

Oracle Big Data SQL

リリース4.1



BIG DATA

IoTや世界的な顧客ベースのソーシャル・ストリームなど、企業が有効に活用できるデータの未曾有の増加により、ビジネスの絶大なチャンスが生まれています。しかし、ビッグ・データの大きな可能性とともに、複雑さも大幅に増大する可能性もあります。新しい膨大なデータを利用するためのビッグ・データ・システムを既存の情報資産に統合することは困難を伴うことがあります。貴重なデータが、ビジネスに不可欠な多くのオペレーションが行われる場所とは異なるシステムに格納されている場合もあります。さらに、このデータにアクセスするには、多額の資金を投じて分析とレポートのためのコードを作成しなおす必要が生じることがあるため、なかなかデータにアクセスできないうえに、ビジネスにもたらされるデータの最終的な価値も低下するおそれがあります。

Oracle Big Data SQLを使用すると、既存のSQLスキル、セキュリティ・ポリシー、アプリケーションを利用して、Apache Hadoop、Apache Kafka、NoSQL、オブジェクト・ストア、Oracle Databaseにまたがるデータを並外れたパフォーマンスで即座に分析できます。データ・サイエンス業務の簡素化からデータ・レイクの解放にいたるまで、Big Data SQLは最大限のエンド・ユーザーがビッグ・データの利点を享受できるようにします。

おもな機能

- Oracle Database、Hadoop、オブジェクト・ストア、Kafka、NoSQLソース全体でシームレスにデータの間合せを実行
- 変更なしですべてのOracle SQL間合せを実行アプリケーションの投資を維持
- Hadoop、Kafka、NoSQL、オブジェクト・ストアでSmart Scanを使用して、ファンアウトの並列化によってデータを処理することで、スケーラビリティとパフォーマンスを向上
- 主要なHadoopディストリビューションにOracle Database 19cからアクセス可能
- Oracle Database、Hadoop、オブジェクト・ストア、Kafka、NoSQLに分散する機密データへのアクセスをOracle Database Securityの機能で一元的に制御
- Copy to Hadoopを使用してOracle DatabaseからHadoopに容易にデータをコピー

すべてのデータでの高度なSQL処理

Oracle Big Data SQLはオラクルのデータ仮想化を刷新するものです。異種のデータセットに対するSQLや他のデータAPI（REST、Node.jsなど）用の新しいアーキテクチャおよびソリューションであり、Apache Hadoop、Apache Kafka、オブジェクト・ストア、および多数のNoSQLデータベースのデータをOracle Databaseに格納されたデータとシームレスに統合します。Oracle Big Data SQLを使用すると、以下のことができます。

- **Oracle SQLを使用して**、Apache Hadoop、オブジェクト・ストア、Apache Kafka、およびNoSQLの**データの間合せと分析**を行う。
- 分散アーキテクチャでSmart Scan、集計オフロード、パーティション・ブルーニング、ストレージ索引、Bloomフィルタ、条件プッシュ・ダウンなどの高度な技法を利用して、あらゆるデータに対する**間合せパフォーマンスを最大化**する。
- **ビッグ・データ分析**を既存のアプリケーションおよびアーキテクチャに**統合**する。
- Oracle Databaseの**セキュリティ**とアクセス・ポリシーをApache Hadoop、オブジェクト・ストア、Apache Kafka、およびNoSQLのデータに**拡張**する。



おもな利点

- Hadoop、オブジェクト・ストア、Kafka、NoSQL、Oracle Databaseにまたがるデータセットを透過的に分析
- データのローカル処理を利用することにより、問合せの高速化を達成
- ユーザーの既存のSQLスキルを使用してビッグ・データのソース全体でデータを分析
- 現在のSQLベースのアプリケーションでシームレスに新しいデータを統合可能
- 比較的低コストのHadoopストレージとオブジェクト・ストレージが活用されるように、情報ライフサイクル管理戦略をシームレスに拡張
- Oracle Databaseのセキュリティ・ポリシーを使用してあらゆる機密データを保護

強化された外部表

異なるシステムに格納された大規模なデータセットを扱うとき、データの構造を知ることはもちろん、データの格納場所を知ることも難しい場合があります。Big Data SQLでOracle Database 19cのビッグ・データ対応外部表を使用すると、Oracle Databaseが、Hadoop、オブジェクト・ストア、Kafka、およびNoSQLシステムのデータをカタログに追加し、保護するための単一の場所となります。Big Data SQLでは、クラスターとクラスター内の表の両方について、データの移動またはコピーを行わなくても、外部データソースに関するメタデータが追跡されます。Big Data SQLの外部表は以下に対応します。

