

ストリーミング/IoTデータを使ったリアルタイムな業務分析

メリット

- ・ストリーミングアナリティクスでリアルタイムのオペレーショナルインテリジェンスを獲得
- ・使いやすい設計インターフェイスで開発者の生産性を向上
- ・統一したアプローチであらゆるレイテンシで情報を提供
- ・複数のソースとターゲット間の新しいストリーミングデータへ迅速に適応
- ・リアルタイムストリーミングのコンフィグレーション、導入、管理、モニタリングを簡略化

幅広いデータソースに接続し、マルチレイテンシで処理するインフォマティカのエンタープライズ向けストリーミングデータ管理

ストリーミングデータの収集と取り込みには新しいアプローチが必要

今日のデータ主導型の経済環境において、企業はストリーミング、IoT、リアルタイムデータを利用して、新たなビジネスチャンスを開き、競争優位性を獲得しています。ストリーミングデータにより、予防保全、不正行為の検出、リアルタイムなサービス提供による顧客体験向上などのイニシアチブを通じて、データ主導型の意味決定をリアルタイムに行い、業務の生産性を強化できます。

ストリーミングデータの収集と管理には、従来のバッチ指向のデータ収集プロセスに加え、新しいアプローチが必要です。データの格納などの中間ステップが必要なバッチソリューションは、データをリアルタイムで取り込むようには設計されておらず、注意深く監視してログの停滞を防ぐ必要があります。ただし、既存のバッチパイプラインはストリーミングデータの収集と管理に活用できるので、ストリーミングデータのために新しい設備へ投資する必要はありません。

ハイパフォーマンスなエッジデータ処理にはフィルタリングとデータドリフトの管理が不可欠

ストリーミングデータはリアルタイムに収集し、ピークと異常なトラフィックを管理する能力を持つデータレイクに移動する必要があります。データ量が非常に多い場合もあるため、データをレイクに取り込む前に、重要度の低いデータの一部をアプリケーションから除去する必要があります。

ストリーミングシステムとIoTシステムから収集したデータは、ファームウェアのアップグレードやソースシステムへの変更といったさまざまなイベントにより、時間の経過とともに変化する可能性があります。この現象をデータドリフトと呼びます。ストリーミングソリューションには、データ管理プロセスを中断することなく、データのドリフトへ自動的に対処できる機能が必要です。

マルチレイテンシのデータ管理をサポートするデータ統合

ストリーミングデータは速度が速く量が多いため、リアルタイムでの統合と変換が困難です。リアルタイムのストリーミングデータおよびバッチデータの処理をサポートするには、デュアルモードアプローチを同時に実行する必要があります。ピーク時のリソース制限によってストリーム処理のパフォーマンスが低下する場合は、処理の一部をバッチに移行する必要があります。開発者が必要としているのは、データストリームの設計は1回のみで、後は運用環境にある数千のIoTデバイスへ実装するだけで、データ処理を自動的に管理し、サービスレベル契約を満たすことができるシステムです。

さまざまなソースからデータを収集することが必要

ストリーミングデータの世界には、数多くの規格とプロトコルがあり (MQTT、OPC、AMQPなど)、データタイプもさまざまです (JSON、XML、Avro)。さらに、データは構造化形式、半構造化形式、非構造化形式が混在しています。また、ファイル内、データベース内、アプリケーション内に存在します。このようなデータをストリーミングソリューション全体の一部として収集する必要があります。また、業務イベントをトリガーにする変更をとらえるために、ストリーミングデータをメッセージング環境やデータベース環境から収集する必要があります。

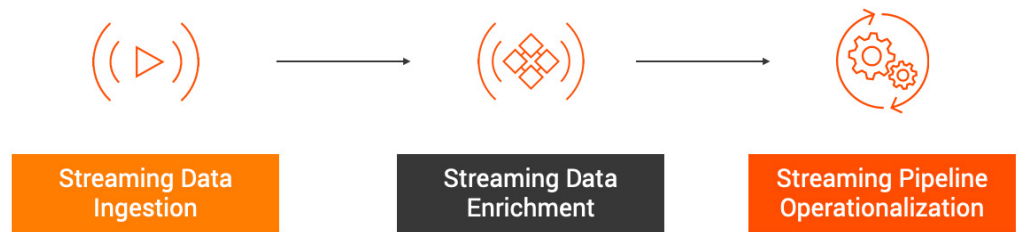


図1: ストリーミングデータのフロー

主なメリット

リアルタイムストリーミングのイニシアチブからより迅速により多くの価値を引き出す

ストリーミングデータの取得、意味付け、実行が可能な単一のストリーミングアナリティクスソリューションによって、リアルタイムのオペレーショナルインテリジェンスを実現できます。このソリューションにより、変更データだけでなく、リアルタイムデータも取り込んで、データ管理を通じて、データのフィルタリング、変換、集計、エンリッチ化、処理を行えます。そして、データをデータレイクに送信し、アナリティクスやリアルタイムのアクションに使用できます。

