

InterSystems IRIS Data Platform: リアルタイムの データ集約型アプリケーションを実現する統合プラットフォーム



InterSystems IRIS Data Platform: リアルタイムのデータ集約型アプリケーションを実現する統合プラットフォーム

エグゼクティブサマリ

あらゆる業界の組織は、イベント、洞察、行動の間で生じる遅延の短縮と、遅延を解消することでもたらされる戦略および運用上の恩恵を活用したいと考えています。さらに、データ駆動型のインテリジェンスを、リアルタイムのビジネスプロセスに組み込む努力しています。

こうしたゴールを上手く実現することができると、以下に示すような多くのメリットが生まれます。

- 今までにない革新的なビジネスサービスの提供
- 収益の増加
- 顧客体験の向上
- 運用の効率化
- リスクの特定と削減
- 絶え間なく変わる、新しい業界規制の順守
- コストの削減

このホワイトペーパーでは、こうした遅延の短縮や解消によって生じる機会と課題について明らかにし、さまざまな業界でリアルタイムのデータリッチなソリューションの開発、展開、保守を簡素化する、新しいテクノロジーについて解説します。

はじめに

組織は、かつてないほど多くのデータを自由に使うことができるようになりました。しかし、多くの組織は競争上の優位性を得るため、こうしたデータから洞察を引き出し、それに基づいてリアルタイムで行動するよう迫られています。企業は、リアルタイムのデータ集約型アプリケーションを構築することで、こうした機会をフルに活用したいと考えています。それを支えるのが、以下の機能を備えたテクノロジーです。

- イベントとトランザクションのリアルタイムデータを、大規模な履歴/参照データセットと合わせて、遅延なく分析する。
- リレーショナル、ドキュメント、キーバリュー、オブジェクト、非構造化テキストなど、さまざまなデータモデルとデータ表現を利用する。
- 異種のアプリケーションとデータソースを統合する、シームレスでリアルタイムの複合プロセスを構築する。

- ワークロード、データサイズ、ユーザボリュームの増加に対応できるよう拡張する。
- SQL クエリ、機械学習、予測分析、自然言語処理(NLP)などの分析処理を、データ駆動型アプリケーションに組み込む。
- オンプレミス、クラウド、ハイブリッドの展開を可能にする柔軟な展開の選択肢と、継続的な配信と DevOps 方法論を利用する機能を活用する。
- さまざまな分野の専門家の雇用を必要とせず、こうした機能をコスト効率の良い方法で提供する。

リアルタイム組織の実現

テクノロジー業界のアナリスト会社である IDC は、最近、世界中の 500 社を超えるさまざまな業界の企業に聞き取り調査を実施しました。調査企業の 75%を超える企業は、最新のライブデータを分析できないことが、新たなビジネスチャンス活用の大きな阻害要因になっていると伝えています。さらに、半数以上の企業は、それによって業務効率が低下していると述べています。¹

この調査により、企業の 64%では、ETL(抽出、変換、ロード)処理を行って、運用システムからデータウェアハウスにデータを移動し、運用データを分析するまでに、5 日以上 の遅延が生じていることが判明しました。

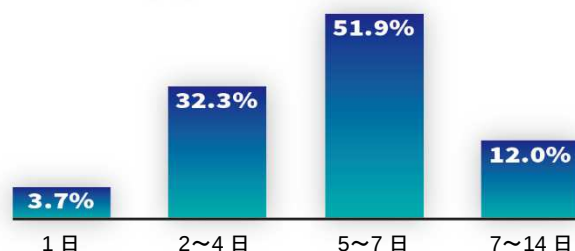


図1:ELT による分析データベースへの運用データの移動に要する平均時間

出典:3rd Platform Information Management Requirements Survey, IDC, October, 2016, サンプル数=502

変更データキャプチャ(CDC)処理に依存するリアルタイムアプリケーションでは、データを分析可能にするまでの所要時間が 1 分を超える CDC プロセスが 96%を占めており、65%は 10 分を超えていると報告されています。重要なリアルタイムのユースケースにとって、この遅延は限度を超えています。

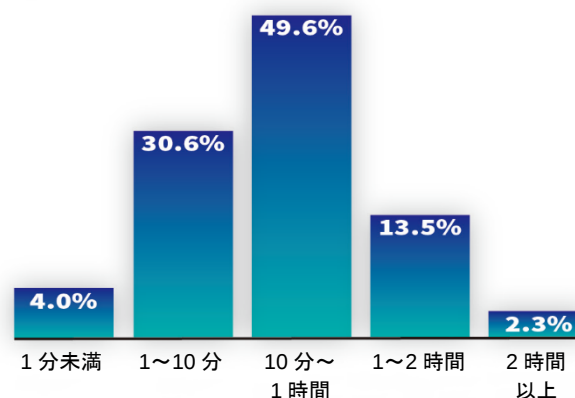


図2:CDC 処理の平均所要時間

出典:3rd Platform Information Management Requirements Survey, IDC, October, 2016, サンプル数=502

