

Oracle ホワイト・ペーパー

2019年7月

Rのエンタープライズ対応

*Rを使用してエンタープライズレベルのパフォーマンス、
スケーラビリティ、容易な本番デプロイメント、
セキュリティを実現*

免責事項

下記事項は、弊社の一般的な製品の方向性に関する概要を説明するものです。また、情報提供を唯一の目的とするものであり、いかなる契約にも組み込むことはできません。マテリアルやコード、機能の提供をコミットメント（確約）するものではなく、購買を決定する際の判断材料になさらないでください。オラクル製品に関して記載されている機能の開発、リリースおよび時期については、弊社の裁量により決定されます。

概要：ラップトップを超えて.....	4
Oracle DatabaseでのRコードの実行.....	5
CRANパッケージでのインデータベースのサポート.....	6
本番環境へのR分析のデプロイ.....	7
HadoopとSparkを使用したRの実行.....	7
Oracle Databaseを使用したビッグ・データのIoTユースケース.....	8
Oracle DatabaseとHadoopを使用したビッグ・データのユースケース.....	9
ユースケース1：信用リスクの分析.....	9
ユースケース2：不正の検出.....	9
ユースケース3：顧客離れの防止.....	10
まとめ：R for the Enterprise.....	10

概要：ラップトップを超えて

専有の統計環境に代わる選択肢として1994年に開発されたRは、統計分析、データ操作、機械学習、グラフィカル表示向けのソフトウェア機能の統合スイートです。このオープン・ソースのスクリプト言語は、データ・サイエンティスト、ビジネス・アナリスト、データ・アナリスト、および統計家にとって、重要な分析手段の1つとなっています。数百万に上る世界中のRユーザーが、数千ものオープン・ソースRを利用しており、バイオインフォマティクス、空間統計、金融市場分析、線形/非線形モデリングなどの幅広い分野で、Rエコシステムによって生産性を向上しています。

データ・サイエンティストは通常はパソコンやワークステーション上でRプログラムを実行しますが、大量のデータに対して高度な演算処理を素早く実行する必要性が増しています。オラクルは、Rを大規模なデータに使用できるようにするために、HadoopやOracle Databaseに格納されたデータに対して、統計分析やグラフィカル分析を実行できる幅広いオプションを開発しました。これにより、高いレベルのセキュリティ、スケーラビリティ、パフォーマンスを必要とするプロジェクトにエンタープライズレベルの機能がもたらされ、オンプレミスまたはOracle Public Cloudの本番環境にRスクリプトを迅速かつ容易にデプロイできるようになりました。

Oracle Machine Learningプラットフォームは、Rの利点と、Oracle DatabaseとHadoopが持つパワーとスケーラビリティを併せ持っています。Oracle Machine Learning for R (OML4R) は、Oracle Database Enterprise EditionのOracle Advanced AnalyticsオプションのコンポーネントであるOracle R Enterprise (ORE) でサポートされており、Oracle Cloud製品とともに使用できます。Oracle Machine Learning for Sparkは、Oracle Big Data ConnectorsのコンポーネントであるOracle R Advanced Analytics for Hadoop (ORAAH) でサポートされており、Oracle Big Data Serviceとともに使用できます。RプログラムとRパッケージをこれらのアセットとともに使用することで、大量のデータをセキュアな環境で処理できます。顧客は統計モデルを構築してローカル・データ・ストアに対して実行したり、セキュアな企業データベースに格納されたデータに対してRコマンドやRスクリプトを実行したりすることができます。Rスクリプトは、Oracle Advanced Analyticsを使用してSQLから実行できます。構造化された結果やイメージといったRスクリプトの結果は、アプリケーションやダッシュボード・ツールに即座に取り込まれます。

