータベースおよびCPUベースのソリューションなどのより新しいアプローチでは、パフォーマンス要件に対処できるよう、高価で特別に設計されたハードウェアが必要です。新しいNoSQLシステムは、安価なシステムのファームを使用してこれに対処しようとしていますが、たとえ安価な低レベルのハードウェアシステムまたはクラウド化したとしても、必要なパフォーマンス要件を満たすには、ファームの使用に膨大な数のサーバーが必要となり、最終的には全体の運用コストが最初に見積もられたものよりも大幅に増加してしまいます。

当社では、必要に応じてクラスターのサイズを縮小拡大することにより、この問題に対する賢明な解決策を見つけました。そういった意味で、CPUを最も消費するアクティビティは、最初の情報の取り込みに関連付けられています。ここでは、膨大な量の履歴情報が顧客から取得され、内部データベースシステムにインデックス付けられ保存されます。この最初の取り込みが完了すると、その後の少量の情報の取り込みに必要な計算能力は、最初の取り込みよりもはるかに低くて済みます。そのため、**ほとんどのクラスターの使用を停止でき、システムをサポートするハードウェアの実行コストは、他の同等のプラットフォームよりもはるかに低くなります**。

データベース比較表

| データベース | GEOSPOCK | RDBMS(関係データベース管理)システム(従来のデータベース) | キーバリュー型 | ワイドコラム型 | ドキュメント指向 |
|------------------|--|---|--|--|---|
| 必要なハードウェア | <ul style="list-style-type: none"> 購入可能 安価 成長し易い | <ul style="list-style-type: none"> 購入不可能 高価 成長が困難 | <ul style="list-style-type: none"> 購入可能 安価 | <ul style="list-style-type: none"> 購入可能 安価 | <ul style="list-style-type: none"> 購入可能 安価 |
| OLAP | <ul style="list-style-type: none"> 対応 | <ul style="list-style-type: none"> 非対応(制限あり) | <ul style="list-style-type: none"> 対応 | <ul style="list-style-type: none"> 対応 | <ul style="list-style-type: none"> 非対応(この用途には不向き) |
| OLTP | <ul style="list-style-type: none"> 非対応 | <ul style="list-style-type: none"> 対応 | <ul style="list-style-type: none"> 非対応(バッチ処理の対応) | <ul style="list-style-type: none"> 非対応 | <ul style="list-style-type: none"> 対応(機能に制限あり) |
| データ量 | <ul style="list-style-type: none"> 多量 - 数100TB程度 原則無制限 | <ul style="list-style-type: none"> 最大数TBに制限あり 成長を確実なものとするため、追加の外部レポジットが必要 | <ul style="list-style-type: none"> 数100TB程度のスケールアウトが容易 | <ul style="list-style-type: none"> 多量 - 数100TB程度 | <ul style="list-style-type: none"> 多量 - 数100TB程度(パフォーマンスの低下が見られるため非推奨) |
| 応答時間 | <ul style="list-style-type: none"> コンスタント 多くの場合、BigQueryより速い | <ul style="list-style-type: none"> システムが大きくなるにつれ遅延 クエリの複雑さと分析量により、通常数分から数時間 | <ul style="list-style-type: none"> 1つのディメンションに限定されていると非常に速い(一意のインデックス)が、その他の場合はフルスキャンが必須となり、パフォーマンスは非常に低くなる | <ul style="list-style-type: none"> クエリのサイズまたは複雑さが増すにつれ、(分、時間、日数単位で)急激に増加する可能性あり | <ul style="list-style-type: none"> インデックスをメモリに保持できる場合は高速(DBの有効なボリュームを削減)だが、その他の場合は非常に遅い可能性あり |