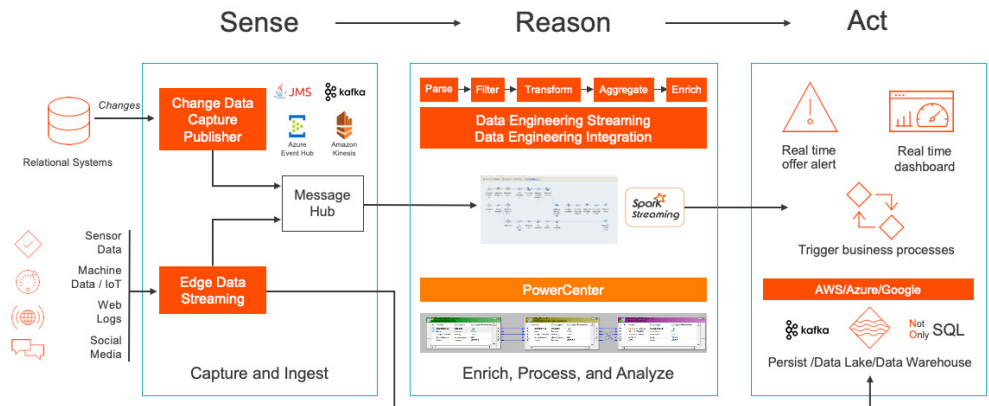


図2: リアルタイムのストリーミングデータ管理

リアルタイムのオペレーショナルインテリジェンスを獲得

ストリーミングデータの収集と統合を行うことで、業務活動の状況を総体的かつリアルタイムに把握できます。Informatica® Edge Data Streamingは、コモディティハードウェアでストリーミングデータの継続的な取り込みとタイムリーな処理を行えるように設計されており、データの間中ステージングやブローカーによる仲介は必要ありません。新しいデータソースの追加やターゲットの処理を迅速かつ簡単に実行でき、ソースとターゲット間のデータフローが簡素化されるため、IT部門はストリーミングデータの管理に費やす時間減らし、戦略的な利益の創出や業務効率化のためにデータを活用する方法の探索に集中できるようになります。さらに、Edge Data Streamingによってエッジ変換も可能になります。その結果、質の低いデータがデータレイクに加えられることはなく、データの泥沼化を回避できます。

Informatica Data Engineering Streamingにより、迅速にデータをインテリジェンスに変換して、パターンを特定し、異常を検出し、差し迫ったリスクや機会をもたらすイベントを意思決定者に通知することができます。また、データをHadoopなどの処理環境に保存して継続的に使用できるようにし、履歴情報と関連付けることで予測アナリティクスを実行できます。

一元化したマルチレイテンシのアプローチで将来にわたり投資効果を確保

利用可能なシステムリソースとSLA（サービスレベル契約）に基づいて、ストリームおよびバッチデータ処理を最適化できます。Spark Streamingでの1秒以下の高速ストリーム処理からHadoopでの大規模なバッチ処理まで幅広いデータ処理が可能で、データパイプラインの再設計や再構築は不要です。いったんデータパイプラインを構築すれば、特別な開発を行わなくても、あらゆるレイテンシで実行することができます。

多様なソースとターゲットをサポート

Informatica Edge Data Streamingには、現在も増え続けている幅広いストリーミングデータソースとターゲットをサポートする、すぐに利用可能な軽量のエージェントが含まれています。これらのエージェントは、ソースとターゲットのアダプターを内部的に開発する必要性を最小限に抑え、処理環境にストリーミングデータを統合するプロセスを高速化します。エージェントは、ユーザー定義のトポロジ構成に基づいて自動的に配備され、高性能のメッセージバスを介して直接接続されます。そのため、Edge Data Streamingではデータのステージングが不要で、ワンホップのフローでソースからターゲットに簡単にデータを移動させることができます。ソース側で、エージェントはファイルに記録されたデータであっても読み取ってディスパッチします。ターゲット側で、エージェントは受け取ったデータを適正な処理環境に書き込みます。