これらの機能は、今日の多くのビッグ・データ・プロジェクトでは特に重要です。データ・サイエンティストは、制御下でOracle Database内のデータにアクセスできるため、ITセキュリティ・ポリシーを徹底しながら、生産性を大幅に向上させることができます。オラクルの統合型アプローチでは、データ分析が簡素化され、データ移動が最小限に抑えられるか排除され、生データをすぐに使用可能な情報に変換する時間が短縮されます。Oracle Big Data SQLを使用すると、Oracle Machine Learning for Rで利用できるデータソースの範囲が広がり、Rのユーザーは、Oracle Database、Hadoop環境、およびNoSQL内のデータを操作できます。

Oracle Machine Learningには、SQLユーザーとRユーザーの両者をサポートする2つのコンポーネントがあります。Oracle R EnterpriseでサポートされるOracle Machine Learning for Rを通して、Rユーザーはデータを移動することなく、標準的なR構文を使用してOracle Database内のデータを透過的に操作できるため、Oracle Databaseを高パフォーマンスの計算エンジンとして活用できます。ユーザーはR関数呼出しをSQLに変換することによって、インデータベース統計手法とデータの並列処理を利用して、スケーラビリティとパフォーマンスを向上させることができます。Oracle Machine Learningでは、Oracle Database内で実行される強力なインデータベース機械学習アルゴリズム一式が提供されます。Oracle Databaseの制御の下で、ユーザー定義のRスクリプトをデータベース・サーバー・マシンRエンジンで実行することもできます。このユーザー定義のRスクリプトでは、RまたはSQLのいずれからも呼び出すことができるオープン・ソースのRリポジトリであるComprehensive R Archive Network (CRAN) のパッケージを利用できます。Oracle Machine Learningのユーザーは、ビジネス・アナリストからデータ・サイエンティストに至るまでのユーザーを対象とする、オラクル製およびサードパーティ製のさまざまなR GUIやIDEオプションを使用できます。

さらに、Oracle Big Data Connectorsの1つであるOracle R Advanced Analytics for HadoopでサポートされるOracle Machine Learning for Sparkにより、Rユーザーは標準的なR構文を使用して、Apache HiveやApache Impalaのデータを透過的に操作できます。緊密に統合されたRインタフェースを介して多数のSpark MLlibアルゴリズムを公開するなど、Sparkベースの並列分散機械学習アルゴリズムも豊富に提供されます。

ユーザーは、RからカスタムのMapReduceジョブを実行することもできます。マッパー関数とリデューサ関数はRで記述されており、CRANパッケージを利用できます。

OML4Rは、Oracle R Distributionと呼ばれるオラクルのオープン・ソースR再配布、およびオープン・ソースRディストリビューションと組み合わせて使用できます。Oracle R Distributionは、高度な線形代数や行列処理に対応したIntelのMKLなどの高パフォーマンス・ライブラリを使用して、簡単に拡張できます。

Oracle R Distributionはオラクルによってサポートされています。

このホワイト・ペーパーでは、開発者が以下を行うことができるようにすることで、オラクルがオープン・ソースRをいかに拡張しているかについて説明します。

- Oracle DatabaseまたはHadoopのデータを透過的に分析および操作する
- データとタスクの平行処理によりデータベースでRスクリプトを実行する
- Rを通してインデータベースSQLベースのアルゴリズムをシームレスに使用する
- データベースでRモデルをスコアリングする
- SQL文からRスクリプトを容易に実行する
- RをITソフトウェア・スタックに統合する

Rを、オラクルのデータベースおよびインフラストラクチャのコア製品とHadoopの両方と統合することで、2つの異なる利点、すなわち使い慣れた強力な統計環境と、スケーラビリティ、パフォーマンス、セキュリティの大幅な向上を実現できます。