| データベース | GEOSPOCK | RDBMS(関係データベース管理)システム(従来のデータベース) | キーバリュー型 | ワイドコラム型 | ドキュメント指向 |
|----------------------------------|---|--|--|---|--|
| クエリの豊富さ/機能性 | <ul style="list-style-type: none"> SQLをサポート データの取り込みとロードマップの結果用のAPI | <ul style="list-style-type: none"> 非常に完全かつ柔軟 より高い柔軟性のため、通常はSQLと他の言語のAPI | <ul style="list-style-type: none"> 非常に不十分な機能 単純なクエリのみ | <ul style="list-style-type: none"> 豊富な機能性 実行できるクエリにおいて非常に柔軟 | <ul style="list-style-type: none"> 関連性のある単純なクエリのみ 他のストレージ量に柔軟性あり |
| 管理のしやすさ | <ul style="list-style-type: none"> 完全管理 サービスロードマップのSaas | <ul style="list-style-type: none"> パフォーマンスと定期的なインストールを維持するために、通常は複雑な管理が必要(クラウド展開でも同様) | <ul style="list-style-type: none"> 中 - インデックスの一意性の維持が複雑な管理につながる可能性があるため その他のクエリは追加のインデックスが必須であり、管理はより複雑となる | <ul style="list-style-type: none"> 非常に複雑な管理 - 特に情報量が増加し、クエリの複雑性が増した場合 | <ul style="list-style-type: none"> 追加のインデックスやインフラストラクチャーの維持のためにクエリがより複雑になる場合が課題 |
| クラウド展開 | <ul style="list-style-type: none"> ネイティブ | <ul style="list-style-type: none"> ネイティブではない(定期的なソフトウェアインストールでは、通常可能) | <ul style="list-style-type: none"> 多くの製品でネイティブ | <ul style="list-style-type: none"> 対応 - クラウド開発に簡単に適合 | <ul style="list-style-type: none"> 多くの場合ネイティブ |
| オンプレミス展開 | <ul style="list-style-type: none"> ロードマップ | <ul style="list-style-type: none"> 対応(多くの場合) | <ul style="list-style-type: none"> 可能/複数の製品においてネイティブ | <ul style="list-style-type: none"> 中程度 オンプレミスに最適だが、コストや複雑性が大幅に増加する可能性あり | <ul style="list-style-type: none"> ネイティブ/多くの場合容易 |
| 地理的サポート | <ul style="list-style-type: none"> 対応(ネイティブサポート) | <ul style="list-style-type: none"> 対応(ネイティブではなく、多くの場合限定されたサポート) GISシステムでさえ、RDBMSを出発点として開発されている。つまり、ジオサポートは後の段階で追加される。 | <ul style="list-style-type: none"> 明示なし | <ul style="list-style-type: none"> ネイティブではない/追加困難 | <ul style="list-style-type: none"> ネイティブではない |
| テンポラルのサポート(ディメンションとしての時間) | <ul style="list-style-type: none"> 対応(ネイティブサポート) | <ul style="list-style-type: none"> 対応(ネイティブではない・限定サポート) ほとんどのシステムが、他のデータと同様に時間を保存できるが、このフィールドに基づいてクエリを高速化したり、時系列やテンポラルシリーズの比較などに沿って情報の進化を示したりするために多次的に処理されない | <ul style="list-style-type: none"> 明示なし | <ul style="list-style-type: none"> ネイティブではない(他の情報のフィールド) | <ul style="list-style-type: none"> ネイティブではなく、他のコラム/ディメンション |

GeoSpockとは

GeoSpockは、複雑な繋がりを、より大きな全体像が明らかな有意義なビジュアライゼーションへと変換することで、センサーデータと生活をつないでいます。その最先端の巨大空間データプラットフォームには、交通渋滞を緩和することで排気ガスを削減したり、流通業務を最適化することにより収益性を最大限にしたりと、生活やビジネスを革新する力があります。GeoSpockは、ほんの数秒だけで何兆ものデータを活用し、海上物流からスマートシティとデータ技術まであらゆる市場において、隠されたパターンを見つけだし、価値のある新しい観点を作り出しています。

www.geospock.com



CAMBRIDGE | LONDON | SINGAPORE | TOKYO