

Oracle Advanced Analyticsによる ビッグ・データ分析

ビッグ・データと分析をシンプルに



Oracleホワイト・ペーパー | 2015年7月



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。以下の事項は、マテリアルやコード、機能を提供することをコミットメント（確約）するものではないため、購買決定を行う際の判断材料になさらないでください。オラクル社の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。

目次

免責事項.....	2
概要：Oracle Advanced Analyticsによるビッグ・データ分析	1
ビッグ・データと分析 — 新しい機会と新しい課題.....	4
予測分析.....	4
データではなく、アルゴリズムを移動	6
SQLとRのサポート	6
Oracle Advanced Analyticsのデータベース内処理.....	8
自動的なデータ準備、データ型、スター・スキーマ、「ネストした表」	9
Oracle SQL Developerの拡張機能、Oracle Data Miner Workflow GUI	10
Oracle R Enterprise — オープン・ソースRとOracle Databaseの統合	11
Hadoop、Oracle Big Data Appliance、およびOracle Big Data SQL.....	13
社内全体で使う予測分析アプリケーションを開発するためのプラットフォーム	14
結論.....	16

基本的に、すべてのモデルが誤っているが、いくつかは役に立つ」



George Box

20世紀の有名な統計学者

概要：Oracle Advanced Analyticsによるビッグ・データ分析

「ビッグ・データ」と「クラウド」の時代が到来し、企業の変革が進んでいます。後れを取らないためには、新しいスキルを習得し、これらの新しいデータソースとテクノロジーを利用する新しいプラクティスを実装する必要があります。顧客とのコミュニケーションを改善し、製品への知覚価値を強化する見返りとして、企業が顧客の Digital Exhaust(機器ログ等)を共有しているため、顧客の期待がますます高まり、それが企業の推進力となっています。ビッグ・データと分析は、こうした新しい要件を満たす有望なツールです。クラウド、競合他社、ビッグ・データ分析、次世代型の「予測」アプリケーションの存在により、企業は「実践的知見」を改善して、高い成果を上げるという新たな目標の達成に向かって突き進んでいます。従来のBIおよび分析アプローチでは、そうした詳細な予測知見は生成されず、ビッグ・データとクラウドが創造した新世界秩序の新たな顧客の期待にはまったく対応できません。

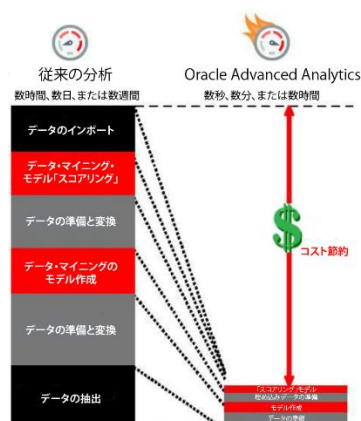
残念ながらビッグ・データの場合、データの3つのV (Velocity：速度、Volume：量、Variety：データ型) が拡大、増加すると、新しい問題が生じます。データ量が増えると、データの管理と移動ができなくなります。スケーラビリティ、セキュリティ、情報の待機時間が新たな問題になります。非構造化データ、センサー・データ、空間データを処理すれば、これらすべてのデータが新しいデータ型の複雑さをもたらします。

従来型の高度な分析の場合、データ抽出とデータの移動、データの重複によるデータソースの不整合、データ・セキュリティ上の弱点といった短所が複数の情報テクノロジーに内在しています。これらの欠点は、関係するデータ・アナリスト/サイエンティストのスキル、複数の分析ツール（業務用とオープン・ソース）、言語（SAS、R、SQL、Python、SPSSなど）に応じて、個別に何度でも現れます。データ分析と情報管理の二者が衝突すると、デプロイメント・フェーズで問題がひどく悪化します。

従来のデータ分析は通常、個々の分析サーバー、統計学者やデータ・サイエンティストによるデータ分析向けに設計された分析ツール（SAS、R、Python、SPSSなど）へエクスポートされたデータの代表サンプルデータまたはサブセットから開始されます。専門家が実施する分析は、シンプルな記述(descriptive)統計分析から、高度な予測(predictive)分析および処方的(prescriptive)分析までさまざまです。

データ・サイエンティストが、便利で役立つようにと意図した予測モデルを作成したら、ITが関与してデプロイメントについて理解する必要があり、エンタープライズ・デプロイメントとアプリケーション統合の問題が次の大きな課題になります。予測モデルを適用し、データウェアハウス内にある大規模データセットに関する予測を行うには、予測モデル、および関連データの準備と変換の全ステップをSQLに変換し、データベース内で再現する必要があります。このモデル・デプロイメント・フェーズには、元の統計言語（SAS、R、Python）からSQLへの、時間とコストのかかる面倒な手動コード作成ステップがあります。DBAとITは何らかの方法で、データベースやデータウェアハウス内のこれらの個別の統計モデルを「生産可能」な状態にして、全社に配布する必要があります。一部のベンダーは、予測モデルのデプロイメントのみを実行する専用の製品とオプションの料金を請求します。多くの高度な分析プロジェクトはここで失敗します。Hadoop、センサー・データ、ツイート、拡大し続けるビッグ・データ・リザーバ、および「データを実践的な知見にするまでの」全体的なプロセスを加えることはさらに難しい課題となります。

オラクルではそのようなことはありません。オラクルは、従来の抽出、移動、ロード、分析、エクスポート、移動/ロードのパラダイムを排除するビッグ・データおよび分析プラットフォームを提供します。Oracle Database 12cおよびOracle Advanced Analyticsオプションの場合、ビッグ・データの管理と分析は最初からデータ管理プラットフォーム内に設計されています。オラクルは数十年にわたって、業界最先端のデータ管理プラットフォーム、Oracle SQL、Oracle Big Data SQL、Oracle Exadata、Oracle Big Data Appliance、およびオープン・ソースRとの統合への研究開発に投資してきました。それらがシームレスに組み合わせられ、単一のプラットフォームであるOracle Databaseに結集しました。



オラクルは、ビッグ・データとクラウドの時代のビッグ・データおよび分析プラットフォームで以下を実現することをビジョンとして掲げています。

- ・ ビッグ・データと分析をシンプルにする（あらゆるコンピュータ・インフラストラクチャ上であらゆるデータのサイズと種類、組み合わせに対応）
- ・ ビッグ・データと分析のデプロイをシンプルにする（サービス型、プラットフォーム型、アプリケーション型）

Oracle Advanced Analyticsは、データを移動する必要性を排除し、ビッグ・データ管理とビッグ・データ分析を組み合わせます。



Oracle Advanced Analyticsが提供する強力なデータベース内アルゴリズムの幅広いライブラリ、オープン・ソースRとの統合を併用すれば、広範なビジネス課題を解決できます。また、ライブラリにはSQL、R、またはGUIでアクセスできます。Oracle Database Enterprise Edition 12cのオプションであるOracle Advanced Analyticsはデータベースを社内全体の分析プラットフォームに拡張することで、解約予測、顧客のセグメント化、不正と異常の検出、クロスセルとアップセルの機会の特定、マーケットバスケット分析、テキスト・マイニング、センチメント分析など、データ駆動手法による問題に対応します。データ・アナリスト、データ・サイエンティスト、ビジネス・アナリストは、Oracle Advanced Analyticsを使用することで、より正確な知識の獲得、新たな知見の発見、情報に基づく予測が可能になり、Oracle Databaseの大量データを直接処理できます。

