

Oracle Big Data SQL 3.0



おもな機能

- Oracle Database、Hadoop、NoSQL ソース全体でシームレスにデータの問合せを実行
- 変更なしですべての Oracle SQL 問合せを実行—アプリケーションの投資を維持
- 設置場所でデータを解析し、インテリジェントにフィルタリングすることにより、Hadoop および NoSQL 上の Smart Scan はパフォーマンスを強化
- Oracle Database 12c の主要 Hadoop ディストリビューションへのアクセスをサポート
- Oracle Database Security 機能により機密データへのアクセスを制御
- Copy to Hadoop を使用して Oracle Database から Hadoop に容易にデータをコピー
- Oracle Table Access for Hadoop and Spark を使用して Hadoop から Oracle Database 内のデータに動的にアクセス

モノのインターネットから世界的な顧客ベースのソーシャル・ストリームまで、企業が有効に活用できるデータの未曾有の増加により、ビジネスの強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。これらの巨大な新しいデータ・リソース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な可能性があります。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、分析とレポートのためのコードを作成しなおすために多額の投資が必要となることがあります。

Oracle Big Data SQL を使用すると、組織は既存の 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、NoSQL、Oracle Database にまたがるデータを並外れたパフォーマンスで分析できます。Big Data SQL を使用することで、データ・サイエンス処理の単純化からデータ・レイクの開放まで、できるだけ多くのエンドユーザー・グループでビッグ・データの利点が利用できるようになります。

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、Apache Hadoop や多数の 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 Database、Apache Hadoop、NoSQL のデータを 1 つの SQL 問合せ内で結合する。
- Apache Hadoop および NoSQL でデータの問合せと解析を行う。
- 分散アーキテクチャ内で Smart Scan、ストレージ索引、条件プッシュ・ダウンなどの高度な技法を利用してすべてのデータの問合せの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 のセキュリティおよびアクセス・ポリシーを Apache Hadoop および NoSQL のデータに拡張する。

強化された外部表

おもな利点

- データのローカル処理を利用することにより、高い問合せのパフォーマンスを達成
- 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
- Oracle Database セキュリティ・ポリシーを使用して機密データの安全性を保護

異なるシステムに格納されたビッグ・データを扱うとき、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 では、Oracle Database 12c に新しい外部表タイプを追加することで、Oracle Database が、Hadoop システムおよび NoSQL システムのデータのカタログ作成および保護を行うための単一の格納場所となります。Big Data SQL では、クラスタとクラスタ内の中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 のデータと Hadoop および NoSQL データベースのデータを結合するシームレスなメタデータおよび問合せ
- 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 クラスタへの問合せを可能にする複数クラスタのサポート
- 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に強化されたアクセス・パラメータ

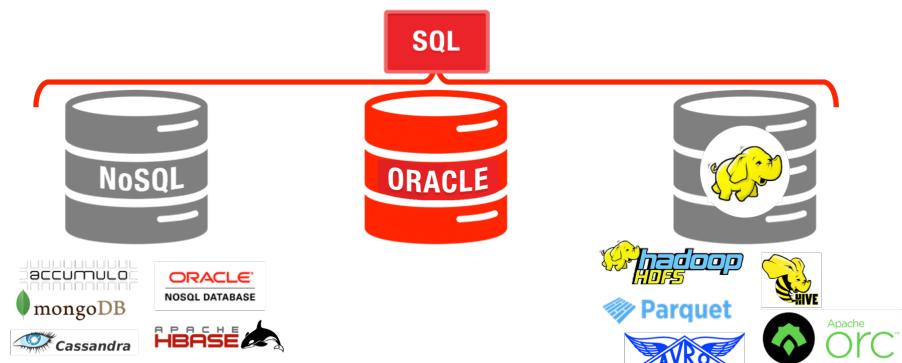


図1：Oracle Big Data SQLを使用すると、Oracle Database、Apache Hadoop、選択されたNoSQLのデータ・ストアで Oracle SQLの問合せを実行できます。

