



これからの 基幹系システムデータ分析

—「IBM z/OS Platform for Apache Spark」活用の4つのメリット



これまでの 基幹系システムデータ活用手法の課題

「基幹系システムデータ」の迅速な分析が、ビジネスを加速させる

多くの企業がビッグデータの活用を検討し実施し始めています。しかし、ひとくちにビッグデータといっても、さまざまな種類のデータがあります。企業は、どのような種類のデータを活用したいのでしょうか。

調査によると、対象として分析・活用したいというデータの1位は、「自社製品の販売データやサービス履歴データ(POSデータなど含む)」、いわゆる「基幹系システムデータ」です。

こうした基幹系システムのデータは、これまでも分析され活用されてきましたが、経営判断のスピード化や市場の多様化に対応するため、迅速な経営へのレポートや業績ダッシュボード化、マーケティング業務での活用など、これまで以上に迅速に、より多くの基幹系システムデータを活用したいというニーズが高まっています。

これまでの基幹系システムデータ活用手法 5つの課題

基幹系システムデータは非常に重要なデータであるため、直接データを参照せずに、必要なデータをいったんデータストアへコピーし、ETL処理を行いデータウェアハウスに格納。それを再度コピーして分析のためのデータマートに格納して活用する手法をとるのが一般的です(図1)。

これには、基幹系システムデータを直接参照・分析する作業によって発生する基幹業務への影響を排除したいという背景もあります。つまり、分析処理の負荷によって基幹業務の処理速度が低下することを避けたいということです。そのために、データを外部システムにコピーする手法が主流となっています。

しかし、データコピーによる基幹系システムデータ活用にはいくつかの課題があります。

●データの一貫性の欠如

分析用データはバッチ処理やETL処理などの後に利用できるため、データ鮮度が犠牲になっている。また、多くのサーバー

にコピーされる際に、サーバー間のデータ鮮度の差異や加工度合いによって、データの一意性が損なわれる場合がある。

●情報漏えいのリスク

データマートから基幹業務データが漏えいするリスクが発生する。

●データウェアハウスなど関連システム費用

データをコピーするためのハードウェア・コスト、データウェアハウスなどのライセンス・コスト、バッチ処理やETL処理の開発コストが発生する。

●情報システム部門の運用工数の増大

分析用のデータウェアハウスやデータマートのシステム運用が必要になる。

●データ活用の柔軟性の不備

新たなデータを必要とする場合、抽出バッチやETLなどに開発期間がかかるため、柔軟なデータ活用がしにくい。

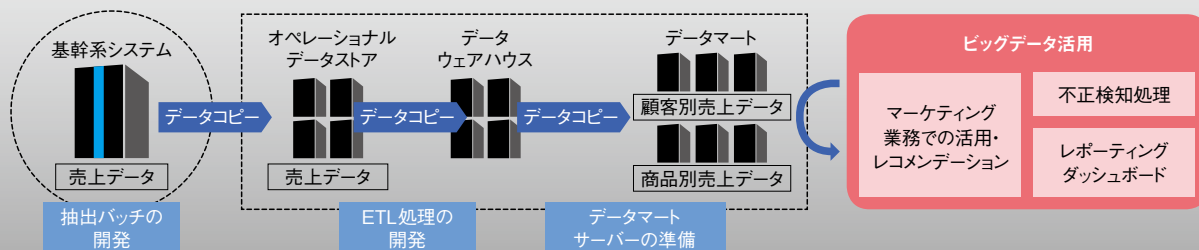


図1:これまでの基幹系システムデータ活用手法

Apache Sparkテクノロジーを活用した、 これからの基幹系システムデータ活用手法

IBM z/OS Platform for Apache Sparkでz Systemsのデータをダイレクトに参照

もし、基幹系システムのデータを負荷をかけずに直接参照でき、分析に活用することができれば、データ漏えいのリスクもなくデータ鮮度もリアルタイムに近い状態で利用することが可能になります。ビッグデータ活用を実施・計画している企業にとって、理想的な方法だと言えるでしょう(図2)。

IBMはApache Sparkテクノロジーを利用した基幹系システム(メインフレーム)データのダイレクト活用を実現しました。Apache Sparkは、オープンソースのクラスター・コンピューティング・フレームワークで、そのインメモリー処理により分析アプリケーションの速度を現在市場に出ているテクノロジーに比べて最大で100倍高速化します。IBMは2015年6月の発表において、Apache Spark に対し大規模な投資を行い、世界各地にあるIBMの研究開発部門に3500人以上の研究者・開発者を配属し、Spark関連のソフトウェア開発プロジェクトに従事させるとしました。これらの取り組みの一環として2016年3月に登場したのが、「IBM z/OS Platform for Apache Spark」(以下、z/OS Platform for Apache Spark)です。

z/OS Platform for Apache Sparkは、Apache Sparkコア、Spark SQL、Spark Streaming、機械学習ライブラリー(MLib)、GraphXで構成されるSparkオープンソースと業界唯一のメインフレーム常駐のSparkデータ抽象化ソリューションを組み合わせています(図3)。