データ・アナリスト/サイエンティストは、Oracle Advanced Analyticsの使用方法を柔軟に選択できます。Oracle Data Minerは、データ・アナリスト向けに設計されたOracle SQL Developerの拡張機能で、Oracle Advanced Analytics SQLのデータ・マイニング関数（Oracle Data Mining）への使いやすい「ドラッグ・アンド・ドロップ」ワークフローGUIが提供されます。Oracle SQL Developerは、従来の環境とクラウド環境の両方でOracle Databaseの開発と管理を簡素化する無償の統合開発環境です。Oracle Data Minerユーザーは、自分の分析手法が気に入ったら、そのワークフローを他のアナリストと共有したり、SQLスクリプトを生成してDBAに渡すことでモデル・デプロイメントを加速化したりすることができます。Oracle Data MinerはPL/SQL APIも搭載しているため、ワークフローのスケジューリングと自動化も実行できます。

Rプログラマーとデータ・サイエンティストは、馴染みのあるオープン・ソースRの統計プログラミング言語コンソールのRStudio、あるいは自分で選んだIDEを使って、データベース内のデータを直接操作し、データベースとOracle Advanced AnalyticsのRとの統合（Oracle R Enterprise）を活用できます。Oracle Advanced AnalyticsのOracle R Enterprise（ORE）は、SQLとRの変換を透過的に実行して、SQLとOracle Data Mining（ODM）の関数と同等の関数を提供し、データベース内のパフォーマンス、並列処理、スケーラビリティを実現します。これにより、企業でいつでもRを利用できるようになります。

ODM SQLデータ・マイニング関数とORE R統合を使うアプリケーション開発者は、データベースの強みとRの柔軟性を活用する完全自動の予測分析ソリューションを構築して、Oracle Advanced Analyticsと分析ソリューションをBIダッシュボードとエンタープライズ・アプリケーションに統合できます。

ビッグ・データ管理とビッグ・データ分析を同じ強力なOracle Database 12cデータ管理プラットフォームに統合することで、オラクルはデータの移動を排除し、総所有コストを軽減し、予測分析ソリューションとアプリケーションを全社に最速の方法で届けることができます。

ビッグ・データと分析 — 新しい機会と新しい課題

Gartnerはビッグ・データの特性について、「意思決定、ビジネスの知見、プロセスの最適化を強化するために、新しい革新的な形の処理を必要とする大量、高速、多様な情報資産」と定義しています。ただし、多くの人にとってこのようなデータは別に新しいものではありません。企業は何年にもわたって、大量のデータのデータ・マイニングを行ってきました。では、何が新しく大変なのかというと、「ビッグ・データ」の量、速度、ソースの種類がますます速く増え続けていること、新しい「実践的知見」で達成可能なことに対して顧客から新しい期待が寄せられていることです。その結果、情報テクノロジー (IT) 部門、データ・サイエンティスト、データ・アナリスト、彼らがサポートする各種業務部門（マーケティング、カスタマー・サービス、サポート、R&D、オペレーションズなど）に対して、新しい要望が生まれます。

残念ながら、ビッグ・データの3つのV（Velocity：速度、Volume：量、Variety：データ型）が長期にわたって拡大、増加していくと、新しい問題が生じます。データ量が増えるにつれ、最終的にはほとんどデータを動かさなくなり、結局、ある時点になると、大量のデータを個別のサーバーに移動してデータ分析を実行することが困難になります。ビッグ・データが爆発的に増える中、データの移動、データの重複、セキュリティ、「データ分析のスプロールマート（不規則な増大）」の発生、データ分析とデータ管理の分離など、多くの問題が生じており、さらに悪いことに、情報の待機時間が数日、数週間に延びることがほとんどです。

従来のデータ分析方法がこれらの問題の原因となっています。データ・アナリストとデータ・サイエンティストには通常、使い方を習得した自身の特別な「ツール」（SAS、R、SPSS、Pythonなど）があります。そのため、データベース/データウェアハウスからデータを抽出し、変換して、個別かつ専用の分析サーバーへロードすることが必要になります。データ専門家が「優れた」予測モデルを作成しても、新しい問題が浮上します。そのモデルをもっとも必要な箇所に適宜デプロイすること、BIダッシュボード、コール・センター、Webサイト、ATM、モバイル機器などのアプリケーションに統合することが、ITにとって次の大きな課題になります。大規模データ表で予測を実行するには、予測モデル、および関連データの準備と変換の全ステップをターゲット・プラットフォーム上で再現する必要があります。Oracle環境の場合、このエクスポート、データ分析、インポート結果を外部ループに対して行くと、データ分析が不必要に複雑化し、モデル・デプロイメント・フェーズが時間とコストのかかる作業となってしまいます。ITは、モデルを「生産可能」にして、SQLを使ってデータベース内で再実装する必要があります。

問題なのは、モデルが元々、統計プログラミング言語（SAS、R、SPSS、Python）を使って作成されており、そのモデルを生産可能にするには、データベース内でSQL関数として機能させる必要があることです。この段階で時間が大量に費やされ、エラーが起こる可能性があります。効率的なデータ収集、データ管理、分析、予測モデルのデプロイメント、知見、実践的なビジネス・インテリジェンスがリーダーを目指す組織にとっての成功の鍵となります。従来のデータ分析方法では、まったく不十分です。Hadoop、センサー・データ、ツイート、増える一方のデータ・リザーバを追加しても、問題全体が悪化するだけです。

予測分析

予測分析は、大量のデータを自動的にふるいにかけることで、それまで隠れていたパターンを見つけ、貴重な新しい知見を発見し、次のようなデータ駆動手法による問題について情報に基づく予測を行うプロセスです。

- 顧客行動の予測、クロスセルとアップセルの機会の特定
- 顧客離れ、従業員の離職率、学生保持率の予測
- 異常検出、税、医療、または費用に関わる潜在的な不正への対策
- 隠れた顧客セグメントの理解、顧客感情の理解

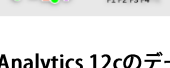
- 成果を上げるおもな要因の特定、品質向上の実現

テクノロジーとしての予測分析は長年にわたって、非常に重要な価値を発揮してきました。予測分析はGartnerの『Hype Cycle for Emerging Technologies』で取り上げられるまでになり、2013年には、Gartnerが定義する「Plateau of Productivity（安定期）」に達しました。2015年の今日、予測分析は企業やアプリケーションで実装、デプロイされており、その範囲は顧客離れと従業員の離職率の予測から、医療詐欺や脱税の警告、ターゲットを絞った販売戦略やリアルタイムのリコメンデーション・エンジンまでさまざまです。ビッグ・データ分析テクノロジーとユーザーの採用が進化、成熟、発展するに伴い、「可能性の技術」を推し進める予測分析のユースケースおよび統合された「予測」アプリケーションが毎日新たに登場して、ユーザーの新たな期待レベルを常に引き上げています。