Oracle DatabaseでのRコードの実行

オラクルは、Oracle Database内でR演算を実行できるR/パッケージ一式を開発しました。この透過レイヤーにより、オラクルの表およびビューはネイティブのRオブジェクトであるかのように、data.frameのサブクラスであるore.framesを介して、R環境からアクセスできます。これにより、ユーザーは幅広いR機能を実行できます。データ・サイエンティストはSQLの知識がなくても、好きなR IDEを使用して探索的データ分析を行い、自然統計言語であるRの幅広い分析機能にアクセスできます。Rユーザーは、一般的なオープン・ソースR dplyrパッケージからオーバーロードされた機能を提供するパッケージOREdplyrを利用することもできます。これらの機能を使用することで、ユーザーはデータ・アクセス、スケーラビリティ、パフォーマンスなどの課題ではなく、データ分析に集中することができます。

透過レイヤーのおかげで、R開発者は使い慣れた環境、言語、ツールを使用できます。内部では、オーバーロードされたR関数が、データベースの平行処理、問合せの最適化、列の索引付けとパーティション化などの利点を生かし、統計機能の豊富なインデータベース・ライブラリを活用しながら、Oracle Database内で実行されます。Rユーザーは、標準的なRの開発スキルやツールを使用して、このような複雑な演算をデータベース内で実行できます。

ユーザーは、予測分析手法を構築、評価、共有、デプロイしながら、パフォーマンスに優れたオラクルの並列分散機械学習アルゴリズムも活用できます。これらのインデータベース・アルゴリズムが表およびビューからテキスト列をアクセプトすることで、テキスト・マイニングが統合され、用語とテーマの抽出が自動化されます。抽出されたデータは、その後、モデルの構築やデータのスコアリングにおいて他の予測変数と結合されます。

データ・サイエンティストの生産性をさらに向上させるために、パーティション・モデルと呼ばれるモデル一式を自動的に作成できます。このパーティション・モデルでは、各コンポーネント・モデルがユーザー指定のデータ・パーティションに構築されます。単一の統合されたモデルを使用して、スコアリングを有効化して簡素化します。

データ・サイエンティストとアプリケーション開発者は、データのある場所にアルゴリズムを持ち込むことで、データ量の増加に合わせて分析プロジェクトを容易に拡張できます。たとえば、ディシジョン・ツリー、サポート・ベクター・マシン、k-means、ニューラル・ネットワーク、段階的回帰、スケーラブルな機械学習のためのランダム・フォレストといったネイティブな並列分散インデータベース・アルゴリズムを使用できます。オラクルの強力なエンジニアド・システムの1つであるOracle Exadata上で実行する場合は、さらに高速にデータを分析し、予測を行うことができます。この処理がストレージ層で行われるためです。そのため組織は、オラクルのエンジニアド・システムがもたらす優れたパフォーマンスの利点をいっそう享受できます。オラクルのアプローチには、次のような明白な利点があります。

- R内でデータの準備、分析、可視化を単独で行うことが可能
- 問合せの最適化、列の索引付け、パラレル処理、インメモリ実行とパーティション化のオプション機能を備えた高パフォーマンスの計算エンジンとしてデータベースを利用可能
- フラット・ファイル・データの管理を排除し、ストレージ、バックアップ、リカバリ、セキュリティに付随する複雑性を解消
- Rのメモリ制約を最小限に抑えることで、ビッグ・データ要件に対応
- RスクリプトをSQLから実行して、エンタープライズ・アプリケーションとダッシュボードの容易なデプロイメントと統合を実現

CRANパッケージでのインデータベースのサポート

オラクルの**組み込みR実行機能**を使用すれば、データ・サイエンティストは、Comprehensive R Archive Networkリポジトリにある何千もの専門アルゴリズムを活用できます。自身でアルゴリズムを作成するか、既存のアルゴリズムをダウンロードしてから、これらのパッケージをデータベース・サーバー側のRエンジンにインストールできます。このアーキテクチャにより、データベースに対するデータの送受信、および選択したアルゴリズムへのデータの直接フィードが容易になります。

索引付けを通したパラレル・フィードにより、高度で複雑なアルゴリズムの実行が可能になります。たとえば、顧客表を郵便番号で分割し、複数のRエンジンをパラレル実行することで、R環境から離れることなく、郵便番号の異なる複数の顧客グループを同時に処理できます。透過レイヤーからアクセス可能なものから、CRANパッケージからアクセス可能なものまで、多様な統計技術を提供するRスクリプトは、オラクルのインデータベースRスクリプト・リポジトリ内に構築して保存できます。