ソースが多岐にわたるライブデータのリアルタイム分析を必要とするアプリケーションは、ほとんどすべての業界で実装されています。業界とその用途の例を以下に示します。

- **金融サービス**:州と連邦政府の必須規制の順守、不正の検出、リスク管理の取り組み
- **組立製造/相手先商標製品製造(OEM)**:予知保全
- **運輸/物流**:コンテナと積荷のリアルタイム追跡
- **小売業**:顧客と訪問者のターゲティングとパーソナライズ
- **公衆安全**:初期対応者の状況認識
- **医療**:診療現場での個別化された事前対応型の処置

こうしたアプリケーションでは、同じエンジンでトランザクションと分析のワークロードを同時に処理し、データの移動、マッピング、変換を不要にして、遅延と複雑さを解消するデータプラットフォームを必要とします。

InterSystems IRIS Data Platform™は、そうしたニーズに応えるプラットフォームです。InterSystems IRIS は、多種多様な異種データソースの取り込み、組み込み型のリアルタイム分析のサポート、データ量とユーザ数の増加に応じた容易な拡張性、他のシステムとのシームレスな相互運用性、DevOps に適合した柔軟性と俊敏性に優れた展開機能を持っています。

¹ "Choosing a DBMS to Address the Challenges of the Third Platform" (IDC, 2017)



InterSystems IRIS Data Platform

InterSystems IRIS Data Platform は、リアルタイムのデータリッチなソリューションの開発、展開、保守を簡素化する、包括的な統合プラットフォームです。その主な機能には、トランザクションと分析の同時処理機能、完全に同期化された、多様なデータモデル(リレーショナル、階層型、オブジェクト、ドキュメント)のサポート、異種のデータサイロとアプリケーションを統合する完全な相互運用プラットフォーム、バッチおよびリアルタイムのユースケースをサポートする高度な構造化/非構造化分析機能などがあります。さらに、ベストオブブリードの分析機能を InterSystems IRIS ソリューションに取り込むオープンな分析環境を実現し、クラウド展開とオンプレミス展開の任意の組み合わせが可能な、柔軟な展開オプションを提供します。

InterSystems IRIS は、単一のアーキテクチャに基づいて新規に構築された、さまざまなアプリケーションとシナリオをサポートする単一製品です。

InterSystems IRIS Data Platform は、以下のような特徴を備えています。

- リアルタイムアプリケーションをサポートする、ハイブリッドトランザクション/分析処理
- 多様なデータモデル
- 組み込み型のオープンな分析機能
- Apache Spark との統合
- ビジネスインテリジェンス(BI)
- 高度な分析をリアルタイムプロセスに取り込む機能
- 自然言語処理(NLP)
- 相互運用性
- 統合開発環境
- 柔軟な展開オプション

リアルタイムアプリケーションをサポートする、ハイブリッドトランザクション/分析処理

InterSystems IRIS Data Platform の中核を成すのが、実績のある、エンタープライズクラスの分散型ハイブリッドトランザクション/分析処理(HTAP)データベースです。HTAP は、リアルタイムデータ(ACID 準拠のトランザクションデータを含む)と非リアルタイムデータに対する大量の分析ワークロードを同時に処理しながら、トランザクションデータの取り込みと格納を高速に実行できます。このアーキテクチャでは、分析処理を目的とした、異なる環境へのリアルタイムデータの移動に伴う遅延が排除されます。

InterSystems IRIS の HTAP が実現した、必要な規模で高性能を発揮する能力は、多くの技術革新による成果です。

シャーディングの向上

InterSystems IRIS は、大規模なデータセットに対してクエリを実行する、効率的で高性能な手法を提供します。InterSystems IRIS のシャードクラスタは、ワークロードとデータセットをアプリケーションサーバ層で水平方向に分散し、特定の大きなテーブルのデータを複数のノード(これをデータシャード²と呼びます)に分割します。

シャーディングはさまざまなアプリケーションで有利に働きますが、最大の恩恵を受けるのは、以下のいずれかの作業を含むユースケースです。

- 非常に大規模なデータセットを検索するクエリ
- 大規模なデータセットでの複雑なクエリ
- 高速/大量のデータ取り込み

InterSystems IRIS のシャードクラスタがアプリケーションクエリを受信すると、**シャードマスタ**は分解されたクエリを並列実行のためにデータシャードにプッシュし、個々のシャードが返した結果を集約して、最終的な結果をアプリケーションに返します。あるシャードが作業を完了するために他のシャードのデータを必要とする場合、そのシャードは、他のシャードにある**必要とするデータのみ**に、シャードマスタを介さずに直接アクセスできます。