Oracle Advanced Analyticsは、リレーショナル・データベース（RDBMS）に実装されてきた幅広い強力な主力データ・マイニング・アルゴリズムを提供して、これらのデータ駆動手法の問題に対応します。アルゴリズムはSQL関数としてデータベース内に実装されています。したがって、Oracle Advanced Analyticsのデータ・マイニング・アルゴリズムは、関連するSQL機能をすべて利用し、標準的な構造化表とビュー、トランザクション・データと集計、非構造化データ型であるCLOBデータ型（Oracle Textを使って「トークン」を解析）、空間データなど、元のスター・スキーマ表現のままデータをマイニングできます。Oracle Advanced Analyticsのデータベース内SQLデータ・マイニング関数は、データベース内の並列処理を利用してモデル構築とモデル適用の両方を実行し、セキュリティとユーザーのすべての権限のスキームを遵守し、リビジョン管理と監査追跡のデータベース機能に従います。また、暗号化された可能性のあるネイティブ形式のまま、データをOracle Database内でマイニングできます。



OOAのデータベース内データ・マイニング・アルゴリズム - SQL、R、GUIアクセス

関数		アルゴリズム	応用
分類		ロジスティック回帰 (GLM) ディビジョン・ツリー Naive Bayes サポート・ベクター・マシン	従来型の統計手法 人気度/ルール/透明度 埋め込みアプリ 範囲が広い/狭いデータ/テキスト
回帰		線形回帰 (GLM) サポート・ベクター・マシン (SVM)	伝統的な統計手法 範囲が広い/狭いデータ/テキスト
異常検出		1クラスSVM	未知の不正ケースまたは異常
属性の重要度		最小記述長 (MDL) 主成分分析 (PLC)	属性縮約、データ・ノイズの低減
相関ルール		アプリアリ	マーケットバスケット分析/改善策の提案
クラスタリング		階層型K-Means 階層型Oクラスタ化 期待値最大化 (EM) クラスタリング	製品分類/テキスト・マイニング 遺伝子およびタンパク質の分析
特徴抽出		非負行列因子分解 (NMF) 特異値分解 (SVD)	テキスト分析/特徴削減

Oracle Advanced Analytics 12cのデータ・マイニング関数は、SQL関数として実装され、SQL、PL/SQL、R、Oracle Data Miner GUIでアクセスできます。

データではなく、アルゴリズムを移動

データは大きいですが、アルゴリズムはそうではありません。したがって、データをアルゴリズムに移動するよりも、アルゴリズムをデータに移動する方が理にかなっています。オラクルは1999年、Thinking Machines Corporationのデータ・マイニング・テクノロジーおよび開発チームを買収したときに、このビッグ・データと分析データの課題に気付きました。当時、オラクルは、SQL言語を完全にサポートしたネイティブSQL関数として、従来および最先端の機械学習アルゴリズムと統計関数を開発するという戦略に着手していました。

Oracle Advanced Analyticsは、データ・マイニング・アルゴリズムをPL/SQLのスクリプト、コールアウト、または拡張フレームワークのアドインとしてではなく、ネイティブSQL関数として実行します。マイニングのモデルは、構築、適用、共有、監査が可能なファーストクラス・データベース・オブジェクトです。

2000年代の初め、Oracle Data Mining Release 9.2iから始まったオラクル初のデータ・マイニング・アルゴリズムは、利用可能なデータベースの中核的な強み、具体的には、集計、並列処理、スケーラビリティ、その他のデータベース・アーキテクチャの土台を駆使しました。オラクルの最初の2つのデータ・マイニング・アルゴリズム、Naive BayesアルゴリズムとAprioriアルゴリズムは基本的に、計数の原理に基づいています。これらはあらゆるものを非常に速く計算してから、条件付き確率予測モデルを組み立てます。さらに、100%すべてをデータベース内で実行します。データ、予測モデル、あるいは結果がデータベースから離れることは絶対にありません。

OAAのNaive Bayesアルゴリズムは、予測モデルを迅速に構築して、「どの顧客が離れそうか」、「どの顧客が製品Aを買う可能性がもっとも高いか」、あるいは「ある項目が失敗する確率はどのくらいか」といったことを予測できます。分かりやすくするために、もう少し詳しい例を取り上げます。たとえば、製品A（オートバイ、500ドルの靴など）を販売したいとします。Oracle Advanced Analyticsのデータ・マイニング・アルゴリズム、具体的にはNaive Bayesアルゴリズムは、製品Aを購入した全顧客のうち、男性と女性が何人ずつかを計算します。

アパートを借りている顧客と持ち家所有の顧客は何人か。子供がいる顧客は何人で、子供の数は何人か。これらの答えそれぞれを計算して合算すると、複雑な条件付き確率モデルになって、製品Aをもっと売る確率を高めるためには誰に的を絞ればいいのかを正確に予測できます。

OAAのApriori「マーケットバスケット分析」アルゴリズムは各顧客の取引「バスケット」の項目を計算しながら、AとBが頻繁に出現するなどといった同時発生する項目を探し、条件付き確率ARルールをアウトプットします。たとえば、次のようになります。

IF: 「シリアル」と「バナナ」が同じ顧客のバスケットに入っている場合（論理和）、
THEN: 「牛乳」もバスケットに入れられる確率が高い。
WITH: 確率 = 87%、および支持 = 11%

Oracle Advanced Analyticsのこうした新しいタイプの顧客知見で備えれば、店舗はシリアルとバナナの横に牛乳を置くかどうか、プロモーションとして新しい「朝食キット」製品群を提供するかどうか、あるいはその顧客に合ったお勧め商品をレジの支払い時にリアルタイムで提示するかを決定できます。ここでは、ビッグ・データ分析によってデータから「実践的知見」を引き出すさまざまな手法の中から、ごく簡単な例を紹介しました。当然のことながら、より多くのデータ、より高度な分析手法と迅速な社内全体のデプロイメントにより、多くの新しいビッグ・データおよび分析のアプリケーションとソリューションの可能性へと新たな扉を開くことができます。

SQLとRのサポート

SQLがデータ管理の標準言語として、40年以上にわたってデータ分析に利用されてきた一方で、R、SAS、Python、



SQLなどさまざまな言語が競い合っています。SAS、S+、SQL、SPSS、Matlabは長年愛用されてきましたが、この何年かは、特にオープン・ソースのRがトップに躍り出ており、Pythonなども新たに登場しました。KDD Nuggetsのデータ・マイニング業界コミュニティの年次投票 (<http://www.kdnuggets.com/polls/>) によると、RとSQLが現在、それぞれ1位と2位を争っています。

幸運なことに、Oracle Advanced Analyticsでは、SQLとRの両方がサポートされています。オラクルは、データ管理用のSQLを知っている多数の開発者のために、Oracle Advanced AnalyticsのSQLデータ・マイニング関数を介してデータ・マイニングと高度な分析をサポートし、オープン・ソースのR統計プログラミング言語との業界最先端の緊密な統合を行います。

ほとんどのオラクルのお客様は、構造化データの問合せ、レポート作成、分析に使う言語としてSQLにかなり精通しています。ほとんどのBIツールの基盤となる分析とテクノロジーの事実上の標準です。

