

IBM SPSS Analytic Server
バージョン 3.2.1

概要



注記

本書および本書で紹介する製品をご使用になる前に、5 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM SPSS Analytic Server バージョン 3、リリース 2、モディフィケーション 1、および新しい版で明記されていない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： IBM SPSS Analytic Server
Version 3.2.1
Overview

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

概要	1	特記事項	5
アーキテクチャー	2	商標	7
Spark および Analytic Server	2		
3.2.1 の新機能	3		

概要

IBM® SPSS® Analytic Server は、IBM SPSS テクノロジーをビッグデータ・システムと組み合わせたビッグデータ分析ソリューションであり、使いやすい IBM SPSS ユーザー・インターフェースを使用して、以前は達成不可能であった規模で問題を解決できるようにします。

ビッグデータ分析が重要な理由

組織で収集されるデータ・ボリュームは飛躍的に増加しています。例えば、金融業や小売業では 1 年間 (または 2 年間や 10 年間) のすべての顧客取引があり、通信プロバイダーではコール・データ・レコード (CDR) やデバイス・センサーのデータがあり、インターネット会社では Web クロールの結果があります。

ビッグデータ分析は、以下が存在する場合に必要です。

- 特に構造化データと非構造化データの混合である場合の大量 (テラバイト単位、ペタバイト単位、エクサバイト単位) のデータ。
- 急速に変化/蓄積するデータ

また、ビッグデータ分析は、以下の場合に役立ちます。

- 多数 (数千) のモデルが構築されている
- モデルが頻繁に構築/更新される

課題

次のように多様な理由から、大量のデータを収集しているのと同じ組織で、そのデータを実際に利用するのが困難であることがよくあります。

- 従来型の分析製品のアーキテクチャーが分散型計算に適していない
- 既存の統計アルゴリズムが、ビッグデータを処理するように設計されていない (このようなアルゴリズムはデータがアルゴリズムに移動されることを予期しますが、ビッグデータの移動には高額なコストがかかります)。そのため、次のような問題が起こります。
- ビッグデータに対して最先端の分析を実行するには、新たなスキルとビッグデータ・システムについての徹底した知識が必要である。このようなスキルを持つ分析者は非常にわずかです。
- インメモリ・ソリューションは中規模の問題に対しては機能するが、本当のビッグデータに対応して十分に拡張することはできない。

解決策

Analytic Server は以下を提供します。

- Hadoop Map/Reduce などのビッグデータ・システムを HDFS 内のデータと共に活用するデータ中心型アーキテクチャー
- データに向かうよう設計された新しい統計アルゴリズムを取り込むための定義済みインターフェース
- 分析者がデータの分析に集中できるようにビッグデータ環境の詳細を非表示にする、使いやすい IBM SPSS ユーザー・インターフェース
- どの規模の問題にも対応して拡張できるソリューション