そのため、InterSystems IRIS は、複数のテーブルを含む複雑なクエリでも、高性能、効率性、信頼性を一貫して実現します。一方、シャードアーキテクチャをサポートする他の多くのデータベースプラットフォームは、テーブル全体のブロードキャストを使用しているため、パフォーマンスの低下やタイムアウトが起きることがあります。

シャーディングはデータの非結合パーティションを生成するため、それぞれのデータサーバのキャッシュは完全に独立し、データサーバを直線的に追加することでクラスタの全体的なメモリが増加します。したがって、サイジングを適切に行うことにより、すべてのデータをメモリに適合させなくても、InterSystems IRIS はインメモリデータベースが持つパフォーマンスの恩恵を実現します。

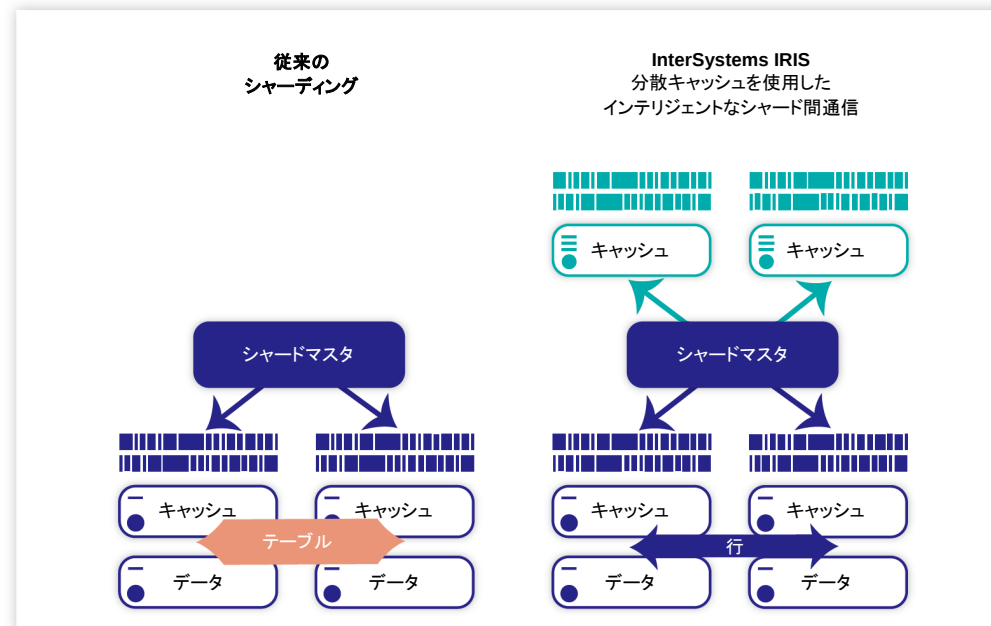


図3: インテリジェントなシャード間通信に基づく大規模な分散データセットの分析

² データシャードは、クラスタのシャードマスタで定義された、それぞれのシャードテーブルに属する1つの水平パーティションを格納する、InterSystems IRIS のインスタンスです。このインスタンスをホストするノードを、シャードデータサーバと呼びます。

InterSystems IRIS のシャードクラスタには、パフォーマンス面で次のような付加的なメリットがあります。

- InterSystems IRIS の Java Database Connectivity (JDBC)ドライバの透過的な並列ロード機能は、シャード全体で同時かつ高速にデータを取り込む Java ベースツールの使用をサポートします。
- 大規模なマルチユーザクエリのワークロードがシャードマスタのボトルネックとなる場合は、シャードマスタの前にアプリケーションサーバ層を追加し、分散アプリケーションのロジックとキャッシュによって、ユーザボリュームに応じた拡張を行うことができます。

シャーディングはアプリケーションに対して透過的に行われるため、アプリケーションコードの変更はほとんど、または、まったく必要ありません。アプリケーションは、シャードテーブルと非シャードテーブルの違いをまったく認識しません。この相違は、設計時の考慮事項にすぎません。

InterSystems IRIS のアーキテクチャは、分割された分散データセットに存在するパターンや関係を特定するための、複雑な複数テーブル結合が実行可能です。その際、コシャーディング³やデータの複製、ネットワークでのテーブル全体のブロードキャストは不要です。