VSAM、SAMなど多種多様なファイルも取り込み活用可能

z/OS Platform for Apache Sparkは、MDSS(Mainframe Data Service for Apache Spark z/OS)と呼ばれる変換ツールのライセンスを無償で利用可能です。VSAMやSAMなど多種多様なデータを利用することもz Systemsユーザーには大きなメリットとなります(図4)。

z/OS Platform for Apache Spark活用による4つのメリット

●コスト削減

データ抽出のバッチ処理、ファイル転送処理、ETL処理などの開発コストやシステム・コストを削減できる。

●データの一貫性

単一のデータを元に分析できるようになる。多数のサーバーに、複数回データをコピーすることで、分析用データを作成してきた従来のようなデータ鮮度のばらつきやデータ不整合がなくなる。

●セキュリティー・リスクの軽減

基幹系システムデータそのもののコピーがなくなり、データ漏えいのリスクが大幅に軽減される。

●ビジネスのスピード・アップ

これまで多くの工数と時間がかかっていた分析用データの作成について、工数と期間の双方が削減できるため、事業部門の要請にも短期で対応可能になる。

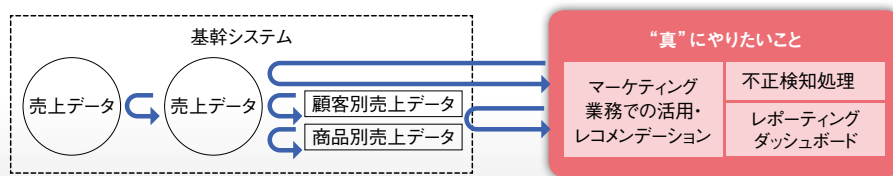


図2:基幹系システムデータのダイレクト活用

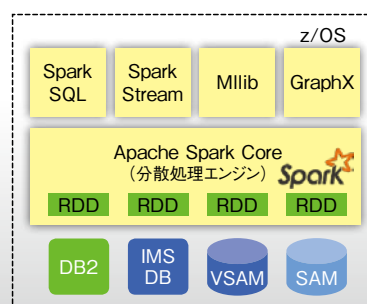


図3:Apache Spark

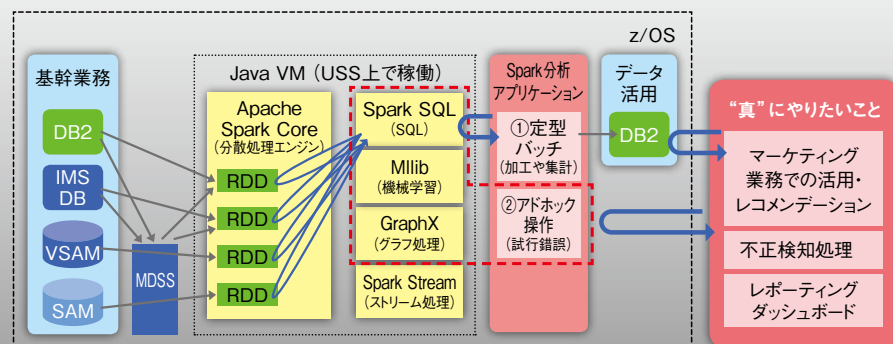


図4:z/OS Platform for Apache Spark

z/OS Platform for Apache Sparkの特徴

パフォーマンス低下とCPU課金を抑える専用プロセッサ「zIIP」

本来、基幹系システムにアクセスし分析処理を行えば、CPU負荷が高まり処理速度の低下を招きます。IBM System z9以降のモデルでは「zIIP」(z Systems Integrated Information Processors)と呼ばれる専用プロセッサが処理を行うため、基幹系業務の処理速度に影響を与えません。

zIIPは2004年に発表されたz/OS用Java実行処理専用プロセッサである「zAAP」(System z Application Assist Processor)の後継ともいえるプロセッサであり、Javaに加えてDB2リモートアクセスとBI/DW、IPSec暗号化、z/OS XML、z/OS グローバル・ミラー、HiperSocketsラージ・メッセージ、VTFMサーバー、DB2先読み・遅延書込みI/O、DB2 RUNSTATS、DB2 ソート・ユーティリティ、z/OS CIMサーバー、RMFモニターIIIデータ収集処理をオフロードすることができます。z/OS Platform for Apache SparkはJava VM上で実行されるため、zIIPを利用します。