Informatica Change Data Captureは、顧客の新規登録や出荷データの作成といった業務イベントを認識し、データベースへの追加や更新、削除といった処理を発生時に取り込みます。データベース変更データは、中間のステージングテーブルを通さずにメッセージングハブへ直接パブリッシュできます。

インフォマティカについて

デジタルトランスフォーメーションによって我々の期待値が変化しています。より良いサービスを、素早く、便利に、低コストで利用したいという期待が高まっているのです。企業も状況に応じて変化する必要があります。そしてそのヒントは「データ」にあります。

エンタープライズ向けクラウドデータ管理で世界をリードするインフォマティカは、俊敏性の向上、新たな成長機会の獲得、新しいソリューションの開発を実現するための洞察を通じて、あらゆる産業や分野の企業がインテリジェントにビジネスをリードできるよう支援します。インフォマティカは、あらゆるデータを徹底的に重視し、企業の成功に必要なとされる汎用性を提供します。

インフォマティカは、企業がこれからのインテリジェントな破壊的イノベーションを推進できるよう、当社が提供するあらゆるサービスを通じてデータの力を継続的に引き出すことを支援します。

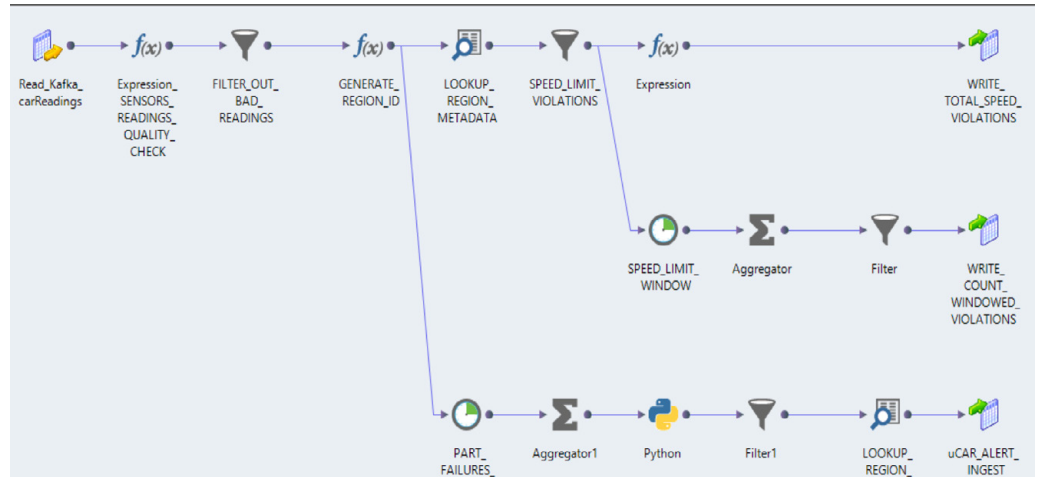


図3: データパイプラインワークフローの開発環境

使いやすい設計インターフェイスで開発者の生産性を向上

価値を獲得するまでの期間とは、設計、構築、テストから導入、メンテナンスまでを、いかに迅速に進めることができるかということを意味します。Informatica Data Engineering Streamingでは、手作業のコーディングに比べて、開発の生産性を最大5倍まで高めることができます。開発者は、ビジュアルな開発環境とあらかじめ組み込まれた動的なテンプレートを使用することで、Spark Streamingの概念や言語に関する専門知識がなくても、データストリームを構築することができます。また、シンプルなコンフィグレーションパラメータによって、データストリームを迅速に運用環境へ展開できます。ビジュアル開発環境と基盤となる処理エンジンが高度に抽象化されているため、データストリームを任意の場所（オンプレミスまたはクラウド）に展開できます。

今後のステップ

詳細は、[Informatica Data Engineering Streaming](#)と[Data Engineering Integration](#)の製品ページをご覧ください。



〒105-6226

東京都港区愛宕2-5-1 愛宕グリーンヒルズMORIタワー26階 電話：03-6403-7600(代表) FAX：03-3433-1021

IN17_0919_03714

© Copyright Informatica LLC 2020. Informatica, Informaticaロゴは、米国およびその他の国におけるInformatica LLCの商標または登録商標です。インフォマティカの商標の最新版は、<https://www.informatica.com/jp/trademarks.html>をご覧ください。その他すべての企業名および製品名は、各社が所有する商号または商標です。本文書に記載されている情報は、予告なく変更されることがあり、現状のまま提供され、明示または黙示を問わず一切の保証を伴いません。