³コシャーードデータとは、共通のキーに基づいて分割された分散データを意味します。

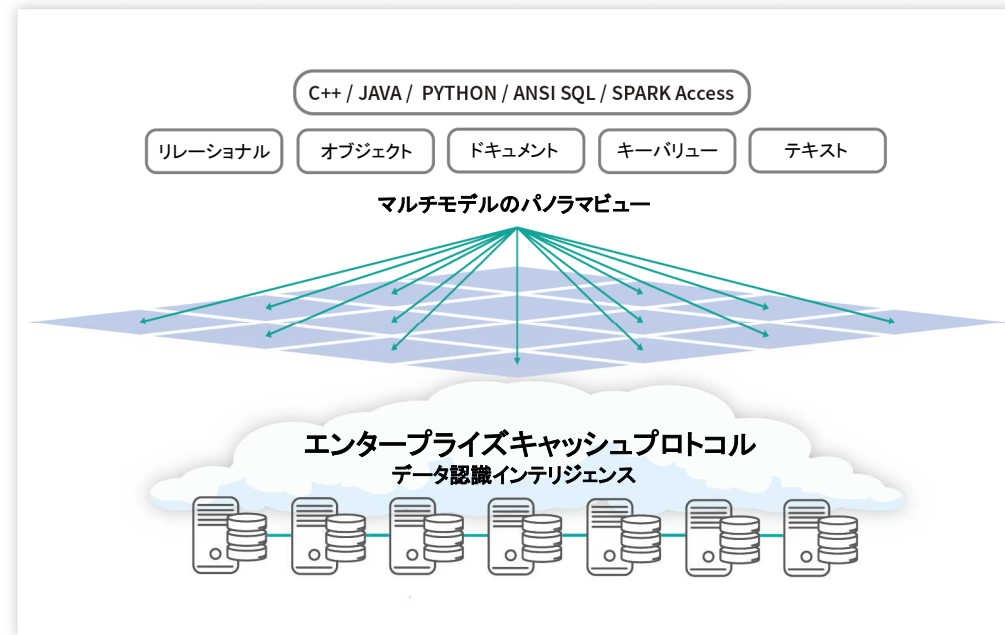


図4: InterSystems IRIS によるマルチモデル、分散データへの統合アクセス

パフォーマンス向上とコスト削減

分析処理の効率的な実行に加えて、同時に存在するトランザクションと分析のワークロードを必要な規模で高速に処理します。分析のためにトランザクションデータを別の環境に移動する必要はありません。トランザクションの処理、永続ストレージによるデータの永続化、分析クエリへのトランザクションデータの提供——そのすべてを、InterSystems IRIS は市販のハードウェアで数十ナノ秒以内に遂行します。

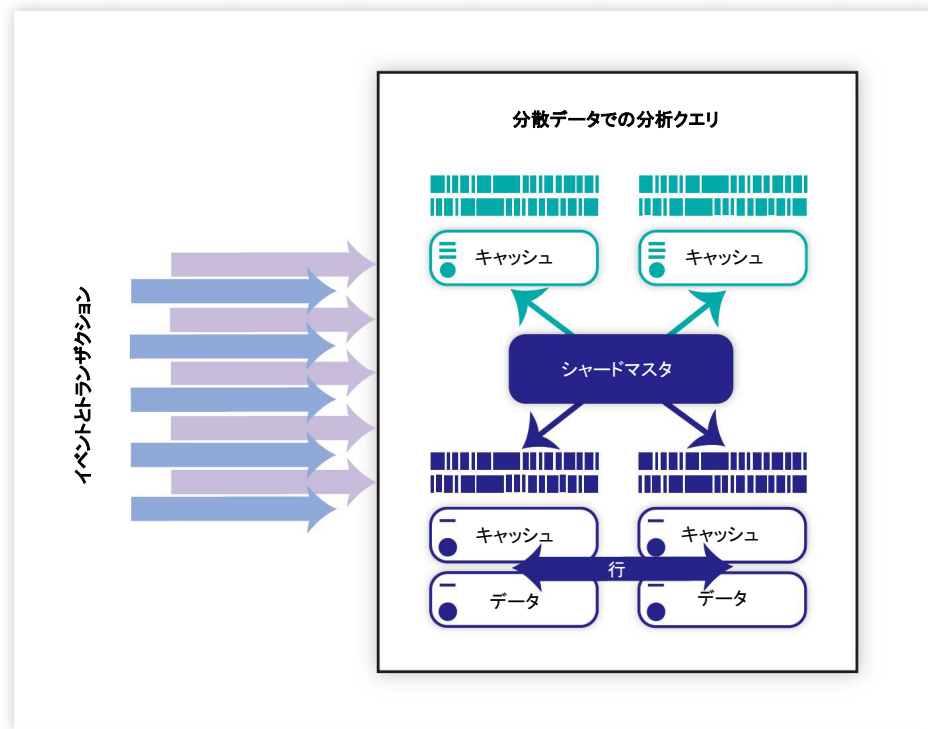


図5: 水平方向に分散したHTAP

InterSystems IRIS は、ダイレクトな共有メモリアイトとクライアント/サーバ分散 SQL 処理を同時にサポートして、高性能なトランザクションと分析を同時に実行することができます。その結果、分割された分散データセットに格納されているデータと一緒にリアルタイムデータを確実に処理、分析できるようになり、時間と運用コストが削減されます。

非シャードテーブルとシャードテーブルの両方で高可用性を確保するため、データを格納しているすべてのノードをミラーリングすることができます。コンピュータードは、ユーザワークロードの変動に応じて容易に追加、削除できます。InterSystems IRIS は、エンタープライズレベルの強固なセキュリティ、Kerberos、LDAP、KMIP (Key Management Interoperability Protocol) との統合、ロールベースのアクセス制御、移動中と保存中のデータの暗号化を提供します。

