

Oracle Big Data Connectors



「Oracle Big Data Applianceを特別に魅力的にしていたのはその拡張性です。また、Oracle Big Data Connectorsを使うことで、Oracle Big Data ApplianceとOracle Database Appliance間のデータ移動をシームレスに実行でき、高スループットと低レイテンシも実現できました」

CRAIG FRYAR

Wargaming、ビジネス・インテリジェンス担当主任

おもなビジネス上の利点

- データ検出アプリケーションをビジネス・ユーザーに迅速に提供
- HadoopからOracle Databaseへの高パフォーマンスの大量ロード
- データベースでOracle SQLを使ったHadoop上のデータ問合せ
- データ・サイエンティストが慣れ親しんだR向けIDEを使用してHadoop上でRを実行することで、迅速な分析を実現

Oracle Big Data Connectors は、Apache Hadoop 上の処理を Oracle Database 上の処理と統合するソフトウェア・スイートです。Hadoop および Oracle Database 向けに最適化された Oracle Big Data Connectors を使用すると、構造化データと非構造化データを使用して重要なビジネス上の洞察を獲得できます。Oracle Big Data Connectors は Hadoop から Oracle Database への超高速データ・ロードを実現するとともに、Oracle Database から HDFS データへのアクセスを提供し、使い慣れたツールでの Hadoop 上の R および XML 処理をサポートします。このコネクタは既存のスキル・セットで簡単に使用できるため、ビッグ・データ・ソリューションの開発が大幅に簡単になります。Oracle Big Data Connectors は、豊富な機能と高速接続、高いパフォーマンス、セキュリティをビッグ・データ・アプリケーション向けに提供します。

Oracle Big Data Connectors

企業はビッグ・データ・ソリューションを開発して大量の生データを処理することで、競合他社に影響を与え、個人顧客に対する洞察を獲得し、新しい分析を活用しようとしています。ビッグ・データ・ソリューションでは、Hadoop を使って大規模なデータ処理、データ変換、生データの探索的分析を実行し、その結果を Oracle Database 内のビジネス・クリティカルなデータと統合することで、リアルタイムの問合せ、高度な分析、複雑なデータ管理を実現します。Oracle Big Data Connectors は、Oracle Database に Hadoop を接続することで、ビッグ・データ・ソリューションに不可欠なインフラストラクチャを提供します。これには、高速ロード、HDFS 上のデータへの Oracle SQL アクセス、Hadoop での R および XML 分析、データ統合が含まれ、情報の検出、詳細な分析、企業内の全データの迅速な統合を可能にします。このソフトウェア・スイートのコンポーネントは次のとおりです。

- Oracle SQL Connector for Hadoop Distributed File System (Oracle SQL Connector for HDFS)
- Oracle Loader for Hadoop
- Oracle Advanced Analytics for Hadoop
- Oracle XQuery for Hadoop
- Oracle Data Integrator¹

Oracle Big Data Connectors はオラクルのエンジニアド・システムと連携しており、Oracle Exadata、Oracle SuperCluster、Oracle Database Appliance に対して Oracle Big Data Appliance を接続します。また、汎用ハードウェア上の認定 Hadoop ディストリビューションと Oracle Database に対しても接続できます。最新の認定情報については、[こちら](#)をクリックしてください。または、好みの検索エンジンで「**Oracle Big Data Connectors**」を検索して、「**Certifications**」タブをクリックします。

¹ Oracle Big Data Appliance に対してのみ、Oracle Big Data Connectors には制限付きの Oracle Data Integrator ライセンスが含まれています。その他の Hadoop プラットフォームでの使用には追加のライセンスが必要です。

- XMLアプリケーション開発者が使い慣れたXQuery言語を使って、HadoopでXMLデータを処理
- グラフィカル・ツールによって、Hadoopの複雑さを緩和
- Oracle Big Data Appliance向けの最適化
- Hadoop開発者とOracle開発者にとって使いやすいツールによる、ビッグ・データ・ソリューションの迅速な開発と導入
- 複数のデータベースとHadoopバージョンに対する認定