Rは広く人気を集めている統計分析用のオープン・ソース・プログラミング言語で、無料のため、ほとんどのデータ科学教育プログラムで教えられています。最初にRの使い方を学ぶデータ・アナリスト、データ・サイエンティスト、研究者、学者たちはますます増えており、SQL言語またはR言語のいずれかを使ってOracle Database内でデータを操作できるRプログラマー集団の数が増えています。

Oracle Advanced Analyticsは過去15年以上にわたる成熟を遂げ、今ではOracle 12c内で開発されています。Oracle Advanced Analytics Optionは、パラレル化したスケーラブルな20近くのデータベース内実装の主力予測分析アルゴリズムを提供します。Oracle Advanced Analyticsはこれらのデータ・マイニング・アルゴリズムを、SQL、R言語、Oracle Data Miner GUIを使ってアクセスできるSQL関数として公開します。Oracle Data Miner GUIは、Oracle SQL Developerの拡張機能で、クラスタリング、回帰、予測、相関分析などのもっとも一般的なデータ駆動手法の問題に対応します。Oracle Advanced Analyticsのアルゴリズムはすべて、データベース内に深く実装されており、Oracle Databaseの業界最先端のスケーラビリティ、セキュリティ、SQL関数、統合、ETL、クラウド、構造化/非構造化/空間データ型の機能と強みを活用します。これらのアルゴリズムには、SQLとR、GUIの両方でアクセスできます。

したがって、Oracle Advanced Analyticsは次のように考えることができます。

従来のSQL

- 「手動型」の問合せ
- ドメインに関する専門知識
- すべての「ルール」を定義し、管理しなければならない

SQL問合せ

- SELECT
- DISTINCT
- AGGREGATE
- WHERE
- AND OR
- GROUP BY
- ORDER BY
- RANK



+

Oracle Advanced Analytics (SQLとR)

- 自動的な知識の発見、モデルの構築と展開
- マイニング/分析する「適切な」データを組み立てるためのドメインに関する専門知識

分析SQLの「動詞」

- PREDICT
- DETECT
- CLUSTER
- CLASSIFY
- REGRESS
- PROFILE
- IDENTIFY FACTORS
- ASSOCIATE



Oracle Advanced AnalyticsはSQLを拡張して、予測、検出、相関付け、クラスタ化など、強力な分析動詞を追加します。

Oracle Advanced Analyticsのデータベース内処理

Oracle Advanced Analyticsはデータベースを包括的で高度な分析プラットフォームに拡張して、ビッグ・データ分析に対応します。オラクルでは、強力な分析がデータベース内のデータ上で直接実行されます。結果、知見、リアルタイムの予測モデルをデータベースで利用、管理できます。

データ・マイニング・モデルはPL/SQL APIで構築された、データベース内のスキーマ・オブジェクトで、データを準備し、隠れたパターンを学習してOAAモデルを構築します。構築したOAAモデルは組込みのOAAデータ・マイニングSQL関数を使ってスコアリングできます。モデルを構築する際、Oracle Advanced Analyticsは既存のスケラブルなテクノロジー（パラレル実行、ビットマップ索引、集計の手法など）、新たに開発された他のOracle Advanced AnalyticsとOracle Databaseのテクノロジー（並列インフラストラクチャ内の再帰、IEEE浮動小数点、自動データ準備で対応するビンニング、欠損値の処理、テキストなどの非構造化データのサポートなど）を利用します。

データ・マイニング関数をSQL関数としてデータベース内に埋め込むことの真の力は、データ・マイニング・モデルのスコアリングでもっとも大きく発揮されます。

履歴データの隠れたパターンを学習することでモデルがいったん構築されると、データベース内の新しいデータにモデルを適用する作業は驚くほど高速に実行されます。スコアリングは、単なる行型の関数にすぎなくなります。Oracle Advanced Analyticsはこうして、数百万のレコードを秒単位で「スコアリング」でき、オンライン・トランザクション処理（OLTP）環境をサポートするように設計されています。

Oracle Exadataの「スマート・スキャン」テクノロジーを使用すると、さらに効果が上がります。Oracle Advanced AnalyticsをOracle Exadata環境で稼働すると、SQLの述部とOAA予測モデルはハードウェア・レイヤーにプッシュダウンされて実行されます。

- Oracle Exadata環境の場合、Oracle Exadataストレージ・レベルにプッシュされて実行されます。
- Oracle Big Data Appliance (Hadoop) 環境の場合は、BDAストレージ・レベルにプッシュされて実行されます。

いずれの場合も、述部を満たすレコードのみがディスクからプルされて、データベース内でさらに処理されます。以下の例ではたとえば、離れていきそうなアメリカの顧客を見つけるとします。

```
select cust_id
from customers
where region = 'US'
and prediction_probability(churnmod,'Y' using *) > 0.8;
```

Oracle ExadataまたはBDA上で実行されたスコアリング関数

自動的なデータ準備、データ型、スター・スキーマ、「ネストした表」

通常、アナリストが適切なデータ分析を実行するには、データ値の区分け方法（ビンニング）、欠損値への対処方法を明示的に決める必要があります。ほとんどの場合、モデルで使用する変数の値（特徴の選択）を減らす必要があります。Oracle Advanced Analyticsは15年以上にわたる進化を遂げた今、データ・マイニング・プロジェクトで必要とされることの多いほとんどのステップ実行を自動化できます。今日、自動データ準備（ADP）は、デフォルトのビンニング戦略とユーザーのカスタマイズが可能なビンニング戦略（等幅、同数、ユーザー定義および同様のビン・カテゴリ属性など）を使って、数値属性をN上位値、「その他の」またはユーザー定義のビンに自動的にビンニングします。欠損値がある場合は通常、そのレコードを分析から削除せずに、統計値（平均値、中央値、最頻値など）に自動的に置き換えます。ADPはモデル構築で使用され、モデルを新しいデータに適用するときに再び使用されます。ユーザーはもちろん、そうしたければADP設定をオーバーライドできます。

Oracle Advanced Analyticsは、属性縮約（最小記述長(MDL)アルゴリズムを使った属性の重要性）と特徴抽出（主成分分析とNon-Negative Matrix Factorization）の手法をサポートします。ただし、Oracle Advanced Analyticsのアルゴリズムそれぞれ（デシジョン・ツリー、一般化線形回帰、サポート・ベクター・マシン、Naive Bayes、K-Means クラスタリング、期待値最大化クラスタリング、異常検出1クラスSVMなど）には、属性縮約と選択に対応するための独自の自動戦略が組み込まれているため、明示的な変数縮約のステップ実行は任意選択で、必須ではありません。ユーザーはもちろん、アルゴリズムとデータ準備の設定を制御することも、インテリジェント機能のデフォルト設定をそのまま受け入れることもできます。

購入、トランザクション、イベントなどのトランザクション・データは、優れた予測モデルを構築する上で重要なデータの多くを代表します。Oracle Advanced Analyticsはこのデータをネイティブのトランザクション形式でマイニングし、データベースの集計関数を使ってまとめてから、データ（項目購入など）のベクトルをフィードし、他の顧客2Dデータに結合して、360度の顧客ビューを提供します。分類、回帰、クラスタリングなどのOracle Advanced Analyticsのモデルは、集約されたこのトランザクション属性を「ネストした表」として取り込みます。