- Oracle Databaseのデータを、Hadoop、オブジェクト・ストア、Kafka、およびNoSQLデータベースのデータと結合する**シームレスなメタデータ統合と問合せ**
- HCatalog（またはHive Metastore）に格納されたメタデータからOracleの表への**自動マッピング**
- 1つのOracle Databaseから複数のHadoopクラスターへの問合せを可能にする**複数クラスターのサポート**
- データベース管理者が列のマッピングとデータ・アクセスを柔軟に制御できるようにするための**強化されたアクセス・パラメータ**

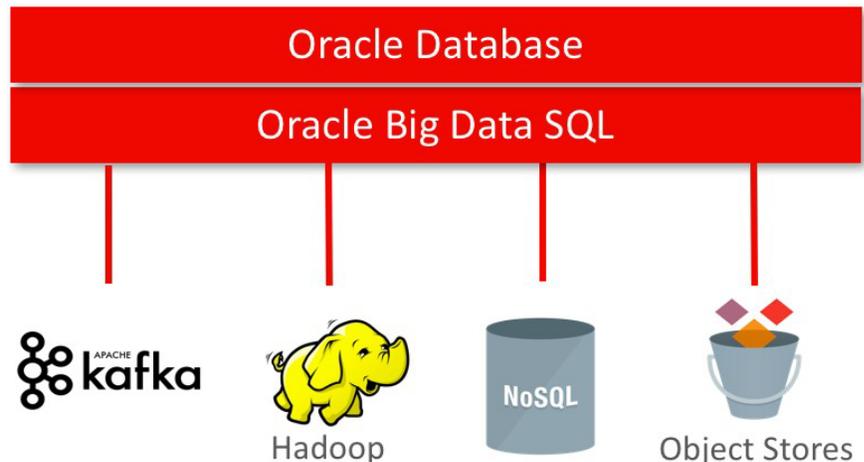


図1：Oracle Big Data SQLを使用すると、Oracle Database、Apache Hadoop、Apache Kafka、選択されたNoSQLデータ・ストア、およびオブジェクト・ストアでOracle SQLの問合せを実行できます

Smart Scan：データ主導の平行処理

ビッグ・データから洞察を得ることは、非常に大量のデータを選別することを意味する可能性があります。ビッグ・データがもたらすデータ量の激増により、優れた分析パフォーマンスを実現するには、データを分析に回すのではなく分析機能をデータ側に移動する以外にありません。Big Data SQLは、オラクルのクラス最高のExadata Database Machineで初めて導入されたSmart Scanの機能を、ビッグ・データ・ストアに適用します。Smart Scanを使用すれば、Oracle SQLの操作をビッグ・データ・システムのストレージ層にプッシュ・ダウンできます。

Smart Scanはこれらのストレージ・システムの水平スケーラビリティを備えているだけでなく、最大のデータセットの平行処理も自動的に行えます。以下のことが可能です。

- 問合せに関連する行と列のみがOracle Databaseに送信されるように行う、**ローカルでのデータのフィルタリング**

- Oracle Databaseのデータと膨大な外部データの結合を高速化する、Bloomフィルタとキー・ベクトルによる**結合の最適化**
- Hadoopクラスタの演算能力を利用してデータをローカルに集計し、要約されたデータをOracle Databaseに戻す**分散集計**
- データ・マイニング・モデルの**スコアリング**、およびJSONやXMLなどで記述されたドキュメント・データセットの問合せ処理の強化
- **Oracleネイティブ演算子**により、Big Data SQLを使用した問合せとOracle Database単体の問合せの間での完全な正確性を実現

ストレージの索引付け：より効果的なI/O

Oracle Big Data SQLは、一連のSmart Scan機能のほか、I/O発生前に処理の高速化を図るストレージ索引テクノロジーも備えています。データにアクセスすると、関連データの格納場所を示すローカルのメモリ内索引がOracle Big Data SQLによって自動的に作成されます。それ以降に同じデータの問合せを行うと、ストレージ索引テクノロジーにより、その問合せに関連しないデータ・ブロックが読み取られないよう保証されます。

ビッグ・データ・システム内のデータ・ブロックは非常に大きくなる（最大で数百メガバイト）可能性があるため、この"I/Oスキップ"戦略により、問合せによってはパフォーマンスが大幅に向上することがあります。

条件プッシュ・ダウン：外部ストレージ・システムの利用

Oracle Big Data SQLでは、HadoopおよびNoSQLソースのデータを容易に統合できるだけでなく、最大のパフォーマンスを得るために基本のストレージ・メカニズムを利用することもできます。Big Data SQLの条件プッシュ・ダウン・テクノロジーを使用すると、Oracle Databaseで発行される問合せの条件をリモート・システムで実行し、特定のファイル形式にプッシュすることができます。条件プッシュ・ダウンを使用することにより、Big Data SQLでは以下を行えます。

