

進化を続けるデータの エコシステムにおいて 情報を最大限に活用するには



InfoSphere Information Server V11.5 の新機能

特長

- 重要なプロジェクトと主要なアナリティクス・イニシアチブの成果を改善します
- ビジネス・イニシアチブと戦略イニシアチブを促進する情報の信頼性を高め、一貫性を維持し、リアルタイムで管理します
- データの理解、クレンジング、モニタリング、変換、および提供を改善します

はじめに

Hadoop のようなスケーラブルなデータ・プラットフォームは、他のプラットフォームを大きく凌駕するコスト・メリットと分析のチャンスをもたらします。IBM は Hadoop の拡張性と可能性を活かして、重要なプロジェクトと主要なアナリティクス・イニシアチブの成果を改善します。IBM® InfoSphere® Information Server V11.5 のエンドツーエンドの情報管理機能を使用すると、データをより正確に理解し、データのクレンジング、モニタリング、変換、および提供を行うことができます。また、IBM はコラボレーションの改善を通じて、ビジネス部門と IT 部門の間に存在するギャップを解消します。

InfoSphere Information Server の柔軟な統合機能を使用することによって、ビジネス・イニシアチブと戦略イニシアチブ (ビッグデータやインパクト・ポイントの分析からマスター・データの管理やデータウェアハウスまでを含む) を促進する情報の信頼性が高まり、一貫性が実現し、リアルタイムで管理できるようになります。

リリース以来、InfoSphere Information Server は小容量のデータから膨大な容量のデータまでをサポートする超並列処理 (MPP) プラットフォームとして、あらゆる複雑性を持つユーザーの要件に対応してきました。InfoSphere Information Server は、あらゆるサイズのプロジェクトを成功に導くために必要な柔軟性 (抽出・変換・ロード機能 [ETL] および抽出・ロード・変換機能 [ELT] により実現)、パフォーマンス、およびスケーラビリティをサポートする唯一のソリューションです。



| | | | |
|---|---|--|--|
| InfoSphere Information Server for Data Integration: <ul style="list-style-type: none">• Hadoop 上でのデータ統合• HDFS ファイルへの接続• 処理をリレーショナル・データベースにプッシュダウン可能• 製品に組み込まれたネットワークを移動中の機密データの保護機能として、業界最高レベルの InfoSphere Optim™ によるデータ・マスキングを活用 | InfoSphere Information Server for Data Quality: <ul style="list-style-type: none">• データの分類• パフォーマンスの強化• InfoSphere Information Analyzer• データ品質のクレンジングと精緻化 | InfoSphere Information Governance Catalog: <ul style="list-style-type: none">• ガバナンス・カタログの拡張可能なフレームワーク• データ・クラスの定義• 多言語対応• XML スキーマの定義のサポート• 資産交換機能が幅広いリネージュ・コンテンツに対応 | InfoSphere Information Server Enterprise Edition: <ul style="list-style-type: none">• Hadoop 上でのデータ統合• HDFS ファイルへの接続• 処理をリレーショナル・データベースにプッシュダウン可能• 製品に組み込まれたネットワークを移動中の機密データの保護機能として、業界最高レベルの InfoSphere Optim によるデータ・マスキングを活用• データの分類• パフォーマンスの強化• InfoSphere Information Analyzer• データのクレンジングと精緻化• ガバナンス・カタログの拡張可能なフレームワーク• データ・クラスの定義• 複数言語のサポート• XML スキーマの定義のサポート• 資産交換機能が幅広いリネージュ・コンテンツに対応 |
|---|---|--|--|

図 1. IBM InfoSphere Information Server V11.5 の 4 種類のパッケージが、幅広い課題に対応します。

InfoSphere Information Server V11.5 の機能は 4 種類のパッケージで提供され、情報に関する重要な課題に対応します (図 1)。

- **InfoSphere Information Server for Data Integration:** 柔軟にあらゆる形式にデータを変換し、あらゆるシステムに提供することで、価値創出までのリードタイムを短縮し、IT リスクを削減します。
- **InfoSphere Information Server for Data Quality:** 高品質なデータの構築と管理を行うことで、大量のデータを信頼性の高い情報に変えます。
- **InfoSphere Information Governance Catalog:** IT 部門とビジネス部門の間でデータをより正確に理解し、コラボレーションを改善することで、コミュニケーション・ギャップを改善し、情報ガバナンスのフレームワークを構築します。
- **InfoSphere Information Server Enterprise Edition:** 上記すべての 3 種類の個別のパッケージの機能が 1 つの包括的なオファリングにまとめられており、ある領域で情報統合プロジェクトを開始した後、必要に応じてプロジェクトを拡張し、結果を最適化できます。

