

IBM DataStage

AI向けの業務対応データを IBM Cloud Pak for Data DataStage
でリアルタイムに提供

データ統合で業務対応 データを提供

今日のデジタル・エンタープライズは、これまでなかった形でデータを生成、消費しています。これには複数のシステム、リポジトリに跨って保管されている消費者、取引、従業員に関するデータが含まれています。これらのデータ・ストアは様々なマルチクラウド、ハイブリッド・クラウド環境やデータ・レイクに分散されているので、組織はそれらの個別ソースと環境を合体させて、AIの力を借りて洞察をすばやく導き、区別され、パーソナライズされたエクスペリエンスを顧客に提供する方法を探しています。Forresterの調査によると、データ・サイエンティストは自分の時間の80%をAI活動向けのデータの準備と管理に費やしています。この結果と、組織の91%がデータを有効に利用していないというIBM調査を組み合わせると、企業はデータ・サイロから価値を得るのに苦労していることとなります。大量データでリアルタイム・データ・アクセス性を達成して、業務対応のデータを提供するのに用いられたアーキテクチャー技法、実践、ツールはデータ統合と呼ばれます。柔軟でスケーラブルなデータ統合テクノロジーで、複数データ・ソースのデータを抽出、変換、ロード(ETL)することで、企業は次善の提案、動揺検出と分析、サプライチェーン予測や瞬時の不正検出の分析を実施できます。

CxOや、エンタープライズ・アーキテクト、複数クラウドやデータ・レイクに跨るデータの管理に苦労していて、AIモデルやアプリケーションの構築、更新の時間短縮を望んでいる運用リーダーに、ETLを超えた信頼できる業務対応データを提供するIBM® InfoSphere™ DataStage、[マーケット屈指の\(要登録、英語\)](#)データ統合ソリューションがスケーラブルなマルチクラウド・データ統合、提供ソリューションで、信頼できる業務対応情報のリアルタイム利用を保証してくれます。DataStageの重要な機能には、一回デザインすれば、どのクラウドでも自動ワークロード・バランシングでワークロードをスケールしながら、実行できるマルチクラウド・ランタイム・サポートや、低レイテンシの平行・エンジンがあります。さらに、リアルタイム・データ提供機能と組み込みレプリケーション・テクノロジーもあり、DevOpsの継続的統合と継続的デリバリー(CI/CD)サポートの時間とコストを削減し、Autonomous Integration Designとインライン・データ品質を用いて、データの問題を自動的に検出し、解決する検証規則でAIモデルを構築する時間を短縮します。

DataStageはIBM DataOps機能の一部で、継続的な高品質データを運用可能にして、AIに対応し、どのデータ・ソースに対しても適切な人に適切なときに自動化されたセルフサービス・データ・パイプラインを提供します。IBM InfoSphere DataStageはオンプレミス、IBM Cloud、どこにでもデプロイできるIBM® Cloud Pak™ for Dataなどのハイパーコンバインド・プラットフォームで利用できます。IBM® Cloud Pak™ for Dataは完全に統合されたデータ、AIプラットフォームでRed Hat® OpenShift® ([英語、外部ページヘルプ](#))上に構築されていて、業務に合わせてスケールできるDataStageの完全クラウド・ネイティブ・アーキテクチャーを提供します。組織にデータ統合、データ・レプリケーション、データ仮想化を含む複数データ提供スタイルをサポートするプラットフォームも提供します。その一方でCDCはログベースの変更を発生度捕捉し、その情報をクラウド上のターゲット・データベースやデータ・レイクにKafkaベースのメッセージ・キューを利用して伝えます。



いちどデザインしたら、どのクラウドでも動作

IDCの調査(英語、外部ページヘルリンク)によれば90%の企業顧客が複数クラウドを利用しています。マルチクラウド・データ統合では、ユーザーはデザインを実行時と分離できます。ETLジョブをひとたびデザインしたら、ランタイム・コンポーネントをコンテナで、どこでもいいのでクラウド環境にデプロイして、大量データ処理によるレイテンシを減らせます。オンプレミスでジョブを作成、テストしてから、Microsoft Azureインスタンスなどのクラウド環境で実行して、クラウド上のAzureデータ・レイクを利用できます。ジョブ・パラメータとその値はDataStageのリモート・インスタンスにKafkaメッセージとして渡されます。