Oracle Advanced Analyticsのデータベース内処理の奥深くでは、レコードがUnique_ID、Attribute_name、およびAttribute_valueの3つの組として処理されます。これこそが、Oracle Advanced AnalyticsがOracle Databaseの中核的な強みを利用する上で秘訣となる部分です。マーケットバスケット分析は当然、このデータをネイティブのトランザクション・データ形式（通常は集計された形）でマイニングして、バスケットに同時に出現する項目を見つけます。

非構造化データ、つまり、テキストも同様に、データベース内で処理されます。Oracle Advanced AnalyticsはOracle Textの処理機能とマルチ言語サポートを使って、テキスト、Word、Adobe AcrobatなどのCLOBデータ型を「トークン化」します。Oracle TextはすべてのOracle Databaseに無料で搭載されている機能なので、Oracle Advanced Analyticsはこの機能を利用して、非構造化データを前処理してから、語や語係数（TFIDF— Term Frequency Inverse Document Frequency）のベクトルをアルゴリズムにフィードします。Oracle Advanced Analyticsは、非構造化属性を単に、警察の所見、医師の注記、履歴書、電子メール、記事、抜粋などの追加の入力属性として処理します。これらの属性は、Oracle Advanced Analyticsのデータ・マイニング・アルゴリズムにフィードされている他のあらゆる属性（年齢、収入、職業など）と結合されます。空間データ、Webクリック、他のデータ型も、Oracle Advanced Analyticsのデータ・マイニング・モデル内で結合して含めることができます。

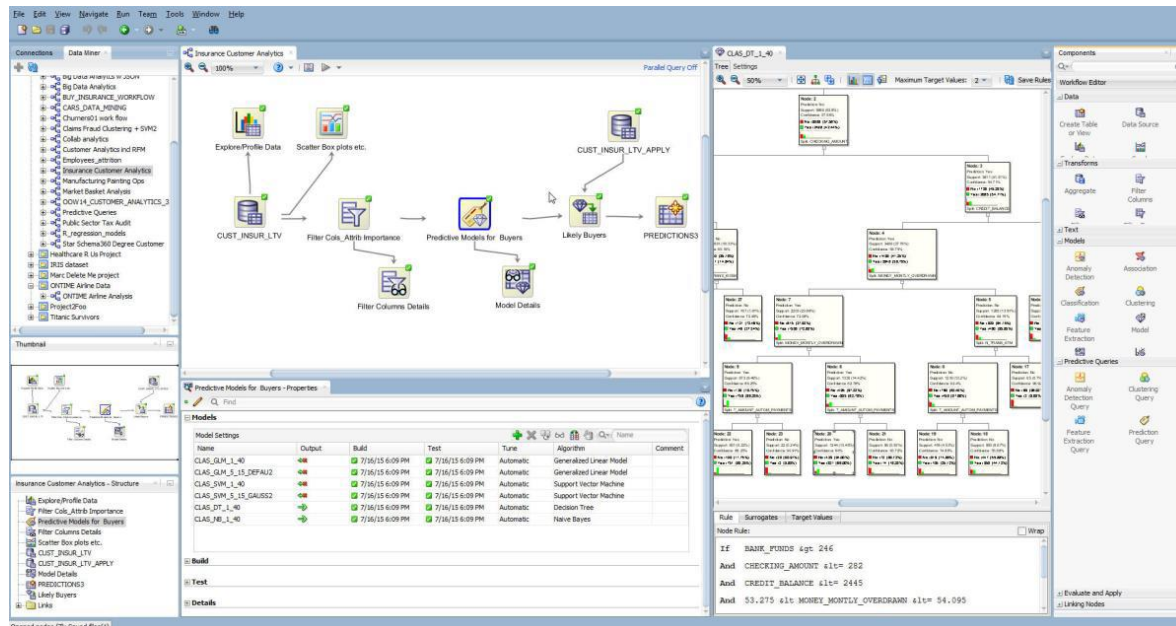
Oracle SQL Developerの拡張機能、Oracle Data Miner Workflow GUI

Oracle SQL Developer 4.1の拡張機能、Oracle Data Miner GUIは、使いやすいデータ分析のGUIを好み、必ずしもSQLやRでプログラミングを行う方法を習得する必要がないか、コードの作成が面倒というユーザー向けに設計されています。Oracle Data Minerにより、データ・アナリスト、ビジネス・アナリスト、データ・サイエンティストはOracle Data Minerのグラフィカルな「ドラッグ・アンド・ドロップ」ワークフロー・パラダイムを使って、データベース内のデータを直接操作できます。

データ・アナリストはOracle Data Minerの使い方を簡単に習得して、データをグラフィカルに素早く視覚化、探索し、データを必要に応じて準備、変換して、広範なモデル表示機能とモデル評価ビューアを使って複数のデータ・マイニング・モデルを構築、評価できます。その後、Oracle Data Miningモデルを新しいデータに適用してデプロイしたり、Oracle Data Miningの予測モデルを全社にデプロイするSQLやPL/SQLのスクリプトを生成したりすることができます。

Oracle Data Minerのワークフローは、ユーザーの分析手法をキャプチャして文書化します。これらのワークフローは保存して他のユーザーと共有することで、高度な分析手法を自動化、公開できます。

データ・アナリストの作業が完了すると、Oracle Data Minerは、SQLスクリプトを生成してDBAに渡します。DBAは、Oracle Databaseを使って即座にスクリプトをデプロイし、データ管理と予測分析の組み合わせに対応できます。アプリケーション開発者は、Oracle Data Miner PL/SQLワークフローAPIを使ってプログラムでワークフローを呼び出すことで、新しいビジネス・インテリジェンスと実践的知見の発見と配布を完全に自動化し、予測手法をアプリケーションに統合して全社で幅広く使用できます。



Oracle SQL Developerの拡張機能、Oracle Data Minerのデータ・アナリスト向けドラッグ・アンド・ドロップ・ワークフローのユーザー・インターフェースを使用すると、データを探索し、予測モデルを構築、評価、適用し、高度な分析手法をSQLスクリプトやPL/SQLスクリプトとしてデプロイできます。

データ・アナリストはOracle Data Minerを使って、非常にシンプルな分析手法から複雑で高度な分析手法まで、さまざまな手法を試して組み立てることができます。たとえば、トランザクション・データ、人口統計データ、カスタマー・サービス・データ、顧客の意見を組み合わせ、360度の顧客ビューを構築するとします。そして顧

客に対してクラスタリングを実行し、顧客セグメントに顧客を事前に割り当ててから、セグメントごとに個別の分類、回帰、または異常検出のモデルを構築することで、精度と有用性を高めることを決定します。