アーキテクチャー

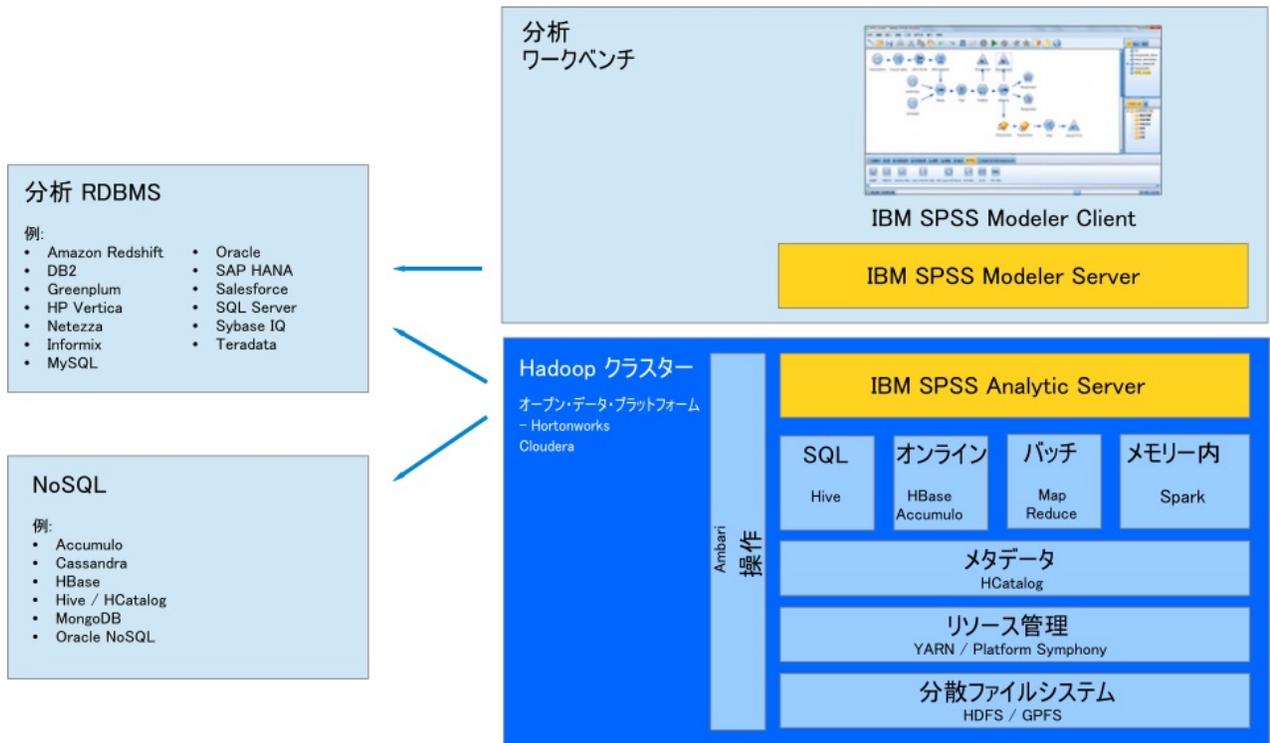


図 1. アーキテクチャー

Analytic Server はクライアント・アプリケーションと Hadoop クラウドの間に位置しています。データがクラウドにあることを想定して、Analytic Server での作業に関する概要は次のとおりです。

1. クラウドのデータに対して Analytic Server データ・ソースを定義します。
2. クライアント・アプリケーションで実行する分析を定義します。現行リリースでは、クライアント・アプリケーションは IBM SPSS Modeler です。
3. 分析の実行時に、クライアント・アプリケーションは Analytic Server 実行要求を実行依頼します。
4. Analytic Server は、Hadoop クラウドで実行されるジョブを調整して、結果をクライアント・アプリケーションに報告します。
5. 結果を使用して、詳細な分析を定義できます。このサイクルが繰り返されます。

Spark および Analytic Server

Analytic Server は Apache Spark と統合することによりパフォーマンスが向上します。

Spark が使用される場合と使用されない場合

Spark が Ambari サービスとして Hadoop クラスタにインストールされている場合、Analytic Server は Spark を使用してビッグデータ・ジョブを処理します。以下のガイドラインは、Spark が使用されない場合を判別するために使用します。

1. データ・セットが 128 MB よりも小さい場合、Analytic Server は Analytic Server JVM の組み込み MapReduce 機能を使用し、Spark も Hadoop クラスタも使用しません。

2. Spark がクラスターにインストールされていない場合、Analytic Server は MapReduce v2 を使用します。
3. Analytic Server は MapReduce v2 を使用して PSM モデルを作成します。ジョブが PSM モデル・ビルドで終了する場合、Analytic Server は Spark を使用してモデル・ビルドに至るまでのすべてのステップにわたるジョブを処理し、続いてディスクへの書き込みを行い、その後 MapReduce を使用して PSM モデルをビルドします。例えば、PSM モデル・ビルドが後に続く結合がジョブに含まれる場合、その結合は Spark で実行され、PSM は結合されたデータに対して MapReduce で実行されます。