InfoSphere Information Server のパッケージ

企業は自社のデータ・ニーズに最も緊密に合致するパッケージを選択することもできれば、一連の完全な機能を提供する包括的なエディションを選択することもできます (図 2)。

InfoSphere Information Server for Data Integration

柔軟なデータ統合機能には以下が含まれます。

- **Hadoop 上でのデータ統合:** データの接続、変換、および提供に関するすべての機能が、Hadoop クラスターのデータ・ノード上で実行できるようになりました。
- **HDFS ファイルへの接続:** このデータ統合パッケージは、Hadoop Distributed File System (HDFS) ファイルに対する拡張サポート (新規のデータ形式、追加の文字セット、および追加のデータ型のサポートを含む) を提供します。本パッケージは、すべてのサポート対象のオペレーティング・システム (Hadoop 環境と非 Hadoop 環境の両方に対応する Windows、IBM AIX®、IBM System z® 上で稼働する Linux、および Linux) で利用できます。

- **処理をリレーショナル・データベースにプッシュダウン可能:** IBM DataStage® Balanced Optimization の機能を使用すると、ユーザーは自動的にジョブ設計を SQL として再作成することによって、リレーショナル・データベース内でデータ統合ワークロードを実行できます。このような機能が、InfoSphere DataStage ライセンスとサポート対象プログラムとして InfoSphere DataStage を含む複数の製品に含まれるようになりました。また、Balanced Optimization の機能が拡張され、System z 上で稼働する IBM DB2® Analytics Accelerator の環境にも処理をプッシュダウンできるようになりました。
- **製品に組み込まれたネットワークを移動中の機密データの保護機能として、業界最高レベルの InfoSphere Optim™ によるデータ・マスキングを活用:** InfoSphere Information Server Enterprise Edition のライセンスに、InfoSphere DataStage の並列ジョブ内の機密情報を保護するための InfoSphere Optim のデータ・マスキング・ライブラリーが含まれるようになりました。

InfoSphere Information Server for Data Quality:

信頼性の高いデータ機能には以下が含まれます。

- **データの分類:** InfoSphere Information Server for Data Quality は個人情報 (PII)、機密データ、およびその他のクラスのデータを保存するロケーションを検出することによって、データ・プライバシー、データ・マスキング、およびテスト・データの管理のイニシアチブをサポートします。
 - クレジットカード、納税者 ID や電話番号などの 30 種類以上の事前に定義したデータ・クラスを使用する列に含まれるデータ型を識別することで、価値創出までのリードタイムを短縮します。
 - 拡張可能な分類を活用して、3 種類のデータ・クラス (有効な値の一覧、正規表現 [Regex]、および Java クラス) の作成とカスタマイズを行います。
 - 列の分析結果によって、データの値に基づくデータ分類を提案します。ユーザーはこのような分類を検証し、各列に保存されるデータを表記する最適なデータ・クラスを選択できます。
- **パフォーマンスの強化:**
 - データのプロファイリング (列の分析) 時にスケラビリティを高め、リソースの使用率を削減します。
 - 保存される個別の値を制限する最新のオプションによって、結果を保存するデータベース (Information Analyzer Database [IADB]) のサイズを大幅に縮小できます。
 - 複数ドメインと外部キーの分析のスケラビリティを改善します。最新のアルゴリズムでは、すべての固有値をキャプチャーする完全な列の分析は必要ありません。

- **InfoSphere Information Analyzer:** InfoSphere Information Analyzer 機能がサポートされるようになり、Hadoop クラスター内で直接実行できるようになりました (IADB をサポートする Oracle Connector の認証を獲得)。
- **データのクレンジングと精緻化:** データの調査、標準化、マッチング、選定、および住所の検証の機能がサポートされるようになり、Hadoop クラスター内で直接実行できるようになりました。InfoSphere US Address Certification および InfoSphere Address Verification Interface による住所のクレンジングと検証のサービスもサポートできるようになり、Hadoop クラスター上で稼働します。