パフォーマンス低下防止に加えて、zIIPにオフロードできるメトリックは、CPU課金を抑えることができます。CPU課金コストがほとんど増加せずに高機能なz/OS Platform for Apache Sparkを利用できることは、企業とIT予算を預かる情報システム部門にとっては大きな魅力となります(図5)。

zIIPでは、以下の処理もオフロード可能です。

● IBM SDK (IBM Java Virtual Machine (JVM)) 経由で Javaを使用

- ▶ WebSphere Application Server
- ▶ IMS
- ▶ DB2
- ▶ CICS
- ▶ Java バッチ
- ▶ CIM クライアント・アプリケーション

● z/OS XML System Servicesを使用

- ▶ DB2 9以降 (ニュー・ファンクション・モード)
- ▶ Enterprise COBOL V4.1以降
- ▶ Enterprise PL/I V3.8以降
- ▶ IBM XML Toolkit for z/OS
- ▶ CICS TS V4.1以降

z/OS Platform for Apache Sparkの使い勝手は他プラットフォーム同様

z/OS用だからと言って、Sparkのインターフェースや使い勝手が変わることはありません。図6のように他プラットフォームと同じインターフェースが提供されています。

使い慣れた言語やSparkを利用して、基幹系システムz Systemsのデータを直接参照し、分析することができます。

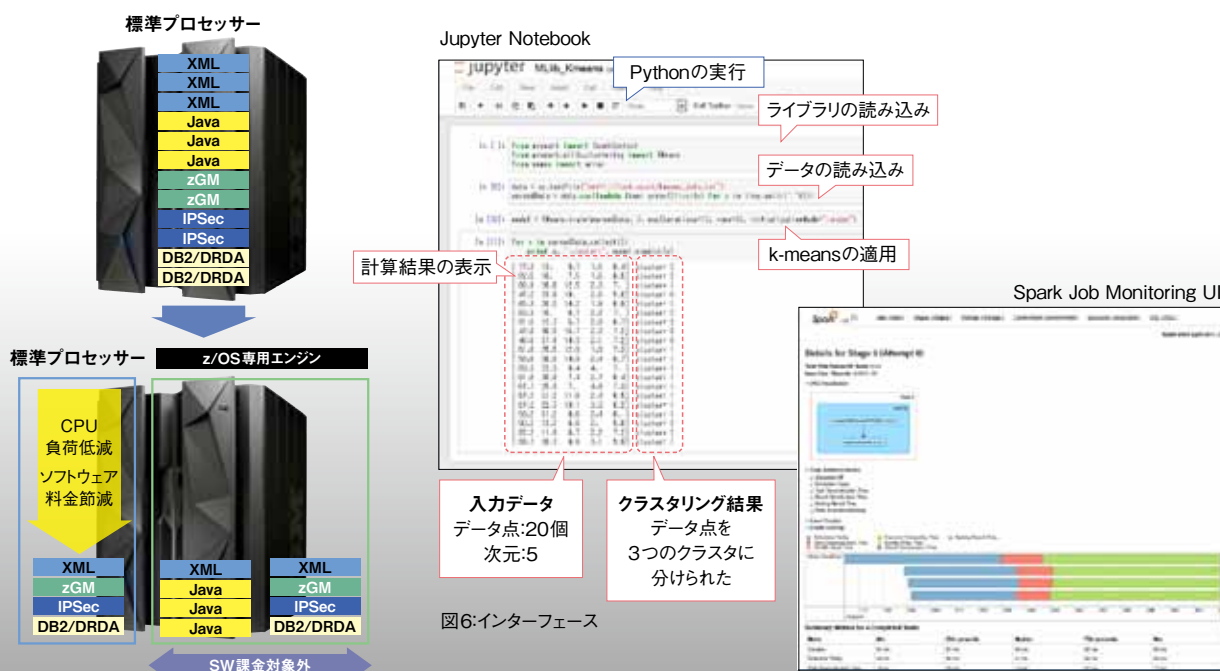


図6: インターフェース

図5: 「zIIP」によるCPU課金の低減

※イメージ図であり、実際の効果はお客さまの状況により効果は異なります

IBM z/OS Platform for Apache Sparkの活用例

活用例1

定型バッチ処理による加工・集計データ提供の短期化

活用法の一つは、従来も行っていた「定型バッチ処理」によるデータ集計や加工の迅速化です。定期的な経営レポートやダッシュボード利用のためのデータは定型化されているため、データ抽出には定型バッチによる処理が効率的です。z/OS Platform for Apache Sparkを利用することでバッチ処理時間を短縮できるようになります。

また、これまでは事業部門から新しい加工・集計データを要請されても、ETLをはじめとして各種バッチ・プログラム開発には長い期間と多くの工数が必要で、迅速に対応することができませんでした。z/OS Platform for Apache Sparkを利用することで、新たな加工・集計データの提供を依頼された場合でも、従来より短時間で提供できるというメリットも生まれます(図7)。