- Rで独自のパッケージを作成し、Oracle Database管理下のデータベース・サーバー・マシンで実行
- CRANオープン・ソース・パッケージを利用
- Oracle Databaseのスケジューリング機能を使用して、SQLインタフェース経由でRスクリプトの"完全自動"実行を実現
- Oracle Database管理下で、データとタスクをパラレル処理するRスクリプトを実行して、大規模ジョブを高速化
- 結果をアプリケーション、BIダッシュボード、レポートと統合

グラフ分析を使用した機械学習の強化

Oracle Spatial and Graphのオプション（別途ライセンスが必要）として提供される強力なグラフ分析機能の利用に関心がある方のために、OML4Rでは、インデータベース機械学習アルゴリズムとParallel Graph AnalytiX (PGX) エンジンに容易に使用できるようにするパッケージOAAgraphをご用意しています。RをOracle Databaseとともに使用して、データの準備、モデルの構築、データのスコアリングを行い、グラフ・データの分析を強化できます。また、グラフ・メトリクスを計算して、インデータベース機械学習アルゴリズムに提供されるデータを強化できます。これらはすべて、モデル品質とグラフ分析を向上させることが目的です。

本番環境へのR分析のデプロイ

Oracle Machine Learningを利用すると、R開発者はデータベースを使用して、SQL問合せ内でRスクリプトを実行できます。そのため、標準的なビジネス・インテリジェンス環境内で容易にRスクリプトを操作できます。Oracle Databaseに対するようなSQL問合せにも、データベースRスクリプト・リポジトリに登録されているRスクリプトの呼出しを含めることができます。ユーザーは、スクリプト名を使用して、そのスクリプトを呼び出す問合せを開始し、新しい表、イメージ、またはXMLで結果を受け取ることができます。たとえば、Rスクリプトを制御するパラメータをランタイム引数として渡すことで、BIダッシュボードとグラフィカルなレポート作成アプリケーションをプログラムによって更新できます。

ある通信プロバイダは、OML4Rを使用して複雑なアンケート調査を強化しました。この企業のアナリストは、分析関数をOracle Database内に保持し、データをフィルタリングして、パラメータ化したBIダッシュボードを介して結果を表示しています。データベースとBIインフラストラクチャはどちらも、このアーキテクチャの標準コンポーネントですが、Rスクリプトと結合することで、さらに強化されました。これらの機能によって、Rは、高度な統計モデルをデータベースで直接実行できる、より強力な言語となります。

HadoopとSparkを使用したRの実行

Oracle R Advanced Analytics for HadoopによってサポートされるOracle Machine Learning for Spark (OML4Spark) は、Rパッケージ・フロントエンドを備えたコンポーネントであり、Hadoopクラスタのデータに対して最先端のSparkベースの機械学習アルゴリズムを提供するとともに、HDFS、Apache Hive、Apache Impala、Spark DataFramesに格納されたデータとHadoopへの透過的なアクセスを実現します。OML4Sparkを使用すると、大量のデータに対して効率的にRモデルを実行でき、R環境を離れることなくSparkインメモリ処理を利用できます。ユーザーはRにより、オラクルが提供する機械学習アルゴリズムやCRAN R/パッケージを利用して、HDFSに格納されたデータを分析できます。

機械学習に関しては、OML4Sparkでは、実行時にHadoopクラスタとSparkの恩恵を受ける複数の並列分散アルゴリズムが提供されます。線形モデル、一般化線形モデル、多層パーセプトロン・ニューラル・ネットワークをはじめとするOML4Sparkのカスタム・アルゴリズムは、すべてのデータをメモリ内に収める必要がないため、Sparkではより適切に拡張されます。また、類似のオープン・ソースSpark MLlib関数よりも高速に実行されます。OML4SparkではMLlibに対する拡張インタフェースが提供されますが、この拡張インタフェースはすべてのR計算仕様を生かしており、Spark MLlibが提供するインタフェースよりも優れています。