InfoSphere Information Governance Catalog

コラボレーションを促進する情報ガバナンス機能には以下が含まれます。

- **ガバナンス・カタログの拡張可能なフレームワーク:** カタログ内で新規の資産の定義と登録を行い、表示と構造の定義を設定し、情報の認識と理解を改善できます。資産を完全に管理でき、データ・リネージュの管理と連携させることができます。
- **データ・クラスの定義:** InfoSphere Information Governance Catalog 内で検出したデータ分類の表示と検索を行います。管理者は、データ・クラスの定義 (正規表現、有効な値、または範囲を含む) の作成の管理を行うことができます。
- **複数言語のサポート:** 異なる言語でビジネス用語を定義し、用語同士を関連付けることによって、関連するビジネスの定義を管理できます。
- **XML スキーマの定義のサポート:** XML スキーマの定義ファイルをインポートし、含まれるエンティティ、属性、データ型を全社的に参照または共有できます。エンティティ、属性、データ型を管理し、注釈を付けることができます。
- **資産交換機能が幅広いリネージュ・コンテンツに対応:** 簡単かつシームレスにすべての拡張データ・ソースと拡張マッピング・コンテンツを InfoSphere Metadata Workbench の旧バージョンから移行し、InfoSphere Information Governance Catalog の複数バージョン間で移行できます。

InfoSphere Information Server Enterprise Edition

このオファリングは InfoSphere Information Server for Data Integration、InfoSphere Information Server for Data Quality、および InfoSphere Information Governance Catalog の機能を組み合わせることによって、柔軟性とスケーラビリティを拡張します。

InfoSphere Information Server の 共通のコンポーネントの拡張機能

共通のプラットフォーム・サービス (接続サービス、管理サービス、および実装サービスなど) を通じて、データの統合機能とガバナンス機能を提供します。InfoSphere Information Server V11.5 は、プラットフォームに関して以下の共通の拡張機能を提供します。

Apache Hadoop 内でエンジン層を実行

データの統合、クレンジング、およびプロファイリングを実行し、Hadoop クラスターのデータ・ノード上で分析ワークロードを実行します。Information Server エンジンを Hadoop クラスターのエッジ・ノードにインストールでき、3 種類のメカニズム (HDFS ディストリビューション、動的なランタイムのコピー、およびネットワーク・ファイル・システム [NFS]) を通じて、必要となるバイナリーをクラスター上のすべてのノードで即時に利用できるようにします。このようなモデルの中からユーザーのアーキテクチャーとセキュリティ要件に最適なモデルを選択してください。

新しいアプリケーション・マスター・コンポーネントが並列ジョブの実行リクエストを Hadoop YARN に送信することで、クラスターのデータ・ノード上でジョブの実行要件に適したリソースを確実に割り当てます。Hadoop のノード・ラベルのような機能を活用して処理を特定のノードでのみ実行し、YARN が適切なワークロードの管理を実現するためのキューのスケジューリングを行います。

その後、ジョブ内のステージがこれらのノード上のデータに割り当てられたデータ・パーティションで処理を行います (YARN によるリソース選択に基づいて、データは特定のデータ・ノードに送信されることがあります)。

InfoSphere Information Server エンジン は自動的にインメモリー機能 (データのパイプライン処理、データ・パーティショニング、および動的な再パーティショニングなど) を活用することによって、ジョブのランタイムのサイズを最小化します。

InfoSphere Information Server が提供するその他の機能には、以下が含まれます。

- Hadoop によるメタデータ管理
- コンテナ・サイズの見積もり
- Kerberos に対応したクラスターをサポート
- Operations Console での可視化
- YARN ジョブのブラウザー

InfoSphere Information Server V11.5 のパッケージ

| | | | |
|---|---|--|---|
| InfoSphere Information Server for Data Integration: 変換と提供 | InfoSphere Information Server for Data Quality: クレンジングと管理 | InfoSphere Information Governance Catalog: データの理解と コラボレーション | InfoSphere Information Server Enterprise Edition: 変換 データ提供 クレンジング 管理 理解 コラボレーション |
|---|---|--|---|

図 2. 企業は適切なパッケージを選択することもできれば、一連の完全な機能を提供する包括的なエディションを選択することもできます。

IBM アナリティクス

ソリューション・ブリーフ

対象となる最新の Hadoop のバージョン

過去のリリースと同様に、InfoSphere Information Server V11.5 は Open Data Platform (ODP) の要件を満たす Apache Hadoop ディストリビューション (IBM BigInsights®、HortonWorks、および Pivotal HD など) に加え、Cloudera のような 非 ODP ディストリビューションもサポートします。

BigInsights 4.x、HortonWorks 2.2 および 2.3、Cloudera 5.3 および 5.4 が Hadoop ディストリビューションをサポートし、Hadoop を使用しない InfoSphere Information Server の実装環境や Hadoop のエンジン層の実装環境からデータの読み書きを行います。

リポジトリとアプリケーション・サーバーの最新のサポート

リポジトリ層が Oracle 12g (RAC を含む) および SQL Server 2014 をサポートするようになり、サービス層が WAS 8.5.5 をサポートするようになりました。