- Apache Hiveによって管理される表の**パーティションのプルーニング**
- Apache Parquet形式およびApache ORC形式で格納されたファイルの**I/Oの最小化**
- Oracle NoSQL DatabaseまたはApache HBaseに格納されたデータの**リモート読取りの有効化**

分散集計：サマリーの問合せの高速化

Oracle Big Data SQLは、オラクルのインメモリ・テクノロジーを活用してSQL集計をOracle Big Data SQLセルにプッシュします。そのため、Oracle Big Data SQLではHadoopクラスタの処理能力を利用して、クラスタ・ノード全体に集計を分散させることができます。これにより、多くの場合10倍を超えるほど、大幅にパフォーマンスが向上する可能性があります。

分散集計は、単一表と複数表の結合の両方に関する問合せに適用されます。一般的な"スター"および"スノーflake"クエリは通常、より小さいディメンション表をより大きいファクト表に結合します。Oracle Databaseのインメモリ集計は、Smart Scanの一部に効率的なキー・ベクトル処理を使用し、スキャン操作の一環としてデータを集計するため、対象となる問合せの処理時間が大幅に短縮されます。

問合せのストリーム：KafkaトピックへのSQLアクセス

Apache Kafkaは、スケーラブルでフォルト・トレラントの分散型メッセージ・システムです。組織は、リアルタイムのデータ・ストリームを配信する中央ハブとしてKafkaを利用します。相互に直接通信するシステムとは異なり、アプリケーションはKafkaトピックにメッセージを公開し、そのメッセージは他のアプリケーションによって取り込まれます。

Big Data SQLでは、Kafkaトピックへの直接アクセスがサポートされるため、SQL問合せで、ほぼリアルタイムのイベントと、Oracle Databaseやビッグ・データ・ストアのデータとを組み合わせることができます。

Query ServerによるHadoopデータへのSQLアクセスの簡素化

Oracle Big Data SQL Query Serverを使用すると、アプリケーションは、別のOracle Databaseがなくても、Hadoop内のデータに問合せを実行できます。Query Serverは、お使いのHadoopクラスタのエッジ・ノードに自動的にインストールされ、構成されるOracle Database SQLエンジンです。メンテナンスが不要で、メタデータと認可ルールはHiveメタストアとHDFSから継承されます。Big Data SQLの外部表は、選択したHiveデータベースと自動的に同期されます。データの認可では、Apache SentryとHDFSのアクセス制御が使用されます。

Query Serverは、HadoopでSQLを処理することを目的としており、Oracle Databaseと統合されたBig Data SQLデプロイメントを強化します。すべての問合せはHadoopクラスタに保存されているデータを対象とするため、データは永続化されません。Oracle Databaseのデータと外部ソースを結合する問合せを実行する場合は、Oracle DatabaseとBig Data SQLの組み合わせが適切な代替策です。Query Serverは、すべてのデータがHadoopに常駐しており、機能が豊富なOracle DatabaseのSQL言語と問合せ実行機能を活用したい場合に最適なソリューションです。

Hadoopへの情報ライフサイクル管理の拡張

Oracle Databaseでは、何年も前から情報ライフサイクル管理 (ILM) のサポートを充実させてきました。データの階層化に使用できる機能が多数あり、アクセス要件とストレージ・コストに基づいて各種の媒体にデータを格納できます。階層は、上位から下位に向かって、1) リアルタイム・データ分析に適したメモリ内、2) 頻繁にアクセスされるデータに適したデータベース・フラッシュ、3) 業務系データの問合せに適したデータベース記憶域やExadataセル、4) アクセス頻度が低い生データやアーカイブ・データに適したHadoopやオブジェクト・ストア、という構成になります。

Copy to Hadoop

Oracle DatabaseからHadoopへのコピーは複雑になる可能性があります。Oracle Big Data SQLには、Oracle Copy to Hadoopユーティリティが用意されています。このユーティリティを使用すると、Hadoop Distributed File System (HDFS) へのオラクルのデータのコピーを簡単に行えます。Copy to HadoopによってHadoopクラスタにコピーされたデータはOracle Data Pump形式で格納されます。この形式はBig Data SQLによる問合せを最適化するもので、1) データがオラクルのデータ型として格納されるため、データ型の変換が不要であり、2) データの問合せが直接行われるため、Java SerDesに関連するオーバーヘッドが不要となります。最適化された入力フォーマット・クラスを使用することで、この同じOracle Data Pumpのエクスポート・ファイルにHiveなどのHadoopネイティブ・ツールで簡単にアクセスできます。