おもな機能

- Oracle Databaseとの緊密な統合
- Hadoopコンピューティング・リソースの活用
- Oracle SQLによるHadoopデータへのアクセスとロード
- データベースのCPU使用を最小限に抑えた、HadoopからOracle Databaseへの高速ロード
- ロードおよび問合せ中のHive表に対するパーティション・ブルーニング
- Oracle Data Integratorのグラフィカル・ユーザー・インタフェースによる、Hadoopでのデータ変換ワークフローの実行
- RプログラムからHadoopジョブへの自動変換
- 大量のXMLファイルの並列処理とXQuery結果のOracle Databaseへのロード
- Kerberos認証によるHDFSデータへのセキュアなアクセス

Oracle SQL Connector for Hadoop Distributed File System (Oracle SQL Connector for HDFS)

Oracle SQL Connector for HDFSを使用すると、データベースからOracle SQLを使ってHadoop上のデータを検索できます。データのアクセスに使用する外部表には、データベース内のその他の表と同様の方法で問合せできます。また、外部表からデータを選択してデータベース内の表に挿入する方法でデータをロードすることもできます。Oracle Big Data ApplianceからOracle Exadataへのロード速度は15TB/時です。

Oracle SQL Connector for HDFSを使用すると、テキスト・ファイルまたはHive表（テキスト・ファイル経由）に対するデータの問合せとロードを実行できます。Hiveパーティション表に対して問合せを実行する場合、Oracle SQL Connector for HDFSがHiveパーティションの一部にのみアクセスするように制限することで、アクセスするデータを最小限に抑え、パフォーマンスを向上できます。また、Oracle SQL Connector for HDFSは、オフライン・ロード用にOracle Loader for Hadoopで作成したデータファイルや、アーカイブ用にデータベースからHadoopにコピーしたデータファイルなどのOracle Data Pump形式のファイルを検索できます。

機能	
Oracle SQLによるHadoop内のデータへのアクセス	Oracle DatabaseからHDFS内のHive表やテキスト・ファイルへの直接的な問合せ
Hiveパーティション表に対するパーティション認識型アクセス	該当するHiveパーティションのみに対するロードまたは問合せによる、パーティション・ブルーニング
並列な問合せとロード	Oracle Databaseに対する高速の並列問合せとロード
セキュリティ	Kerberos認証を使用したアクセス
柔軟性と使いやすさ	外部表の作成ツール
Oracle SQLを使用した、HDFS上のOracle Data Pumpファイルへのアクセス	データベース・データをData PumpファイルとしてHadoop上にアーカイブし、Oracle Databaseから問合せ

Oracle Loader for Hadoop

Oracle Loader for Hadoopは、HadoopからOracle Databaseに高速でデータを移動するための、高パフォーマンスなロード・ツールです。Oracle Loader for Hadoopは、Hadoopのコンピューティング・リソースを利用してデータのソート、パーティション化、変換を行い、データベースへのロード前にデータをOracle対応のデータ型にします。Oracleデータ型へのデータ変換により、ロード中のデータベースのCPU使用量が削減されるため、データベース・アプリケーションへの影響を最小限に抑えられます。このため、大量データを取り込む際に通常問題となる、リソースをめぐる競争が軽減されます。この機能は連続的で頻繁なロードで特に便利です。

Oracle Loader for Hadoopは革新的なサンプリング技術を使用して、並列・データ・ロード中に、Hadoopノード間でデータをインテリジェントに分散します。このため、並列・アプリケーションの共通の懸念事項であるデータの偏りの影響を最小限に減らすことができます。

Oracle Loader for Hadoopは、基本表圧縮、高度な行圧縮、Hybrid Columnar Compressionのいずれのデータベース圧縮を使った表に対してもデータをロードできます。

Oracle Big Data Connectors

Oracle Big Data Connectorsは、HadoopとOracle Database間に高パフォーマンスの統合を提供します。

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 Exadata
- Oracle NoSQL Database
- Oracle Big Data Discovery
- Oracle Exalytics
- Oracle Business Intelligence
- Oracle Data Integrator