Spark が使用される方法

Analytic Server サービスが開始し、Spark が使用可能であることを検出すると、このサービスは「Spark Hadoop ジョブ」を初期化します。このジョブは、クラスター全体の分散タスク間の通信を可能にします。このジョブは、Analytic Server サービスが実行されている限り実行され、Analytic Server のすべての実行に使用されます。この方法は、複数の MapReduce Hadoop ジョブを調整する方法よりもパフォーマンスが向上します。これは、各 Hadoop ジョブのすべての Analytic Server コンポーネントを再ロードするオーバーヘッドが除去されるためです。

Spark は MapReduce ジョブを実行できます。これにより、Analytic Server は、使用可能な場合は「ネイティブ」の Spark アルゴリズム (join、sort、union など) を使用することができます。それに加えて、Analytic Server は、Spark の既存の SPSS Map アルゴリズムおよび Reduce アルゴリズムを、Hadoop API を直接使用しないで実行できます。

3.2.1 の新機能

バージョン 3.2.1

プラットフォーム

Hadoop Data Platform (HDP) 3.0 および 3.1 のサポート。

Cloudera 6.0 および 6.1 のサポート。

Rank 関数

rank 関数は、入力データ・セットを別々のパーティションに分割し、各パーティション行の順位を示す新しいフィールドを生成するために使用します。この機能は、Hive 関数の **rank()**、**dense_rank()**、および **row_number()** に似ています。

UDF Hive プッシュバック

新しい Hive UDF 機能が導入されました。Hive UDF が HiveDB に登録されると、Analytic Server は新しい UDF 機能を使用してプッシュバックを実行できます。

最新のシステム要件情報については、IBM Technical Support サイトの <http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html> にある「Detailed system requirements」レポートを使用してください。このページでは以下を行うことができます。

1. 製品名として SPSS Analytic Server を入力して、「**Search**」をクリックします。
2. 目的のバージョンとレポートの範囲を選択して、「**Submit**」をクリックします。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。この資料は、IBM から他の言語でも提供されている可能性があります。ただし、これを入手するには、本製品または当該言語版製品を所有している必要がある場合があります。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの websites サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Director of Licensing
IBM Corporation

North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

記載されている性能データとお客様事例は、例として示す目的でのみ提供されています。実際の結果は特定の構成や稼働条件によって異なります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者をお願いします。

IBM の将来の方向性および指針に関する記述は、予告なく変更または撤回される場合があります。これらは目標および目的を提示するものにすぎません。

表示されている IBM の価格は IBM が小売り価格として提示しているもので、現行価格であり、通知なしに変更されるものです。卸価格は、異なる場合があります。

本書はプランニング目的としてのみ記述されています。記述内容は製品が使用可能になる前に変更になる場合があります。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する個人や企業が実在しているとしても、それは偶然にすぎません。

著作権使用許諾:

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する個人や企業が実在しているとしても、それは偶然にすぎません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物にも、次のように、著作権表示を入れていただく必要があります。

© IBM 2019. このコードの一部は、IBM Corp. のサンプル・プログラムから取られています。

© Copyright IBM Corp. 1989 - 2019. All rights reserved.

商標

IBM、IBM ロゴおよび [ibm.com](http://www.ibm.com) は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

Adobe、Adobe ロゴ、PostScript、PostScript ロゴは、Adobe Systems Incorporated の米国およびその他の国における登録商標または商標です。

IT Infrastructure Library は AXELOS Limited の登録商標です。

インテル、Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Centrino、Intel Centrino ロゴ、Celeron、Xeon、Intel SpeedStep、Itanium、および Pentium は、Intel Corporation または子会社の米国およびその他の国における商標または登録商標です。

Linux は、Linus Torvalds の米国およびその他の国における登録商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

ITIL は AXELOS Limited の登録商標です。

UNIX は The Open Group の米国およびその他の国における登録商標です。

Cell Broadband Engine は、Sony Computer Entertainment, Inc. の米国およびその他の国における商標であり、同社の許諾を受けて使用しています。

Linear Tape-Open、LTO、LTO ロゴ、Ultrium および Ultrium ロゴは、HP、IBM Corp. および Quantum の米国およびその他の国における商標です。



Printed in Japan

日本アイ・ビー・エム株式会社

〒103-8510 東京都中央区日本橋箱崎町19-21