ハイブリッド・パーティション表

Oracle Partitioningは、1つの表のデータ・パーティションをさまざまな層に格納できるようにするテクノロジーです。表内の不変のアーカイブ・データを、さまざまなファイル形式 (Apache Parquet、Apache ORC、Apache Avro、テキストなど) を使用してHadoopやオブジェクト・ストアに常駐させることができます。このデータはオープン形式のため、Oracle Databaseの処理のために予約されることはありません。データは、お使いのデータ・レイク内で他のアプリケーションと共有される場合があります。データベースの問合せは、ほかのデータへのアクセスと同様に、このアーカイブ・データにシームレスにアクセスします。

Big Data SQLのSmart Scan機能を併せて使用すると、パフォーマンス上のメリットを複合化できます。Big Data SQLのSmart Scanでは、Hadoopクラスタの大規模・パラレル・プロセス機能を利用してソース側でデータをフィルタするため、クラスタとデータベースの間のデータ移動量とネットワーク通信量が大幅に減少します。

ビッグ・データでのOracle Database Security

Oracle Big Data SQL独自のデータ統合アプローチにより、アプリケーションは基盤となるデータ認可ルールを（つまり、HDFSのファイルについてはアクセス権限を、HiveメタデータについてはApache Sentryポリシーを）自動的に活用して、高度なOracle Database Securityのポリシーの上層に重ねることができます。このアプローチを通じて、セキュアな実装が簡素化されると同時に、基盤のストアでは利用できなかったオラクルのセキュリティ機能を利用できるようになります。オラクルのセキュリティ・メカニズムを使用すると、以下に基づいてビッグ・データを保護することができます。

- データへのアクセスを管理するための標準のOracle Databaseのロールと権限
- 権限のないユーザーが機密情報にアクセスした場合に伏せ字で表示されるようにするためのデータ編集
- 管理ポリシーを徹底するための仮想プライベート・データベース
- Oracle Database Vaultによる、特権アカウントからの機密データ保護
- Oracle Database Security Assessment Toolによる、基盤のストア内の潜在的な機密データと総合的なデータベース・セキュリティ・ステータスの特定

一連のビッグ・データの展開のサポート

Oracle Big Data SQLは幅広い展開オプションおよびプラットフォームに対応するよう設計されています。Big Data SQLには、1) Enterprise Linuxが動作するOracle Database、2) ClouderaおよびHortonworksの主要なApache Hadoopディストリビューションが必要です。Big Data SQLをOracleエンジニアド・システムと一緒に使用すると、最高のパフォーマンスを得ることができます。Big Data SQLは、Oracle ExadataとOracle Big Data Applianceの機能を最大限に活用してクラス最高のビッグ・データ管理システムを作成することで、ビッグ・データとOracle Databaseの能力を統合します。

Oracle Database のバージョン	データベース・ ハードウェア	Hadoopクラスタ・ ハードウェア	Hadoopディストリ ビューションと バージョン
19c	Oracle Exadata (Linux OL6、OL7) またはIntel x86 64 ビット・システム (Linux OL6、OL7、 RHEL6、RHEL7)	Intel x86 64ビット・ システム (Linux OL6、OL7、 RHEL6、RHEL7)	<ul style="list-style-type: none"> • CDH* 5x (5.5以降)、 6x • HDP ** 2x (2.3以降)
19c	Oracle Exadata (Linux OL6、OL7) またはIntel x86 64 ビット・システム (Linux OL6、OL7、 RHEL6、RHEL7)	Oracle Big Data Appliance (Linux OL6、OL7)	<ul style="list-style-type: none"> • CDH* 5x (5.5以降)、 6x

* CDH : Cloudera's Distribution Including Apache Hadoop

** HDP : Hortonworks Data Platform

始めるには

Oracle Big Data Lite Virtual Machineで、Oracle Big Data SQLや、オラクルのビッグ・データ・プラットフォームのその他のコンポーネントをお試しください

(<http://www.oracle.com/technetwork/database/bigdata-appliance/oracle-bigdatalite-2104726.html>)。Big Data Liteを使用すると、ラップトップまたはデスクトップ・コンピュータからオラクルのビッグ・データに関する機能の動作をテストできます。機能には、CDH、Oracle Big Data Spatial and Graph、Oracle Big Data Discovery、Oracle Big Data Connectors、Oracle Data Integrator、Oracle Golden Gateなどが含まれています。

お問い合わせ

Oracle Big Data SQLの詳細については、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。

ORACLE

CONNECT WITH US



Integrated Cloud Applications & Platform Services

Copyright © 2019, 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、ここに記載されている内容は予告なく変更されることがあります。本文書は、その内容に誤りがないことを保証するものではなく、また、口頭による明示的保証や法律による黙示的保証を含め、商品性ないし特定目的適合性に関する黙示的保証および条件などのいかなる保証および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。0116



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