多様なデータモデル

InterSystems IRIS は、本格的なマルチモデルデータベースを基盤にして構築されています。すなわち、格納されているデータには、リレーショナルモデルやオブジェクトモデルなどの多様なデータモデルを介してアクセスできます。これらのデータモデルは常に同期されます。その結果、データの複製や異なるデータ表現間のマッピング (オブジェクトとリレーショナルのマッピングなど) が不要になります。複数のデータタイプをネイティブにサポートできるため、組織は、最適な形式と表現でデータをモデル化、保管、使用して、柔軟なソリューション開発、パフォーマンスの向上、複雑さの軽減を図ることができます。

これらの新しいデータタイプはどれくらい重要ですか？

(評価尺度: 1 = それほど重要ではない、5 = 非常に重要)

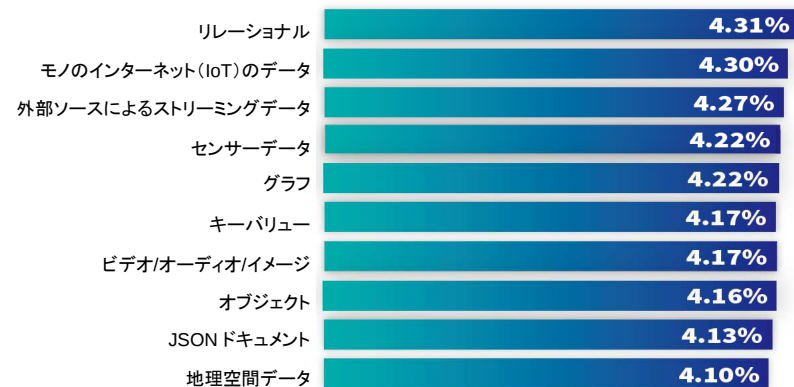


図6: データプラットフォームでの各種データタイプのサポートの重要性

出典: 3rd Platform Information Management Requirements Survey, IDC, October, 2016, サンプル数=502

組み込み型のオープンな分析機能

InterSystems IRIS は、今日のデータ集約型のリアルタイムアプリケーションに求められる多様な要件を満たすため、さまざまな分析機能をサポートしています。最先端をいく、分散 SQL、BI、NLP の組み込み型分析機能を利用できます。さらに、サードパーティやオープンソースの広範な分析パッケージを必要に応じて組み込むことができます。

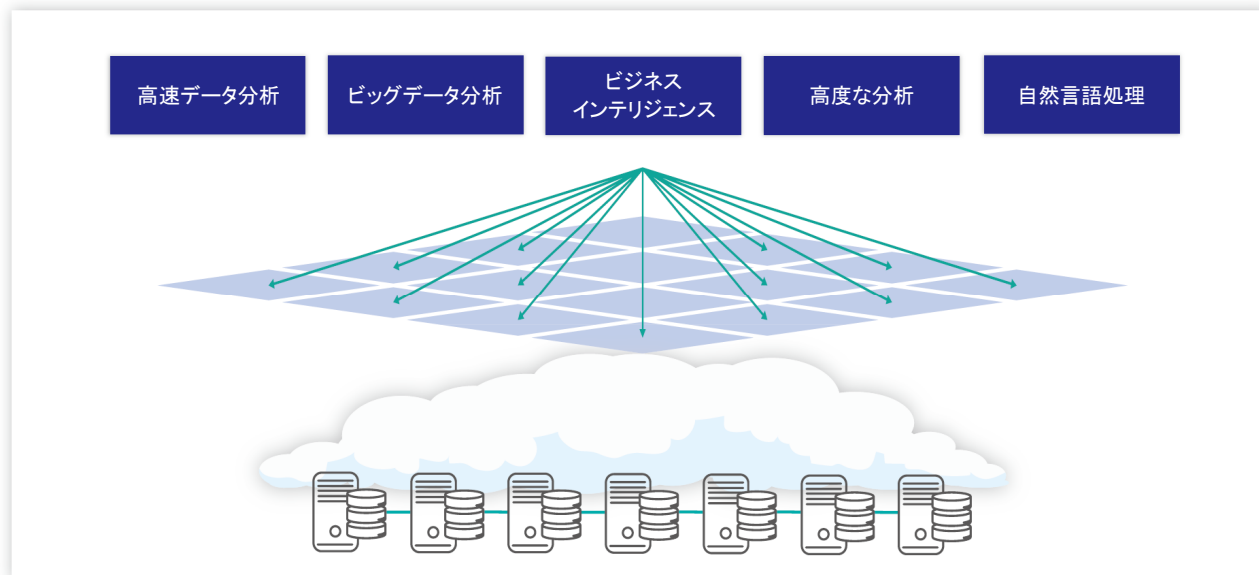


図7: InterSystems IRIS の組み込み型のオープンな分析機能

リサーチ会社のガートナーが大企業を対象にした2017年の調査によると、1,931人の回答者の45%はデータマイニングと予測分析の使用を予定していると述べていました。39%は Apache Hadoop または Apache Spark の使用を予定しており、25%は、Apache Hadoop または Apache Spark が提供する高度な分析機能の使用を予定していました。⁵