関連サービス

オラクルの主要製品をサポートするサービスは次のとおりです。

- アップデート・サブスクリプション・サービス
- 製品サポート・サービス
- プロフェッショナル・サービス

Oracle Loader for Hadoopでは、さまざまな入力形式のデータをロードできます。テキスト・ファイル、Hive 表、ログ・ファイル（解析とロード）、Oracle NoSQL Database などから、ネイティブにデータをロードできます。また、Hive にアクセスできる入力形式（Parquet、JSON ファイルなど）と入力ソース（HBase など）からは、Hive 経由でデータをロードできます。さらに、Oracle Loader for Hadoop は、ユーザーが提供するカスタム入力形式の実装によって、独自のデータ形式を読み取ることができます。

機能	
Hadoop へのデータ型変換のオフロード	データベース CPU への影響を最小限に抑制
パラレル・ロード	高速パラレル・ロードによる Hadoop ノードの活用
ロードバランシング	自動ロードバランシングによるデータの偏りの解消
セキュリティ	Kerberos 認証を使用したアクセス
入力形式	テキスト・ファイル、Hive、Parquet、JSON、シーケンス・ファイル、圧縮ファイル、ログ・ファイルなどの各種入力形式からロード可能
パーティション認識型ロード	Hive パーティション表の対象パーティションのみのロード

Oracle R Advanced Analytics for Hadoop

Oracle R Advanced Analytics for Hadoopは、Hadoop上でRコードを実行することで、スケーラブルな分析を可能にします。このコネクタは、Hadoopベースのコンピューティングが持つ複雑さをRユーザーに感じさせることなく、Rユーザーが選んだ任意のIDEで開発されたスタンドアロンのデスクトップ・アプリケーションを使用して、Hadoop上でRコードをパラレル実行できるようにします。また、Oracle DatabaseのOracle Advanced Analyticsと統合することで、R計算とデータベース内のデータ・マイニング計算をデータベース内で直接実行します。

Oracle R Advanced Analytics for Hadoopは、一般的な統計および予測手法に対する高パフォーマンスでスケーラブルなパラレル実装を豊富に使うことで、別のプラットフォームへのデータ移動なしで、迅速な洞察の獲得を可能にします。サポートされる手法の一覧は以下の表を参照してください。Oracle R Advanced Analytics for Hadoopは、パラレルRコードに対するR方式のデバッグ機能を使ったユーザー・デスクトップでの迅速な開発をサポートしており、内部でパラレル処理をシミュレートします。

このコネクタを使用すると、複数の環境（クライアント・デスクトップ、HDFS、Hive、Oracle Database、インメモリのRデータ構造）からデータを組み合わせ、1つの分析タスク実行のコンテキストで使用できるため、データの集約と準備が大幅に簡素化されます。また、Oracle R Advanced Analytics for Hadoopは、Rコードをパラレル実行するための一般的な演算フレームワークを提供します。

機能	
ビッグ・データ向けのスケーラブルな分散型分析	Hadoop でのネイティブの分散型 R 分析による R コードの透過的なパラレル実行
新しいスキル・セットの要らない迅速なデプロイと使いやすさ	開発者の生産性：パラレル処理スキルを必要としない、ユーザー・デスクトップ上の使い慣れた R 環境での R コードの開発とデバッグ 簡素化されたインタフェースを通じた、R ユーザーによる Hadoop パラレル処理の活用 混成データの集約とスケーラブルなデータ準備のサポート
ネイティブな分散型 R 分析	統計と高度な行列計算 ・ 共分散と相関性の行列計算 ・ リザーバ・サンプリング ・ 主成分分析 ・ 低ランクの行列因子分解を使用した行列完補完 ・ 非負行列因子分解 回帰モデル ・ 線形回帰 ・ 単層のフィード・フォワード・ニューラル・ネットワーク ・ 一般化線形モデル 分類モデル ・ 一般化線形モデルに基づくロジスティック回帰 ・ k-means クラスタリングを使用したセグメント化

Oracle XQuery for Hadoop

Oracle XQuery for Hadoop を使用すると、Hadoop に保存されているテキスト、XML、JSON、Avro などのコンテンツを、XQuery を使って処理、変換できます。Oracle XQuery for Hadoop は Hadoop のパラレル処理機能を利用し、クラスタ内の全ノードを使って W3C XQuery 式を評価します。