Sparkで実行されるアルゴリズムでは、どちらのモデル・ビルドもサポートされ、HDFS、Apache Hive、Apache Impala、Spark DataFrames、およびJDBCデータソースの形式の入力データセットを使用して（予測スコアリングが）適用されます。モデル自体は、別のクラスタで実行するために、HDFSおよびローカル・ファイル・システムにバイナリ形式で保存できます。

OML4Sparkは、OML4Rがサポートする透過レイヤーを利用することで、Apache HiveやApache Impalaの表からアクセスできるデータ上でRコマンドを実行できるようにします。この透過レイヤーのおかげで、R開発者は使い慣れたR環境とRコマンドを使用でき、その内部では、関数が自動的にHQL（Hive Query Language）に変換され、Hadoopクラスタでパラレル実行されます。

MapReduceを活用するRプログラムは、Hadoopクラスタにデプロイできるほか、Hadoopクラスタのデータ並列性がもたらすパフォーマンス上の利点を享受できます。ユーザーは、Hadoop内部の知識や、MapReduce、コマンドライン・インタフェース、その他のITインフラストラクチャの知識がなくても、これらのRスクリプトを作成、実行、保存できます。

Oracle Databaseを使用したビッグ・データのIoTユースケース

モノのインターネット（IoT）は、高度な分析を利用する新たなビジネス機会を創出します。航空機、電車、車、半導体製造機械、大型ハドロン衝突型加速器、そして私たちの家など、あらゆる場所でセンサーがデータを収集しています。そのようなセンサーの一例として、家庭の電力使用量を15分おきにレポートできるスマート・メーターが挙げられます。電力会社は、このデータを使用して、各顧客の電力使用パターンをモデル化できるだけでなく、個々の顧客の使用量を予測できます。

すべての顧客に対して、電力会社は需要の総計を通常は数日または数週間単位で算出できるため、職員の配置、電力の転送や購入などをより効率的に行うことができます。

顧客ごとに1つの予測モデルを構築することは、数百万もの顧客を抱えている可能性のある電力会社にとって、興味深い挑戦です。ある電力会社が100万の顧客を抱えているとします。スマート・メーターは、1年間に350億の読取り値を収集します。ただし、それぞれの顧客が生成する読取り値は35,000ほどにすぎません。Rは、ほとんどのハードウェアで、35,000の読取り値を基に1つのモデルを容易に構築できます。ここで、各予測モデルの構築に必要な時間がわずか10秒であっても、これを順次処理すると、すべてのモデルを構築するにはおよそ116日かかることに注意してください。数日または数週間ごとに結果が必要なため、数か月の遅れは、このプロジェクトが役に立たないことを意味しています。Oracle Exadataのような強力なハードウェアを利用して、これらのモデルをパラレル処理で計算できる場合、たとえば並列度が128であれば、すべてのモデルを1日未満で計算できます。

ユーザーは、さまざまなRパッケージで提供されるパラレル処理を利用できますが、その際に考慮すべき要素がいくつかあります。たとえば、特定のモデルで障害が発生したらどうしますか。モデルは、顧客ごとに100万の個別のフラット・ファイルとして保存しますか。フラット・ファイルの場合、バックアップ、リカバリ、セキュリティにはどのように対処しますか。これらのモデルをどのように使用して顧客の電力使用量を予測しますか。また、予測データはどこに保存しますか。通常はアプリケーションやダッシュボードがSQLと連携する本番環境に、これらのRモデルをどのように組み込むことができますか。