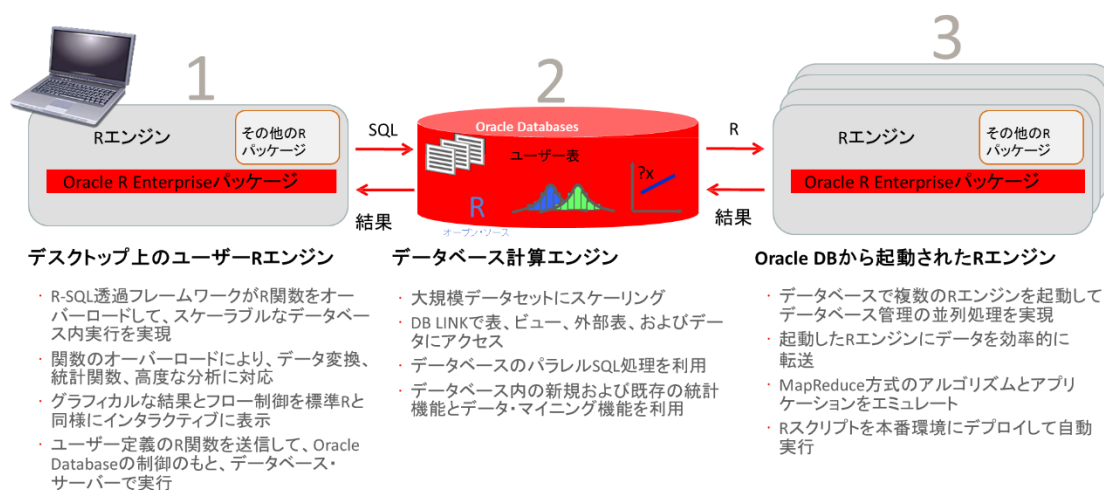
データ・アナリストは、シンプルな分析手法から高度な分析手法まで、さまざまな手法を素早く構築して、Oracle Database内でアクセスできるデータをマイニングできます。すべてのデータ、モデル、結果はずっとデータベース内にあります。

Oracle R Enterprise — オープン・ソースRとOracle Databaseの統合

Oracle R EnterpriseはOracle Advanced Analyticsオプションのコンポーネントの1つで、オープン・ソースの統計プログラミング言語であるRとその環境を、エンタープライズ対応およびビッグ・データ対応にする機能を備えています。「Rは幅広い統計（線形モデルと非線形モデル、従来の統計テスト、時系列分析、分類、クラスタリング）手法とグラフィカル手法を提供し、拡張性に優れている」（<https://www.r-project.org/>を参照）。Rの強みは無料であることです。強力に拡張性に優れたオープン・ソースであるRには、グラフィカルな統計パッケージが多数あり、R「パッケージ」を作成、提供しているRユーザー・コミュニティによって常に増え続けています。Rの課題には、メモリに制約があること、シングルスレッドであること、外部ループを実行して処理の速度低下を招きかねないことが挙げられ、一般に「業務仕様」とは見なされていません。提供されているRパッケージの質はさまざまです。

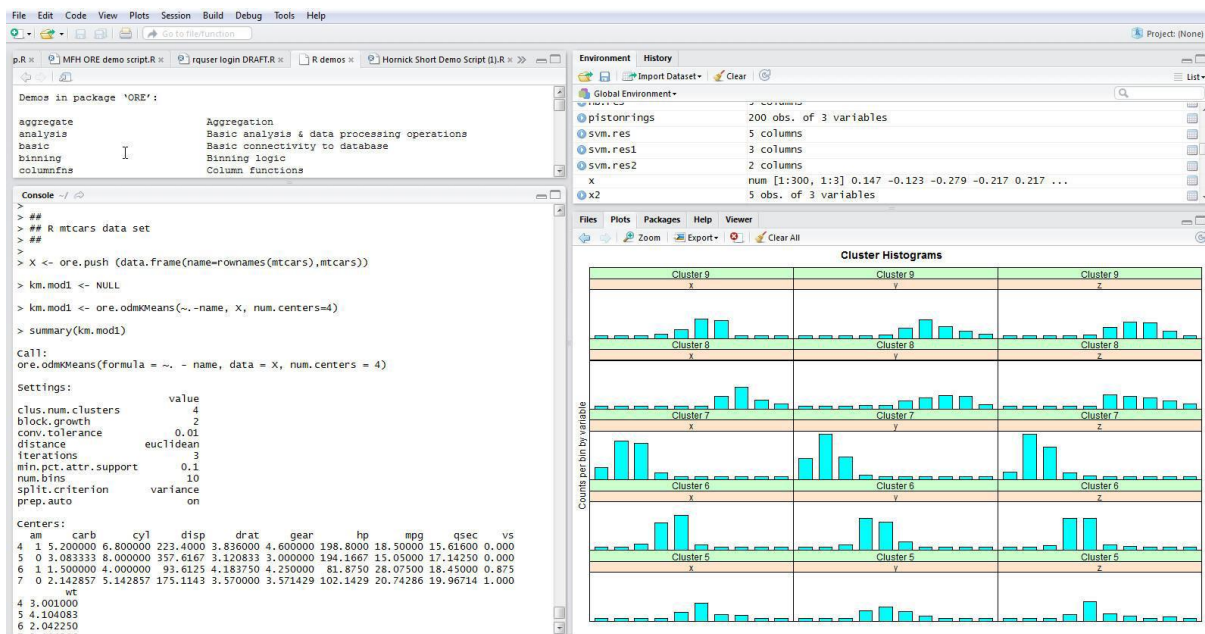
Oracle R EnterpriseはRをOracle Databaseに統合し、R関数をSQL関数とOracle Data Mining SQL関数にマッピングします。また、大量のデータに伴う問題を解決するように設計されています。Oracle R Enterpriseは一連のRパッケージ（ORE）とOracle Database機能で構成されており、Rユーザーはこれらを使って、SQLを使用せずにデータベース内のデータを操作し、データベース・サーバー上で稼働する1つ以上の埋め込みRエンジンでRスクリプトを実行できます。データ・アナリストとデータ・サイエンティストはSQLを習得しなくても、データベースとSQLデータ・マイニング関数の並列処理とスケーラビリティを利用するRスクリプトを開発、完了、デプロイすることで、データ分析を一度に自動化できます。

Oracle R Enterpriseは、オーバーロードされたオープン・ソースRのメソッドと関数を搭載して、標準的なR構文をSQLに透過的に変換します。これらのメソッドと関数は、Oracle R Enterpriseの透過層を実装するORE/パッケージ内にあります。Rプログラマーはこれらの関数とメソッドにより、データベース内にあるデータへのアクセス、分析、操作を行うRオブジェクトを作成できます。データベースはSQLコードを自動的に最適化して、問合せの効率性を改善します。Oracle R Enterpriseは関数のプッシュダウンを実行することで、base R、Oracle SQLの統計関数、Oracle Data Mining SQL関数、および一部の人気の高いR/パッケージをデータベース内で実行します。Oracle R EnterpriseはOracle Databaseの組み込みコンポーネントとして動作するため、Rエンジンに提供されるデータをデータベースで管理しながら、関数のプッシュダウンまたは「埋め込みRモード」を使用して任意のR/パッケージを実行できます。この「埋め込みRモード」の機能により、開発者はOracle Advanced Analyticsのネイティブにサポートされたツールキットをどのオープン・ソースR/パッケージとでも拡張し、完全にデータベースで管理される幅広い高度な自動分析手法を開発できます。