高度な分析技術を採用する動きが急速に広がっています。こうした手法と技術には、機械学習、予測分析、人工知能、ビッグデータのリアルタイム処理フレームワーク（例：Apache Spark）などがあります。

リアルタイム（HTAP）の（分散型）ビッグデータ処理機能に加えて、InterSystems IRIS は、次に述べる分析と統合の機能を提供します。

Apache Spark との統合

オープンソースの高性能なクラスタコンピューティングのフレームワークである Apache Spark は、パフォーマンスを必須要件とする大規模な分散データセットで主に使用されています。Apache Spark は、Apache Hadoop（MapReduce）の100倍に達する高速処理を実現し、一般的な機械学習と統計アルゴリズムの多くを利用できます。

InterSystems IRIS は、シャード対応のネイティブ Spark コネクタを介して Apache Spark と直接統合されるため、InterSystems IRIS のアプリケーションは Spark の処理を取り込むことができ、Spark のアプリケーションは InterSystems IRIS の分散データを取り込むことができます。Apache Spark コネクタは、性能を最大限に高めるため、InterSystems IRIS シャードクラスタのデータシャードをネイティブパーティションとして提示します。このコネクタは、InterSystems IRIS データベースが備えるパーティション化の性質を認識し、Apache Spark のワーカーノードが直接、シャードに自動接続して、データの非結合部分で並列処理を行えるようにします。また、この直接的な並列接続によってスループットも大幅に向上し（各接続で渡すデータ量が減るため）、シャードクラスタへの高速なデータ取り込みをサポートします。

⁵ Rita L. Sallam, et al., "Survey Analysis: BI and Analytics Spending Intentions, 2017" (ガートナー、2017)

ビジネスインテリジェンス

InterSystems IRIS では、BI モデリング、分析、エンドユーザダッシュボードの機能が完全に統合されています。BI モデルは、ビジネスにとって有意義な見方を示します。その例として、集合体概念（製品ライン、販売地域、市場区分など）や数値指標（収益、費用、前年比成長率、欠陥率など）が挙げられます。InterSystems IRIS の BI モデルは、トランザクションデータやその他の必要とされるデータに直接基づいて作成できます。完全に自動化された同期化オプションがあり、ETL 処理が不要になります。ドラッグアンドドロップ方式の分析機能を使用することで、専門知識を持たないユーザでも、あらゆるレベルのデータを調べて複雑なクエリを容易に実行できます。InterSystems IRIS のダッシュボードは、最新のビジネスメトリックを表示して、限定された分析オプションをユーザに提供します。

高度な分析をリアルタイムプロセスに取り込む機能

組織は、外部のツールとアプリケーションを使用して、データマイニングや機械学習のアルゴリズムで作成した予測モデルを取り込むことができます。これは、InterSystems IRIS に組み込まれた、PMML (Predictive Model Markup Language) のサポートによって可能になります。PMML は、外部の分析アプリケーションまたはフレームワークで開発した予測モデルのすべてのパラメータを完全に定義する XML 標準です。InterSystems IRIS に PMML モデルをロードすると、このモデルのリアルタイム実行を可能にするネイティブコードが生成されます。外部ツールは必要とせず、システム間でのデータ受け渡しでパフォーマンスが低下することはありません。こうした統合により、データサイエンティストなどの専門家が作成した予測モデルは、InterSystems IRIS 内のデータ処理パイプラインとビジネスプロセスにシームレスに取り込まれます。

自然言語処理

InterSystems IRIS は、自然言語テキストの意味と感情を推論する NLP 機能を備えています。テキストの概念と関係が自動的に識別され、事前作業や専門知識は必要としません。この高度な NLP 機能は InterSystems IRIS に組み込まれており、ビジネスプロセスに取り込むことができます。組織は、メモ欄、ソーシャルメディア、データリッチなアプリケーション内の他のソースから情報を取り入れることができます。

それぞれが特定の機能や分野を持つ、専門化した NLP ツールは多岐にわたるため、アプリケーションによっては、こうしたツールを正しい順序で使用する必要があります。InterSystems IRIS は、Apache UIMA (Unstructured Information Management Architecture) 標準をサポートしています。UIMA により、標準ベースのプラグイン可能な NLP パイプラインを定義して実行できます。Apache UIMA をサポートすることで、NLP 機能のオープンな相互運用が InterSystems IRIS で実現します。