OML4Rの**組み込みR実行機能**を使用すれば、データ・サイエンティストは顧客ごとに1つのモデルを構築するというタスクに集中できます。このモデルは、Oracle DatabaseのRスクリプト・リポジトリに保存されます。OML4Rでは、このスクリプトを`ore.groupApply`などの単一の関数から呼び出し、Oracle Databaseを利用して複数のRエンジンを作成し、1つのデータ・パーティションを指定のデータベース表からデータ・サイエンティストが作成した関数にロードし、結果モデルを再びOracle Database内のRデータ・ストアに瞬時に保存できます。これにより、モデルを計算して保存するプロセスが大幅に簡素化されます。さらに、すでに実装されている標準的なデータベース・バックアップとリカバリの仕組みを使用することで、別の専門プラクティスを考案する必要がなくなります。これらのモデルを使用した予測は、類似の方法で扱われます。

このようなRスクリプトを本番環境で使用する場合、モデルの構築と予測のどちらにおいても、ユーザーはデータ・サイエンティストが作成したRスクリプトと同じものをSQLから呼び出すことができます。予測データは、アプリケーションやダッシュボードでの読み取りや、他のSQL問合せでの使用が可能なデータベース表として直ちに利用できます。さらに、R関数を呼び出すこのようなSQL文は、Oracle DatabaseのDBMS_SCHEDULERパッケージを使用して、定期的に実行されるようにスケジューリングできます。

データ・サイエンティスト、アプリケーション開発者、および管理者は、Oracle Machine Learningの組み込み機能を利用すれば、通常は新しいプロジェクトごとに作成される複雑なコードやテスト戦略を再作成する必要はありません。代わりに、オラクルによるRとOracle Databaseの統合を活用して、アプリケーションやダッシュボードとともに使用するRベースのソリューションを容易に設計および実装し、エンタープライズ規模に拡張できます。

Oracle DatabaseとHadoopを使用したビッグ・データのユースケース

オラクルのビッグ・データ・テクノロジーは、Hadoop環境、R、およびOracle Database間で容易にデータを移動できるように設計されています。アナリストはJavaに頼ることなく、OracleやHadoopに格納されたデータにアクセスし、MapReduceプロセスや、Apache HiveまたはSparkの問合せをコーディングして、Rの機械学習アルゴリズムを実行できます。以下で説明するように、この柔軟なアーキテクチャを利用することで、組織は大規模な表やデータセットを容易に分析できます。急を要する今日のビッグ・データ課題を解決する上で、RはSQLに続き、エンタープライズ分析における素晴らしい選択肢となりました。

ユースケース1：信用リスクの分析

銀行は絶え間なく新しいサービスを顧客に提供していますが、サービスの提供条件は顧客の信用状態によって異なります。支払っている金額は、貸付残高のうちの期限切れとなった最低金額か、それ以上か。支払いが遅れたことがあるか。与信限度額のどの程度を使用しているか。その他に与信枠をどのくらい持っているか。総合的な負債/収入比はどのくらいか。

これらの変数はすべて、各顧客に設定する与信額や顧客への提示条件に関するポリシーに影響を与えます。EquifaxやTransunionなどの信用情報機関は、個人の総合的な信用履歴を調査しますが、銀行は自行の顧客について、個々の取引レベルに至るまではるかに詳細な履歴一式を調査できます。ビッグ・データ分析においても、このレベルの精度でそのような大量のデータを分析する必要があります。

たとえば、ブラジルのあるオラクル顧客は、数億のレコードに対して複数のニューラル・ネットワーク・アルゴリズムを実行して、各顧客に関する数千もの属性を調査しています。以前は、この銀行は大量のデータを高速処理して意味のある統計情報を生成することに苦心していました。そこで同行は、OML4Sparkを使用して専門アルゴリズムを実行し、Hadoopファイル・システムや、Apache Hive、その他のツールを実行しているクラスタと同じクラスタ上で、このデータをパラレル処理により分析することで、この問題を解決しました。OML4Sparkを使用すると、アナリストは、銀行が持つ大規模なHadoopファイル・システムに格納された表で、Rの分析、統計、モデルを実行できます。現在は、これらのファイル・システムとApache Hive表に対して、複雑な統計アルゴリズムを実行できるようになりました。