Oracle Advanced AnalyticsのOracle R Enterpriseのコンポーネントは、R関数をデータベース内の同等のSQL関数に透過的にプッシュダウンして、スケーラビリティと並列処理を実現します。また、OREユーザーは「埋め込みRモード」を介してどのR/パッケージも利用できます。

Rで作業してデータのアクセスと分析をしたいユーザーの場合は、RStudioまたは任意のR GUIを使ってOracle Databaseに接続し、Oracle Advanced AnalyticsのR統合 (Oracle R Enterprise) にアクセスできます。接続が確立されると、OAA/OREセッションがユーザーのメタデータと同期するため、データベース内のすべての表とビューが表示されます。基本R言語の関数を実行すると、関数は同等のSQL関数に透過的にマッピングされます。OAA/ODMアルゴリズムとOAA/OREアルゴリズムを使用するRユーザーは、スケーラブルなデータ・マイニングをデータベース内で実行できます。



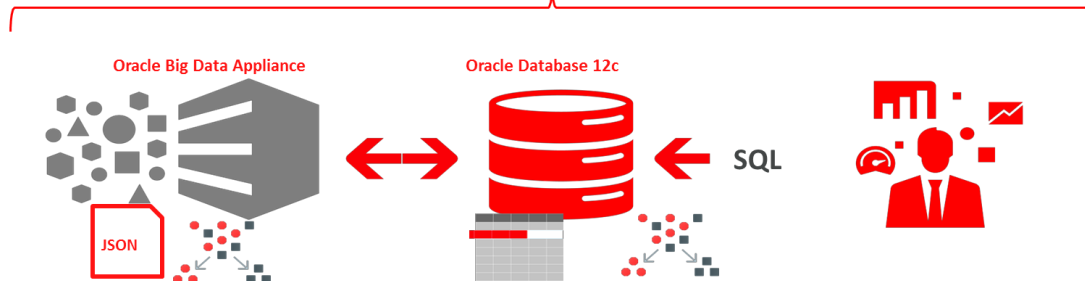
RStudioコンソールからデータベース内ODMアルゴリズム (サポート・ベクター・マシンなど) を起動するOracle Advanced AnalyticsのOracle R Enterpriseコンポーネント

Hadoop、Oracle Big Data Appliance、およびOracle Big Data SQL

ビッグ・データは現在、Hadoopサーバーに保存されていることがほとんどです。データベース外部にある個別データ環境によって、新しいデータ管理とデータ分析の課題が生じています。Oracle Big Data SQLは、Oracle Big Data Applianceを介してSQL処理をHadoopに拡張してこの課題に対処します。Oracle Big Data SQLは、Oracle Exadata用に開発された「スマート・スキャン」テクノロジーを使って、SQLロジックをHive表で動作させるためにプッシュダウンします。データ・アナリストは、ビッグ・データ・リザーバに保存された、未発掘かもしれない有益なデータの新しいビッグ・データソースを簡単に利用して、データベースやデータウェアハウス内で管理されている未知数のデータとそのデータを組み合わせることが可能になりました。

ただし、Hadoopに保存されたデータは膨大な量で、スパース表現（トランザクション形式）になっていて、情報密度が低い可能性があります。データの大半が、センサー、モノのインターネット、「ツイート」、他の大量の情報源に由来する場合、ユーザーはOracle Big Data SQLを利用して、総数、最大値、最小値、ユーザー定義値を上回ったまたは下回ったしきい値の数、平均値、短期の平均値と総数、長期の平均値と総数、スライディングSQLウィンドウの平均値と総数および相互の比較値を収集できます。収集したら、Oracle Big Data SQLを使って「ビッグ・データ」をフィルタリング、縮小して、他のデータベースのデータに結合し、Oracle Advanced Analytics Optionを使ってOracle Database内の「あらゆるもの」をマイニングします。

SQL

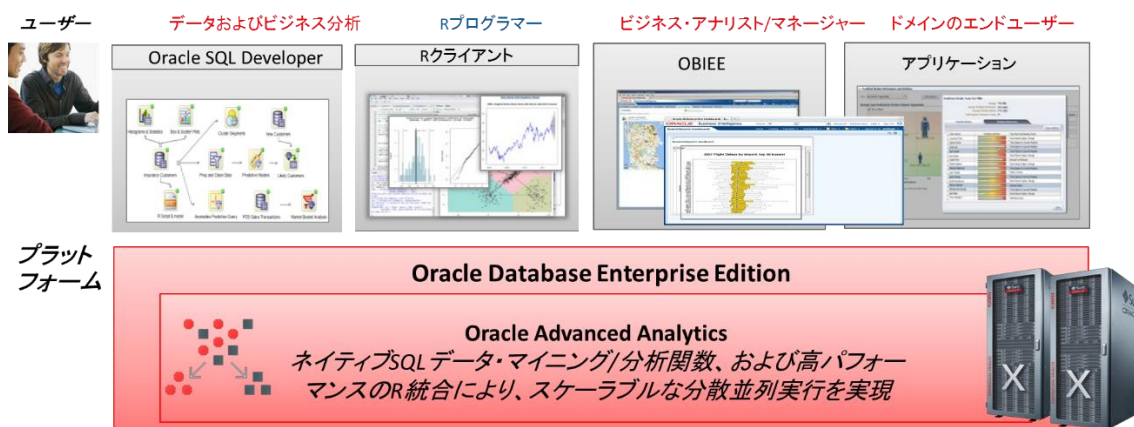


変換されていないJSONデータをHadoopに保存 ビジネスに不可欠なデータをOracleに保存 SQLまたはRで分析したデータ

SQLおよびOracle Big Data SQLを使用すると、データ・アナリストは、Hadoopサーバーとデータベース双方の集計データのアクセス、集約、フィルタリングを実行して組み合わせることで、より完全な360度の顧客ビューを作成し、Oracle Advanced Analyticsを使って予測モデルを構築できます。

社内全体で使う予測分析アプリケーションを開発するためのプラットフォーム

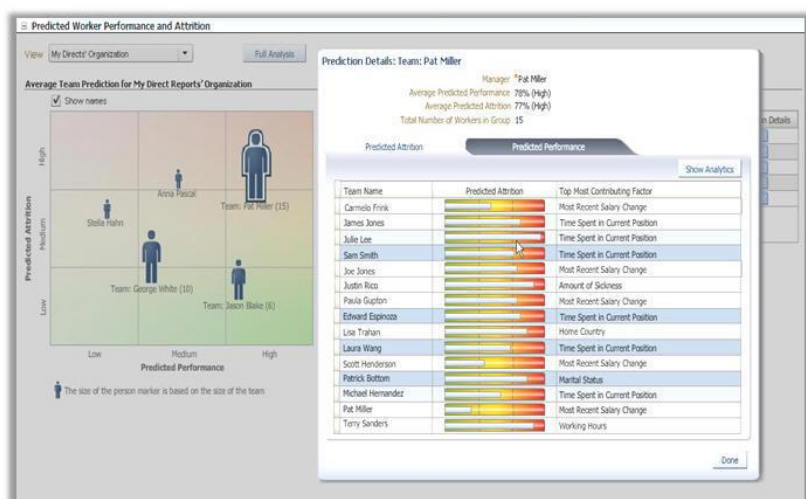
ビッグ・データとビッグ・データ分析をシンプルにするオラクルの戦略は、予測分析アプリケーションの開発、改良、デプロイメントをより簡単にします。すべてがデータベースの機能の一部になっているためです。すべてのデータ、ユーザー・アクセス、セキュリティと暗号化、スケーラビリティ、アプリケーション開発環境、強力な高度な分析をデータ管理およびデータ分析プラットフォーム、Oracle Databaseで使用できます。Oracle DatabaseにSQLで問い合わせることができるあらゆるエンタープライズ・アプリケーション、BIダッシュボードまたはツールに予測知見とリアルタイムの実践的知見を容易に追加できるようになっています。