大企業の 46%は、2017 年に非構造化コンテンツの感情分析を自社のアプリケーションに取り込む計画を立てました。⁶

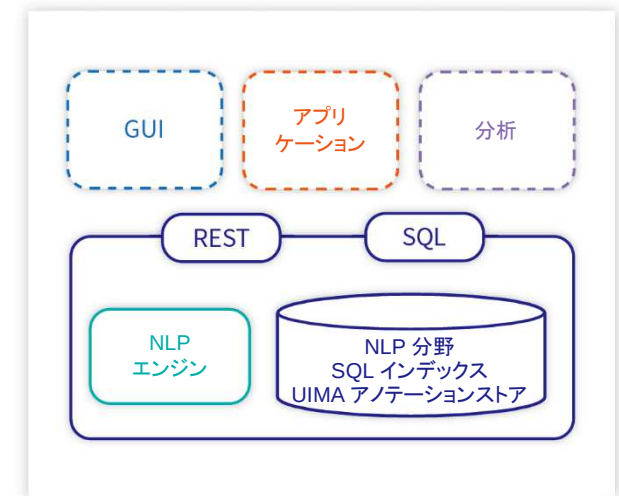


図8: InterSystems IRIS の自然言語処理機能

⁶ Rita L. Sallam, et al., "Survey Analysis: BI and Analytics Spending Intentions, 2017" (ガートナー, 2017)

相互運用性

InterSystems IRIS は、ネイティブな統合および相互運用機能の包括的なセットを提供します。さまざまなパッケージアプリケーション、データベース、業界標準、プロトコル、テクノロジーに伴う接続やデータ変換を、ただちに行うことができます。柔軟性に富むデータ変換機能により、アプリケーション間またはサービス間に存在する、セマンティクスやデータスキーマの相違が解消されます。

アプリケーション開発者は、社内外のデータソース、アプリケーション、サービスとつながるシームレスなビジネスプロセスを構築できます。InterSystems IRIS は、プロセス、ルール、ワークフローを視覚的に表すグラフィカルなツールを提供します。開発者は、システム間の論理的な相互作用に注力できるようになり、アプリケーションのインターフェースやアダプタ、ミドルウェアのメカニズムへの関与を最小限に抑えることができます。このグラフィカルなモデルによって事業部門とIT 部門の連携が実現し、ビジネス要件を満たすソリューションの開発

が促進され、既存プロセスの変更/拡張が容易になります。組み込み型のロールベースのワークフローエンジンは、ビジネスプロセスの手動操作をサポートし、ユーザ間でのタスク分配を自動化して、ユーザの意思決定と行動を取り込みます。