Oracle XQuery for Hadoop は、Oracle Database の実績ある XQuery エンジンに対する Java 実装をベースとしています。XQuery エンジンが Hadoop 向けに最適化されています。XQuery エンジンが、Hadoop 上の XQuery 式を透過的にパラレル評価することで、XQuery 式をクラスタ内の全ノードに配布します。この場合、XQuery プロセッサにデータを移動するのではなく、XQuery 処理をデータの近くに移動します。この問合せ評価方法は、他の XQuery ソリューションよりはるかに高いスループットを実現します。

Oracle XQuery for Hadoop の一般的なユースケースとしては、Web ログの分析、テキストや XML、JSON、および Avro コンテンツに対する変換操作があります。処理されたデータはデータベースにロードするか、または Hadoop プラットフォーム上の Cloudera Search に索引付けできます。

機能

スケーラブルでネイティブな XQuery 処理	XQuery エンジン は Hadoop クラスタ全体に自動的に分配されるため、データのある場所で XQuery を実行できます。
入力データ・ストア	HDFS、Hive、Oracle NoSQL Database に格納されたデータを処理します。
Hadoop テクノロジーの統合	Apache Oozie ワークフローから Oracle XQuery for Hadoop ジョブを実行します。 Cloudera Search Hive 向けの XML/XQuery 拡張機能
パラレル XML 解析	非常に大きい XML ドキュメントの超高速処理
Oracle Database への XQuery 結果の高速ロード	Oracle Loader for Hadoop を使った高速処理

Oracle Data Integrator Enterprise Edition

Oracle Data Integrator には、ネイティブ Hadoop 統合に対する包括的なナレッジ・モジュールが含まれています。このモジュールを使うと、使い慣れたグラフィカル・インタフェースから複雑なデータ変換およびデータ移動処理を指定して実行できるため、Hadoop でのジョブ実行が大幅に簡素化されます。ソース間のデータ移動（HBase から Oracle Database、HBase から Hive、Hadoop からサード・パーティ・データベース、サード・パーティ・データベースから Hadoop など）を、グラフィカルに定義できます。Oracle Loader for Hadoop および Oracle SQL Connector for HDFS 用のナレッジ・モジュールは、Hive および HDFS から Oracle Database への高速データ移動を可能にします。グラフィカル・ツールと宣言的な仕様を備えた Oracle Data Integrator ナレッジ・モジュールを使うと、Hadoop アプリケーションで通常必要になる複雑なコードの作成が不要になります。

Oracle Big Data Appliance に対するライセンスでは、Oracle Big Data Connectors に制限付きの Oracle Data Integrator ライセンスが含まれます。その他の Hadoop プラットフォームでの使用には追加のライセンスが必要です。

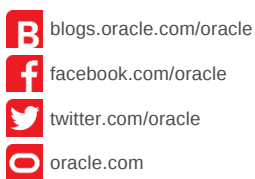
機能

開発者の生産性向上のための最適化	使い慣れた Oracle Data Integrator グラフィカル・ユーザー・インタフェース Hadoop ジョブに対するエンド・ツー・エンドの連携 Oracle Data Integrator による Hadoop ジョブの作成と調整
Hadoop とのネイティブ統合	Oracle Data Integrator 内で Hive メタデータを表示可能 Hadoop 内で変換とフィルタリングを直接実行
パフォーマンス向上のための最適化	最適化された Hadoop Oracle Data Integrator ナレッジ・モジュール Oracle Data Integrator で Oracle Loader for Hadoop と Oracle SQL Connector for HDFS を使用した、Oracle Database への高パフォーマンスなロード



CONTACT US For more information Oracle Big Data Connectors, visit oracle.com or call +1.800.ORACLE1 to speak to an Oracle representative.

CONNECT WITH US



Hardware and Software, Engineered to Work Together

Copyright © 2014, Oracle and/or its affiliates. All rights reserved. Copyright © 2015, 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。オラクル社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。本文書はオラクル社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