マルチクラウド・データ統合には以下のメリットがあります:

- オンプレミスとクラウド環境間でのデータ統合機能
- デザイン・プロセスを単純化する自動ジョブ・デザイン・エクスペリンス
- データ移出のコストを最小化するリモート・ジョブ実行
- 地政学要件の履行
- データを移動する必要がなく、今の場所に置いておけるので、大量データセット処理のレイテンシ削減



自動ワークロード・バランシングとパラレル・プロセッシング

完全なクラウド・ネイティブ・アーキテクチャーなので、DataStageのローカル・コンテナや共有コンテナを利用して、動的にワークロードをスケール・アウトできますし、[ベスト・イン・ブリードのパラレル・エンジン\(PX\) \(英語\)](#)で大きいデータセットに最適化されています。ユーザーはパラレル・シーケンス、またはApache Sparkジョブ作成をIBM DataStage Flow Designerで選択できます。

DataStage Flow Designerジョブは2つのランタイム・エンジンで実行できます:

- ジョブ種別がパラレル、またはシーケンスのジョブはパラレル・エンジンでしか実行できません。一般的にリソース集約型のジョブはパラレル・エンジンで実行されます。結果として、パラレル・プロセッシングを用いた複雑なジョブを完了する時間は平均で2分です。
- ジョブ種別がSparkのジョブはSparkエンジンでしか実行できません。



リアルタイム・データ提供

DataStageとコンテナとしてデプロイされた、リアルタイム・キャプチャー用の完全組み込みChange Data Capture (CDC)テクノロジーはデータ統合とデータ・レプリケーションの世界の両方で最善のものを提供できません。DataStageは大量データセットの複雑な変換を可能にし、その一方でCDCはログベースの変更を発生都度捕捉して、複雑な変換を用いて変換して、クラウド上のターゲット・データベースやデータ・レイクにKafkaベースのメッセージ・キューを用いて提供します。DataStageでは、バッチベースやイベントベースのバルク・データ変換ジョブでデータ・ウェアハウスに供給することも可能です。



DevOpsの時間とコストをCI/CDサポートで削減

さまざまなアプリケーションに跨がって、複数のコンテナ化されたアプリケーションを管理する難問に対応するために、組織は[Cloud Pak for Data](#)で利用できる[Red Hat OpenShift \(英語、外部ページヘルリンク\)](#)などの強固なオープン・ソース・ツールを必要としています。Cloud Pak for Dataプラットフォームはコンテナのスケール、プロビジョンを支援して、マイクロサービスなどの重要なIT活動やクラウド移行戦略を支援します。DataStageコンテナは継続的統合/継続的デリバリー(CI/CD)パイプラインのジョブの開発からテスト、稼働までの作成、自動化を可能にします。GitHubなどのソース管理ツールを高頻度の公開ジョブ、リリース、本稼働までサポートします。



自律統合デザインでAIに燃料を供給

資産を自動的に発見し、分類して内蔵のカスタム・トランスフォーメーション、品質ルールに基づいて、統合フローを生成し、機密情報を検出し、それを保護することで、AI向けのデータの収集、統合をより速いレート、より大規模に加速します。



自動ジョブ・デザインで価値実現までの時間を高速化

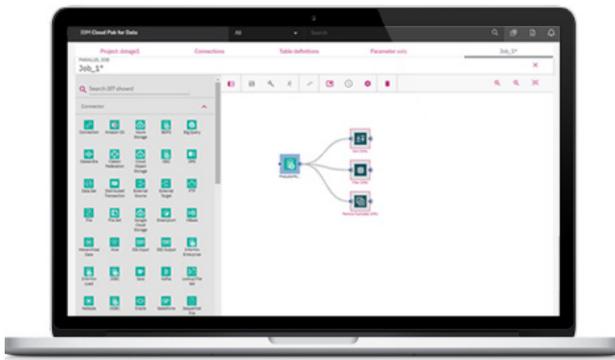


図 1 DataStage Flow Designerと自動デザイン機能

IBM DataStage Flow DesignerはDataStage用のWebベースUIで、機械学習(ML)機能で、技術者以外のユーザーに対してもジョブ内のフローとステージの構築を支援します。

DataStage Flow Designerには以下のメリットがあります:

- 後方互換性。ジョブ移行の必要なし。多くの企業ではひとつのプロジェクトの中に数千のジョブがあり、彼らはこれらのジョブが1日24時間週7日間動作することを前提としています。移行はエラーや停止の可能性が高いので、選択肢にはなりません。こうした企業では既存のDataStageジョブをIBM DataStage Flow Designerに渡すことができます。そうすればこれらのジョブを新たな場所に移行する必要はありません。
- 開発者生産性の向上。IBM DataStage Flow Designerには内蔵検索、企業が急発進するためのクイック・ツアー、自動メタデータ伝搬、スマート・パレット、ステージ提案、すべてのコンパイル・エラーの同時ハイライトなどの機能があります。開発者はこれらの機能を利用して、生産性を上げながら、ジョブをデザインできます。彼らの生産性は従来のジョブのハンド・コーディングに比べて最大9倍になります。
- 徹底的なオペレーターと接続性。デザイン、開発機能の他にDataStageは数百の事前構築されたそのまま利用できるオペレーターを提供しています。これによって開発者が分析アクションに必要なデータを準備するのに費やす時間が劇的に減少します。数週間ごとに新オペレーターが追加されているので、開発者の生産性は時間が経過するほど向上します。



信頼できるデータ提供のための、実行中データ品質とセキュリティ

DataStageはDataStage Flow Designerを利用したデータ統合用に単一ユーザー・エクスペリエンスを提供し、データがデータレイクなどのターゲット環境に提供される瞬間の実行中データの検証、標準化、マッチング規則で、権限を持たないユーザーが機密データにアクセスすることで起きる可能性のあるセキュリティ問題を防ぎます。このデータ品質のコンセプトは、データウェアハウス(DWH)全体の包括的なデータ・ガバナンスのサポートにも拡張できます。

まとめ

DataStageが提供する機能:

- いちどデザインしたら、内蔵の自動ワークロード・パラランシング、パラリズム、スケーラビリティによってどこでも実行
- リアルタイムに更新を捕捉、またはバッチ・ベースの提供スタイル
- 内蔵レジリエンス、容易な運用とCI/CD
- AI向けに最適化されたデータ統合
- ML機能を利用した自動ジョブ・デザイン
- 信頼できるデータ提供のための実行中データ品質とデータ・セキュリティ

IBMは、ハイブリッド・マルチクラウド環境、オンプレミス、IBM Cloud Pak for Dataのようなハイパーコンバージド・システム、または精選のクラウド・プラットフォームに跨がって、広範なデータ統合機能を提供しています。これらの異なった機能は柔軟でスケーラブルなデータ統合ソリューションを提供するので、自社で選択したデプロイメント・モデル上のAI向け高品質データにすばやくアクセスできます。

無償デモを体験して [IBM InfoSphere DataStage](#)の詳細を把握する(要登録)

IBMをお勧めする理由

IBMのDataOps機能は業務対応の分析基盤作成を支援して、AI対応の

自動化、導入されたガバナンス、強力なナレッジ・カタログと協調するマーケット屈指のテクノロジーを提供することで、業務全体にわたる高品質のデータを継続的に運用できるようにします。データ品質を向上して、効率的なセルフサービス・データ・パイプラインを適切な人に適切なときにあらゆるソースで提供します。

DataOpsの詳細情報は

ibm.com/dataopsをご覧ください

IBM InfoSphere DataStageの詳細情報については

ibm.com/jp-ja/products/infosphere-datastageをご覧ください

ビッグ・データとアナリティクス・ハブについては

ibmbigdatahub.com (英語) をご覧ください



© Copyright IBM Corporation 2020

IBM Corporation

New Orchard Road, Armonk, NY 10504

Produced in the United States of America

April 2020

IBM、IBM ロゴ、**ibm.com**、IBM Cloud Pak、DataStage、および InfoSphereは、世界の多くの国で登録された International Business Machines Corp.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、www.ibm.com/legal/copytrade.shtml の「著作権および商標情報」をご覧ください。

Red HatおよびOpenShiftは、Red Hat, Inc.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。

Microsoft 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。

本資料の情報は最初の発行日の時点で最新であり、予告なしに変更される場合があります。すべてのサービスが IBM の操業国すべてにおいて提供されるとは限りません。

本資料の掲載情報は現存するままの状態を提供され、第三者の権利の不侵害の保証、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。