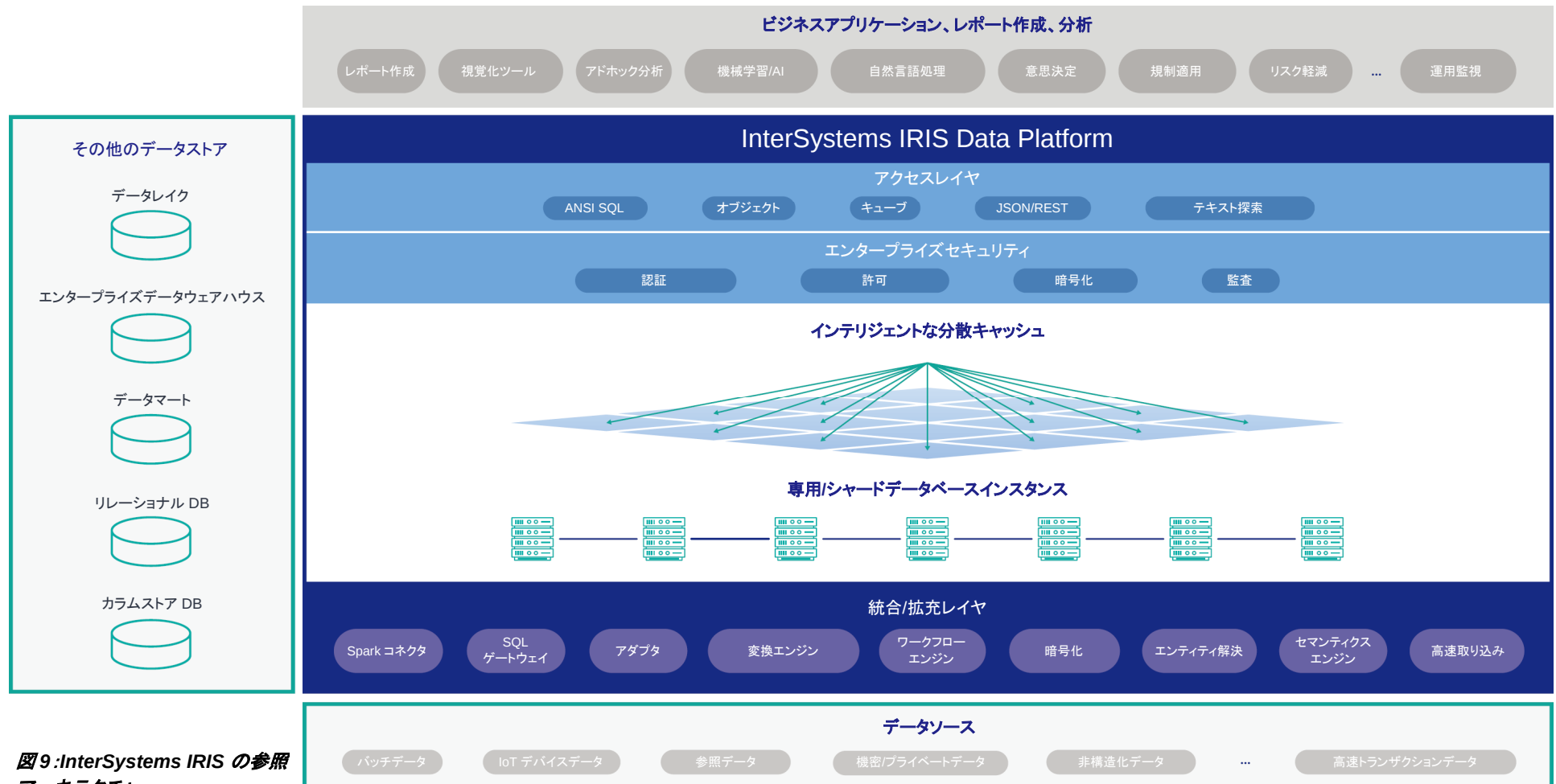


図9: InterSystems IRIS の参照アーキテクチャ

InterSystems IRIS にはデータベースと分析機能が備わっており、高度な分析をビジネスプロセスにシームレスに取り込んで、データベースに格納されているデータとリアルタイムデータの両方を活用できます。長時間実行の非同期トランザクションに伴うデータや伝送中のデータを含むすべてのデータは、データベースで自動的に永続化して、レポートや分析で利用できます。

このプラットフォームは、医療、金融サービス、小売業、通信などの多種多様な業界で使用されている、REST アーキテクチャや Web サービス (JSON、XML、XPath、XSLT、SOAP、DTD など) をはじめとするさまざまな標準をサポートしています。

統合開発環境

InterSystems IRIS が提供するコードベースのグラフィカルな統合環境は、さまざまなプログラミングモデル、プログラミングインターフェース、データ形式を一貫性を持って表現し、すべての機能を包括する単一の開発環境を提供します。

柔軟な展開オプション

InterSystems IRIS は、クラウドベースおよびオンプレミスのインフラストラクチャで、サービスのシンプルで直感的なプロビジョニングと展開を可能にします。さらに、Infrastructure as Code (IaC)、不変のアーキテクチャ、コンテナ化された InterSystems IRIS ベースのアプリケーションの展開などの恩恵をもたらします。新しい技術とそれに伴うトレーニングへの多額の投資が不要になり、さらに試行錯誤のシステム構成と管理に伴う労力を削減します。

組織は、大規模な開発や設備更新を必要とせずに、クラウドコンピューティングとコンテナ化されたソフトウェアが実現する効率性、俊敏性、再現性を活用できます。さらに、InterSystems IRIS は、その構成のプロビジョニングと展開を既存の仮想/物理クラスタで行い、既存のインフラストラクチャや商用のクラウドプラットフォームを含む、エンタープライズレベルの OS プラットフォームでのコンテナ展開をサポートします。

まとめ

InterSystems IRIS は、リアルタイムのデータリッチなソリューションの開発、展開、保守を簡素化する、包括的な統合データプラットフォームです。その主な機能には、トランザクションと分析の同時処理機能、完全に同期化された、多様なデータモデル (リレーショナル、階層型、オブジェクト、ドキュメントなど) のサポート、異種のデータサイロとアプリケーションを統合する完全な相互運用プラットフォーム、バッチとリアルタイムのユースケースが共に可能な高度な構造化/非構造化分析機能などがあります。さらに、ベストオブブリードの分析機能を InterSystems IRIS ソリューションに取り込むオープンな分析環境を実現し、クラウド展開とオンプレミス展開の任意の組み合わせをサポートする、柔軟な展開が可能です。

InterSystems IRIS はさまざまな業界で使用され、戦略および運用面で欠かせないさまざまな恩恵をもたらしています。これは、イベント、洞察、行動の間で生じる遅延を解消し、より多くのデータを活用することで実現します。

さらに、インターシステムズが自信を持って提供する InterSystems IRIS Experience は、InterSystems IRIS のよさをお客様ご自身で体験できる、自主的な実践の機会を提供しています。詳細については、[InterSystems.com/Experience](https://interSystems.com/Experience) をご覧ください。



インターシステムズジャパン株式会社
InterSystems.com/jp/