このアルゴリズムでは、R計算式オブジェクトなどの標準的なRのアプローチが使用されます。内部では、OML4Sparkが提供するインタフェースにより、銀行のクラスタ全体で複数のプロセッサを使用して、Sparkベースの実装またはMapReduceジョブがパラレル実行されます。アナリストは、これらのMapReduceプロセスとSparkアルゴリズムをRで作成して、Hadoopに保存できます。また、Javaを使用することなく、これらのモデルによってレビュー、グラフ作成、分析を容易に行い、結果をOracle Databaseに送信できます。

ユースケース2：不正の検出

もう1つの一般的なユースケースとして、金融取引の分析による不正検出が挙げられます。銀行、小売業者、クレジットカード会社、電気通信事業者、および他の多くの大規模組織が、この問題に取り組んでいます。データをスコアリングして潜在的な不正を検出する場合は一般的に、顧客の口座内で発生した取引について調査します（スコアリングとは、機械学習モデルを使用した結果予測です）。

顧客の通常の行動を把握したら、通常とは異なるパターンや疑わしい取引を識別できるようになります。たとえば、通常はロサンゼルスで買い物をしている顧客が、突然ローマで一連の取引を行った場合、不正である確率が非常に高くなります。ただし、頻繁に旅行する顧客の場合、ローマでのアクティビティの急増は異常パターンと通常パターンのどちらでしょうか。過去の取引すべてを捕捉してそのパターンを調査すれば、通常の行動を反映するモデルを構築できます。

Rは、このような取引を分析できる予測モデルの作成に最適なアルゴリズムと環境を備えています。CRANパッケージのアルゴリズムは、通常はマルチスレッドではありません。そのため、これを実行するマシンのメモリと単一CPUの処理能力による制限を受けます。

Rは通常、特別なパッケージやプログラミングがない場合、マルチプロセッサ・ラップトップのCPU能力を利用しません。

Oracle Machine Learningは、R言語を使用してスクリプトを定義し、Rスクリプト・リポジトリに保存し、データベースで実行することで、顧客の購入パターンの分析に伴う大量の演算処理要求に対処できます。組織は、Oracle ExadataやOracle Big Data Applianceなどのエンジニアード・システムを活用することで、この取組みを拡大し、Oracle Business Intelligence Enterprise EditionやOracle Data Visualization Desktopと統合して、結果を表示できます。R開発者は、オープン・ソースRを使用することで、Hadoopに構築された不正モデルの結果を利用できます。また、そのモデルをOracle Databaseにデプロイすると、モデルは取引レベルで迅速に行動を予測でき、エンタープライズ・アプリケーションの一部となってリアルタイムの予測を実現します。

不正を予防する上で、リアルタイムの分析は重要です。8時間（ラップトップにデータをステージングして、前日の行動の詳細分析を実行するためにかかるであろう時間）より前に発生した不正な取引を特定することも重要ですが、この取引をモデルに対してリアルタイムでスコアリングして、取引を阻止できる、またはさらなる調査のためにフラグ付けできることの方が、はるかに有益であることは明らかです。

ユースケース3：顧客離れの防止

多くの事業にとって、とりわけ電気通信などの競争の激しい市場では、顧客離れは重要な問題です。たとえば、携帯電話のユーザーは、受信状態に問題があったり、通話が頻繁に途切れたりする場合、別のサービス・プロバイダを探すことを検討するかもしれません。現在のサービス・プロバイダは、常にユーザーの行動を分析して、他社に乗り換える可能性を予測しています。サービス・プロバイダには、「問題と行動が類似する顧客の90%が、別のサービス・プロバイダに乗り換えた」ことを示す統計モデルがあります。このモデルをユーザーのデータに適用してスコアリングすることで、乗換えの可能性を明らかにすることができます。ユーザーのスコアリング・データにより、ユーザーは行動が類似する他の数百万の顧客と関連付けられます。