Smart Scan : データ主導のパラレル処理

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、分析のパフォーマンスには、データを分析に移すのではなく、分析をデータに移すことが重要になります。Big Data SQL は、オラクルのクラス最高の Exadata Database Machine で最初に導入された Smart Scan の機能を Hadoop および NoSQL システムに適用しています。Smart Scan を使用して、Oracle SQL のオペレーションをビッグ・データ・システムのストレージ層まで下げることができます。Smart Scan はこれらのビッグ・データ・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パラレル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、ローカルでのデータのフィルタリング
- Bloom フィルタによる結合の最適化により、Oracle Database のデータと Hadoop の巨大データの間の結合を高速化

- データ・マイニング・モデルのスコアリングと、JSON や XML などでドキュメントのデータセットの問合せを行うための強化された処理
- 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、Smart Scan のほか、I/O の発生前にビッグ・データの処理を高速化するためのストレージ索引テクノロジーも備えています。Oracle Big Data SQL はビッグ・データの処理時に、関連データの格納場所を示すローカルのメモリ内の索引を自動的に作成します。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・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・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：ハーネス外部ストレージ・システム

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 では以下が行えます。

- Apache Hive によって管理される表から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

ビッグ・データでの Oracle Database Security

Oracle Big Data SQL の独自のデータ統合アプローチにより、Hadoop データおよび NoSQL データに対して Oracle Database Security 機能が使用可能になります。オラクルの標準のセキュリティ・メカニズムを使用し、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
- ビッグ・データ・システム内の機密データが未承認ユーザーに提示されない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

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 が動作し、Oracle Clusterware に対応する Oracle Database 12c（バージョン 12.1.0.2 以上）、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が得られます。Big Data SQL は Oracle Exadata と Oracle Big Data Appliance の機能を利用してクラス最高の Big Data Management System を

作成し、ビッグ・データと Oracle Database の機能を統合します。クラウドのビッグ・データで Oracle SQL の機能を利用するため、Big Data SQL Cloud Service を Oracle Database Cloud Service (Exadata Edition と Oracle Big Data Cloud Service) と一緒に使用することができます。

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
12.1.0.2 以上	Intel x86 64 ビット・システム (Linux OL6、RHEL6)	Intel x86 64 ビット・システム (Linux OL6、RHEL6)	CDH* 5.5 以上 HDP ** 2.3 以上
12.1.0.2 以上	Oracle Exadata (Linux OL6)	Oracle Big Data Appliance (Linux OL5 および OL6)	CDH 5.5 以上

* CDH : Cloudera's Distribution Including Apache Hadoop

** HDP : Hortonworks Data Platform

Hadoop へのコピー

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が簡単に行えます。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 SerDe に関連するオーバーヘッドが必要となります。Big Data SQL 以外のアクセスは、カスタム SerDe を使用してコピー済みデータを Hive などによって問合せすることにより、容易になります。

Oracle Table Access for Hadoop and Spark

すべてのビッグ・データのワークロードがデータベース内で開始されるわけではありません。そのため、ワークロードでは Hadoop および Spark から Oracle Database に格納されたデータにアクセスする必要があります。Oracle Big Data SQL には、Oracle Database のデータを透過的に問合せするための Oracle Table Access for Hadoop ライブラリが用意されています。

始めるには

Oracle Big Data Lite Virtual Machine でオラクルのビッグ・データ・プラットフォームの別のコンポーネントと併せて、Oracle Big Data SQL を試しに使ってみてください

(<http://www.oracle.com/technetwork/database/bigdata-appliance/oracle-bigdata-lite-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・データの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



Integrated Cloud Applications & Platform Services

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、ここに記載される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默示を問わず、特定の目的に対する商品性もしくは適合性についての默示的な保証を含み、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。0116



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