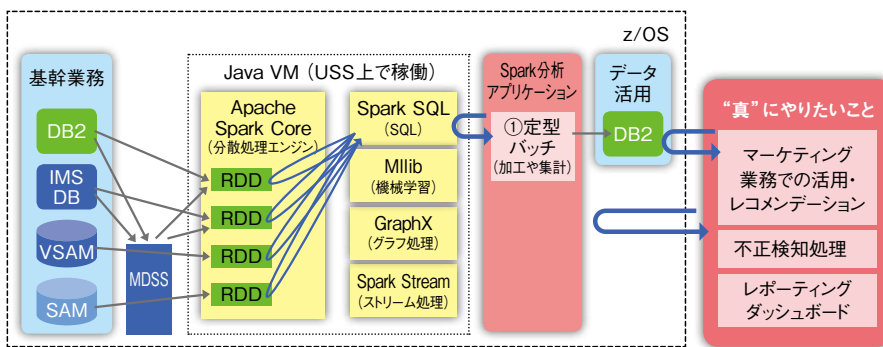


図7:定型バッチ処理

従来型では複数バッチ・プログラム開発と作業が必要であり、コピーを繰り返すためデータのリアルタイム性も低くなっていた。z/OS Platform for Apache Sparkでは、Spark分析アプリケーションによって定型バッチ開発も短期化でき、データも直接基幹系システムから取得し加工するため、高いリアルタイム性を保持できるというアドバンテージが生まれる。

活用例2

アドホック分析による新たな視点での経営可視化

アドホックとは「暫定的な」「臨時的の」を意味するラテン語です。z/OS Platform for Apache Sparkを利用すれば、基幹系システムデータをさまざまな視点から分析することが簡単かつ迅速にできるようになります。

従来は加工データの抽出を依頼、分析作業を行い、足りないデータがあれば再度データの抽出を依頼するというように、試行錯誤を繰り返し新たな分析知見を蓄積するまでに多くの手間と時間が必要でした。

z/OS Platform for Apache Sparkによるアドホック分析では、Spark SQL(表形式データのSQL処理)、Mllib(機械学習処理)、GraphX(グラフ構造データ処理)を活用しさまざまな視点からの分析が簡単に行えます。その都度データ抽出をする必要がないため、数多くの試行錯誤を繰り返すことへの関連部署の負荷が少なくなることが大きなメリットとなります。z/OS Platform for Apache Sparkによって、これまでに得られなかった多くの知見を迅速に得られるようになります(図8)。

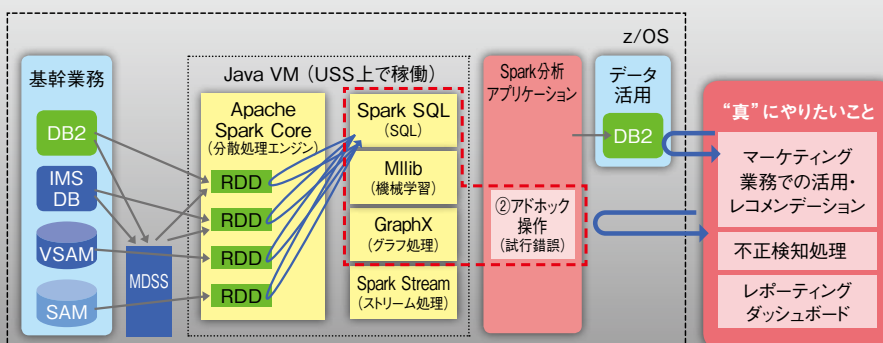


図8:アドホック分析による新たな視点での経営可視化

製品・サービス品質を支えるIBMの豊富な知見とナレッジ

効率的で一貫した サービス・マネジメントの 統合ソリューションを提供

全ての管理をまたがって関連付けられる
情報を一元化し、ナレッジ化することによる
関係者の意思決定を改善

IT スタッフの生産性の強化

豊富な導入実績に基づくベストプラクティ
スの提供

日本だけでなくワールドワイドの実績と豊
富な経験を元にした導入支援サービスを
提供

PDCA サイクルによる 運用改善への対応

運用プロセスや目標に合わせた設定を
GUIの設定レベルで構成、運用上の
KPIの算出や、分析/レポートに利用

管理項目や画面構成、業務プロセス、組
織等の変更を簡単に反映でき、長期にわ
たる継続的使用が可能

お問い合わせ

このカタログの情報は2016年8月現在のものです。仕様は予告なく変更される場合があります。サービスや製品の詳細については、弊社の営業担当員にご相談ください。



日本アイ・ビー・エム株式会社

www.ibm.com/systems/jp-ja/z/os/zos/apache-spark.html

IBM、IBM ロゴ、ibm.com、DB2、IBM z Systems、z Systemsは、世界の多くの国で登録されたInternational BusinessMachines 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、www.ibm.com/legal/copytrade.shtmlをご覧ください。