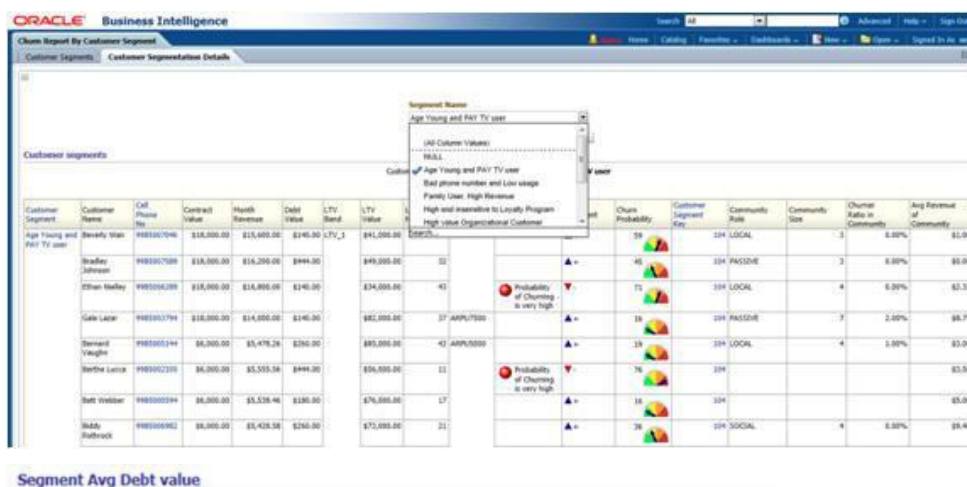
Oracle Databaseの独立したライセンス機能である、Oracle Advanced Analyticsは、予測分析アプリケーションの開発とデプロイメントに最適、強力な分析プラットフォームにデータベースを拡張して、ビッグ・データとビッグ・データ分析の両方に対応します。

オラクルは長年、予測分析アプリケーションを開発し続けており、以下の次世代予測アプリケーションを社内およびクラウドで提供します。

- Oracle Human Capital Management Predictive Workforce
- Oracle Customer Relationship Management Sales Prediction Engine
- Oracle Adaptive Access Management's Identity Management
- Oracle Retail Customer Analytics
- Oracle Predictive Incident Monitoring Premium Service
- Oracle Communications Industry Data Model
- Oracle Retail Industry Data Model
- Oracle Airlines Industry Data Model
- Oracle Utilities Industry Data Model



Oracle HCM Predictive Workforceアプリケーションは、事前構築されたOAA予測分析を実現して、従業員の離職率、従業員の業績、「What if?」分析に対応します。



Oracle Communications Data Modelは、構築済みのOAA予測モデルを埋め込んで、顧客離れの予測、顧客のプロファイリング、顧客離れの要因の特定、クロスセル、顧客生涯価値（LTV）、顧客感情に対応します。

結論

従来のITと分析のアプローチでは、「ビッグ・データ」と「クラウド」の時代の要件にとても対応しきれるものではありません。自らの業界でリーダーを目指す組織にとって、これらの新しいテクノロジー、有益なデータと未知数のデータの迅速なキャプチャと収集、適切なデータ管理、関連データのアセンブリ、容易な詳細分析、実践的知見の自動化とデプロイメントを利用することが、成功の鍵となります。

Oracle Database 12cの価値あるオプション、Oracle Advanced Analyticsは、今日あまりにも一般化し過ぎた従来の抽出、移動、ロード、分析、エクスポート、移動、ロード/インポートのパラダイムを打ち壊します。Oracle Advanced Analyticsは、パラレル化したスケーラブルなデータベース内実装の主力予測分析アルゴリズム（クラスタリング、回帰、予測、相関付け、テキスト・マイニング、相関分析、異常検出など）の幅広いライブラリをSQL関数として、Oracle Database 12c内に実現します。Oracle Advanced Analyticsは、SQL（Oracle Data Mining OAA SQL APIコンポーネント）、Oracle SQL Developer 4.1の拡張機能であるOracle Data Minerの「ドラッグ・アンド・ドロップ」ワークフローGUI、およびオープン・ソースR（Oracle R Enterprise のR統合コンポーネント）との統合を介してアクセスできる、これらの予測アルゴリズムをSQL関数として公開します。

Oracle Advanced Analyticsのデータベース内データ・マイニング機械学習/予測分析アルゴリズムは、Oracle Databaseを徹底活用して構築されていて、Oracle Databaseのスケーラビリティ、セキュリティ、統合、クラウド、構造化/非構造化データ・マイニング機能をフルに利用しているため、オラクルは、ビッグ・データと分析ソリューション、およびアプリケーションの最適なプラットフォームとなります。また、社内での使用またはOracle Cloud



での使用のいずれにも対応します。

オラクルのデータ管理、および記述(descriptive)、予測(predictive)、処方的(prescriptive)ビッグ・データ分析は最初からプラットフォーム内に設計されています。数十年に及ぶオラクル最先端のデータ管理、Oracle SQLとOracle Big Data SQLのすべてが利用され、「データをアルゴリズムに移動」ではなく、「アルゴリズムをデータに移動」するオラクルの設計および開発アプローチと組み合わさっています。オラクルでは、ビッグ・データとクラウドの時代に合ったビッグ・データおよび分析プラットフォームを構築して、以下を実現することをビジョンとして掲げています。

ビッグ・データと分析をシンプルにする：

- ・ あらゆるコンピュータ・インフラストラクチャでどのようなデータ・サイズにも対応
- ・ あらゆる種類のデータのどのような組み合わせにも対応

ビッグ・データと分析のデプロイメントをシンプルにする：

- ・ サービス、プラットフォーム、アプリケーションとして

ビッグ・データ管理とビッグ・データ分析を単一の統合Oracle Databaseプラットフォームに組み入れることで、オラクルは総所有コストを軽減し、データの移動を排除し、予測分析ソリューションとアプリケーションを全社に最速の方法で届けることができます。



CONNECT WITH US



blogs.oracle.com/oracle



facebook.com/oracle



twitter.com/oracle



oracle.com

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

海外からの問い合わせ窓口
電話 : +1.650.506.7000
ファクシミリ : +1.650.506.7200

Hardware and Software, Engineered to Work Together

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、記載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。

facebook.com/oracle 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.0115 の登録商標です。



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