

Oracle Big Data SQL

リリース 3.1



IoT (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、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、NoSQL、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。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、最大のエンドユーザー・グループでビッグ・データの利点を利用できるようになります。

おもな機能

- Oracle Database、Hadoop、NoSQL ソースに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
- Hadoop、NoSQL 上のスマートスキャン機能により、データが存在する場所でパースおよび処理を行うことでパフォーマンスを向上 Oracle Database 12c から主要な Hadoop ディストリビューションにアクセスが可能
- Oracle Database、Hadoop、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 はオラクルによるデータ仮想化のイノベーションです。異種のデータセットに対し別々の SQL や API (REST、Node.js など) によるアクセスを行うのではなく、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 の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、パーティション・プレーニング、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すべてのデータ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

- Hadoop、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
- 比較的低成本の Hadoop ストレージにより情報ライフサイクルマネジメント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

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を取り扱えるようにしました。Big Data SQL では、対象データソースを Oracle Database にコピーや移動する事なく、メタデータの変更だけで、対応する事が可能です。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統合および問合せ
- 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

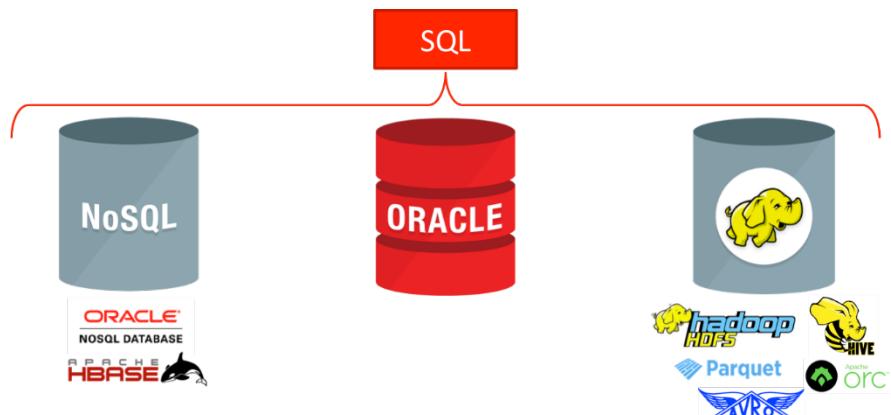


図 1：Oracle Big Data SQL を使用すると、Oracle Database、Apache Hadoop、NoSQL のデータ・ストアに対し透過的に Oracle SQL の問合せを実行できます。

Smart Scan：データドリブンのパラレル処理

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、優れた分析パフォーマンスを実現するには、データを分析する場所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、大規模なデータセットに対する最適な並列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
- 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 などで記述されたドキュメントのデータセットを問い合わせせる場合の**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

Storage Indexing : 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。

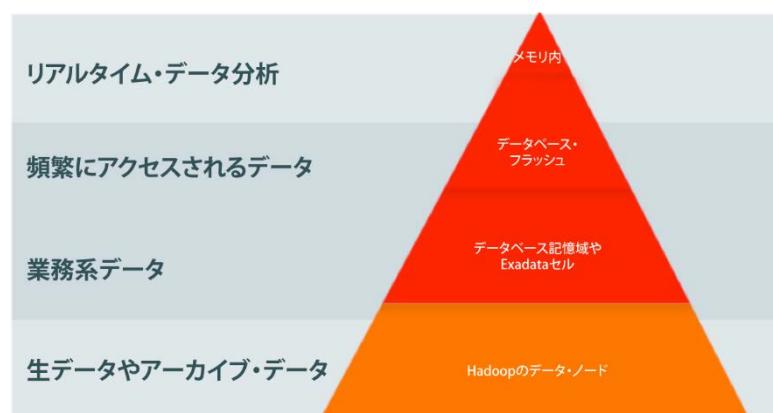
Predicate Push-Down : 外部ストレージ・システムの利用

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**

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、のような構成になります。



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 SerDe に関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。

表領域をHDFSに格納しSmart Scan

Oracle Partitioning は、1 つの表のデータ・パーティションをさまざまな場所に格納できるようにするテクノロジーです。このテクノロジーを使用すると、表に含まれる変更が発生しないアーカイブ・データを、高度に最適化された Oracle Database フォーマットで Hadoop に保存できます。データベース問合せを実行すると、ほかのデータへのアクセスと同様にシームレスに、このアーカイブ・データにアクセスします。また、問合せの高速化を図るために、Smart Scan や Storage Index が利用されます。

Big Data SQL の Smart Scan 機能を使用すると、更なるパフォーマンス恩恵を得ることができます。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 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
- 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・リダクション
- 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 が動作する Oracle Database 12c (バージョン 12.1.0.2 以上) と、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。クラウドでのビッグ・データの処理に Oracle SQL の機能を利用したい場合は、Oracle Exadata Cloud Service および Oracle Big Data Cloud Service に Big Data SQL Cloud Service を組み合わせることができます。

詳しくは、My Oracle Support のドキュメント「Oracle Big Data SQL における互換性マトリックス マスターノート (Doc ID 2130807.1)」をご参照ください。

Oracle Database	データベース・ハードウェアのバージョン	Hadoopクラスタ・ハードウェア	Hadoopクラスタのバージョン
12.1.0.2以上	Oracle Exadata (Linux OL6) または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)	<ul style="list-style-type: none"> • CDH* 5.5以上 • HDP ** 2.3以上
12.1.0.2以上	Oracle Exadata (Linux OL6) またはIntel x86 64 ビット・システム (Linux OL6 RHEL6)	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

始めるには

Oracle Big Data Lite という VM 環境で Oracle Big Data SQL を簡単に試すことが可能です。 (<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について、詳しくは 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。

Integrated Cloud Applications & Platform Services

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. 本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默示的保証を含め、商品性ないし特定目的適合性に関する默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。オラクルは本書に関するいかなる法的責任も明確に否認し、本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

CONNECT WITH US

-  blogs.oracle.com/oracle
-  facebook.com/oracle
-  twitter.com/oracle
-  oracle.com