最新のシステム連携機能

ファイル・コネクタの機能強化によって、Information Server に対応した OS をサポートできます。コネクタがメタデータの収集と列レベルのリネージュのために、IMAM もサポートするようになりました。これによって、Hadoop のメタデータのインポートができるようになりました。

このような最新のコネクタとステージは、以下のような新規のデータ・ソースと最新のデータベースをサポートします。

- Microsoft SQL Server 2014 用の SQL Server ステージ
- Sybase ASE 用の Sybase ステージおよび Sybase IQ、Version 16
- Teradata Database 15.10 用の Teradata Connector
- Netezza Performance Server 7.1 用の IBM Netezza® Connector
- WODM 8.7 用の IBM WebSphere® ILOG® JRules Connector

InfoSphere DataStage および InfoSphere QualityStage® のライセンスには、以下のエンタープライズ・アプリケーション・パックが含まれるため、ユーザーはこれをインストールおよび使用することで、インストール・パッケージ経由でこれらの製品への接続を設定できます。

- Salesforce.com
- Hyperion Essbase
- Siebel
- JD Edwards Enterprise One
- PeopleSoft Enterprise
- Oracle Applications

接続インフラを最適化するため、古いステージの提供を終了し、それに代わって新規のステージを提供します。

- Sybase BCP Load
- IBM DB2 API、DB2 Bulk Load、DB2 zLoad Enterprise Stage
- DRS
- Oracle OCI および Oracle OCI Bulk Load
- Redbrick
- Teradata API および Teradata Bulk Load

古いステージの代わりに提供されるステージの詳細情報については、

ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/com.ibm.swg.im.iis.ds.stageix.doc/topics/master_stagesbytype.html にアクセスして、アルファベット順に表示されるステージの一覧を参照してください。

コネクタ・マイグレーション・ツールを使用すると、最新のコネクタへの移行を自動化できます。本ツールはプロジェクトに接続し、古いプラグインとエンタープライズ・ステージを使用しているジョブを特定するため、ユーザーは最新のコネクタで自動的に更新すべきジョブを選択できます。

IBM を選択する理由

IBM Information Integration and Governance (IIG) は、ビッグデータによって生まれる課題に対応する業界最高レベルの機能を提供します。IIG は大容量のデータのスケラビリティとパフォーマンスを最適化し、加速するデータ速度に対して適切なサイズのデータの統合とガバナンスを俊敏に実現し、幅広いデータ型とビッグデータ・システムのサポートと保護を行います。

InfoSphere はビジネス・ユーザーが確実に洞察に基づいてアクションを実施できるようサポートを提供し、ビッグデータとアナリティクスのプロジェクトを成功に導きます。

詳細情報

IBM InfoSphere Information Server V11.5 に関する詳細情報が必要な場合は、IBM の営業担当員か IBM のビジネス・パートナーまでお問い合わせください。もしくは、ibm.com/software/products/en/infosphere-information-server にアクセスしてください。

さらに、IBM グローバル・ファイナンスは、できるかぎりコスト効率が良く戦略的な方法でビジネスに必要な IT ソリューションを入手するための支援を提供します。お客様のビジネス目標に適合し、効果的なキャッシュ・マネジメントを実現し、TCO が向上するよう、IBM は与信基準を満たしたお客様向けに IT ファイナンス・ソリューションをカスタマイズします。IBM グローバル・ファイナンスは、重要な IT 投資のための資金を獲得し、ビジネスをさらに前進させるための最もスマートな選択肢といえます。詳細情報については、ibm.com/financing の Web サイトをご参照ください。



© Copyright IBM Corporation 2015

IBM アナリティクス

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 番 21 号

Produced in Japan
2015 年 9 月

IBM、IBM ロゴ、ibm.com、AIX、BigInsights、DataStage、DB2、ILOG、InfoSphere、Optim、QualityStage、System z および WebSphere 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://ibm.com/legal/copytrade.shtml> をご覧ください。

Netezza は、IBM のグループ企業である IBM International Group B.V. の登録商標です。

Java およびすべての Java 関連の商標およびロゴは Oracle やその関連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、SQL Server および Windows は、Microsoft Corporation の米国およびその他の国における商標です。

本書の情報は最初の発行日の時点で得られるものであり、予告なしに変更される場合があります。すべての製品が、IBM が営業を行っているすべての国において利用可能なものではありません。本書に含まれる情報は、特定物として現存するままの状態を提供され、商品性の保証、特定目的適合性の保証、および第三者の権利の不侵害の保証を含むすべての明示もしくは黙示の保証責任なしで提供されています。IBM 製品は、IBM 所定の契約書の条項に基づき保証されます。



Please Recycle