Oracle Machine Learningを使用すれば、顧客がWebページを閲覧したり、モバイル・アプリを使用したりする際に、これらのRモデルを実行して、現在の行動、および業務系データ・ストアやデータウェアハウスに対するリアルタイム分析に基づき、その場で提案を行うことができます。

スコアリングはバッチ処理によりオフラインでも実行できます。たとえば、特別なオファーや広告キャンペーンの対象とする顧客を判断するために、1億人の顧客のうちの誰が、多数のオファーのそれぞれに反応するかを予測する必要があるとします。十分な処理能力と適切な予測モデルがあれば、アナリストは解約率だけでなく、解約の背後にある理由も見抜くことができます。ある電気通信会社は、OML4Sparkを使用して、支払い記録や通話プラン、サービス履歴を調査し、顧客データベース内の類似性と傾向を発見することで、より多くの情報に基づき、意味のある決定を下していました。OML4Sparkのおかげで、大規模なHadoopクラスターでバッチ・ジョブをパラレル実行できました。

オラクルは、HDFS、Apache Hive、Apache Impala、Spark DataFrames、Oracle Database、JDBCソース、およびローカル・ファイルに格納されたデータに対して、R環境からこのような演算を実行する選択肢を提供しています。

まとめ：R for the Enterprise

ほとんどの組織は、エンタープライズレベルの厳格な制御を使用して情報をセキュアに保存するために、データベースに依存しています。オラクルは、Rを大規模な機械学習、分析ツール、ビッグ・データへの取組みとともに使用できるようにしました。開発者は、使い慣れたR環境をOracle DatabaseやHadoop、分析ツールとともに使用でき、卓越したスケラビリティとパフォーマンスをビッグ・データ課題に役立てることができます。

ファイル抽出に慣れたアナリストは、データベース中心アーキテクチャを取り入れることで、デスクトップのR実装からデータベースにデータを送信し、デスクトップのモデルを使用してOracle Database内のデータを処理できます。

RとSQL間の透過性のおかげで、アナリストはOracleデータベース、Apache Hive、またはApache Impala内のデータにRを直接使用できるため、効率性が向上します。Rユーザーは、インデータベースSQL分析機能、データ・マイニング機能、およびオープン・ソースのRパッケージをOracle Databaseと組み合わせて使用することで、タスクの平行実行を実現できます。

Oracle Machine Learningは、R環境を離れずに使用できます。CRANパッケージや作成済みの他のR資産を利用して、SQL文からRスクリプトを呼び出すことで、Rベースの分析機能を本番ダッシュボードや本番アプリケーションにデプロイできます。ビッグ・データの問題では、OML4Sparkが持つスケーラビリティを、さまざまなHadoopクラスタや、Oracle Big Data Applianceとその最適化されたHadoopクラスタとともに活用できます。

Oracle Advanced Analyticsオプション、Big Data Connectors、あるいはOracle Linuxを購入するお客様は、Oracle R Distributionに対するエンタープライズ・クラスのサポートを受けることができます。

要約すると、Oracle DatabaseやHadoopとともに使用できるようにRを拡張すると、これまでとは逆に、分析機能をデータ側に持ち込むことができます。Oracle Machine Learningを使用してOracle DatabaseにR機能を送信し、OML4Sparkを使用してHadoopクラスタ・ノード上でRからSparkジョブを呼び出すことで、アナリストはデータ移動を最小限に抑え、生データをすぐに使用可能な情報に変換する際の待ち時間を短縮できます。広く使われているこれら3つの環境を統合すれば、コスト効率に優れた強力なビッグ・データ分析ソリューションが実現します。



Rのエンタープライズ対応

2019年7月

Oracle Corporation World

Headquarters

500 Oracle Parkway

Redwood Shores, CA 94065

U.S.A.

海外からの問い合わせ窓口：

電話: +1.650.506.7000

Fax: +1.650.506.7200

oracle.com



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

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。0513

Hardware and Software, Engineered to Work Together