

# ビッグデータ統合・分析・活用の最前線

## ——「データ統合・分析共通 PaaS」のご紹介

林 直 樹

**要 約** 目まぐるしく変化する市場環境の中、企業が生き残りをかけ戦う際の重要な要素のひとつが「スピード」である。迅速な経営判断に不可欠な ICT 環境もそのスピードに追随する必要がある。昨今では、社内外のビッグデータをリアルタイムで分析し、戦略遂行の PDCA を高速に回すことにより、競争優位を獲得する企業もでてきている。日本ユニシスでは、これらのニーズに対応するために、世界最速の Hadoop デистриビューション「MapR」、リアルタイム分析が可能な DWH ソフトウェア「Vertica」を組み合わせ、オンプレミスだけでなくクラウド上からも提供する「データ統合・分析共通 PaaS」をリリースした。これにより利用者は、環境評価や技術取得に時間をかけずに、少ない初期投資で迅速なシステム構築、安心、安全なシステム運用が得られる。

### 1. はじめに

従来の基幹系システムのような、あらかじめ手順が決まっいて、0か1かで分岐して結果がはっきりしているシステムを Systems of Record (SoR: 記録するシステム) と呼ぶのに対し、必ずしも正解が一つとは限らない、または正解が変化するようなシステムは Systems of Engagement (SoE: 絆を作るシステム) と呼ばれる。顧客接点などフロント系のシステムがそれにあたる。

一方、変化の激しい市場環境の中、企業が競争優位を勝ち取るための重要なポイントは「スピード」であり、IT もそのスピードに追従しなければならない。昨今では、社内外のビッグデータをリアルタイムで分析し、戦略遂行の PDCA を高速に回すことにより、競争優位を獲得する企業もでてきている。

消費者の嗜好の変化や社会の変化に合わせ、試行錯誤しながら動的に正解を見つける SoE には、膨大なデータの活用が必須である。さらに求められるスピードを満たすよう、素早く対応しなければならず、同時にセキュリティやプライバシーの保護も必要である。

日本ユニシスでは、これらのニーズに対応するために、世界最速の Hadoop デистриビューション「MapR」、リアルタイム分析が可能な DWH ソフトウェア「Vertica」を組み合わせ、かつそれらを利用した一連のビッグデータ利活用機能を「汎用データ処理ツール」として仕立て、オンプレミスだけでなくクラウド上からも提供する「データ統合・分析共通 PaaS」をリリースした。本稿では、2章にてビッグデータ利活用を取り巻く市場環境の課題と、解決策としての「データ統合・分析共通 PaaS」の概要を、3章にて「データ統合・分析共通 PaaS」の特徴と利用の流れを紹介し、4章で今後の展望を述べる。

### 2. 情報利活用の課題と解決方法

図1は、顧客との対話の中でよく聞く企業の情報利活用における課題である。主要課題とし

て以下の五つが導ける。

- ・そもそもデータ収集がきちんとしてきてない。
- ・データは蓄積しているが、データベースがサイロ化して、統合的に管理できない
- ・データの有効な活用方法がわからない
- ・専門の技術者が居ない、構築に時間もコストもかかる
- ・ROI がはっきり見えない中、最初から大きな投資はできない

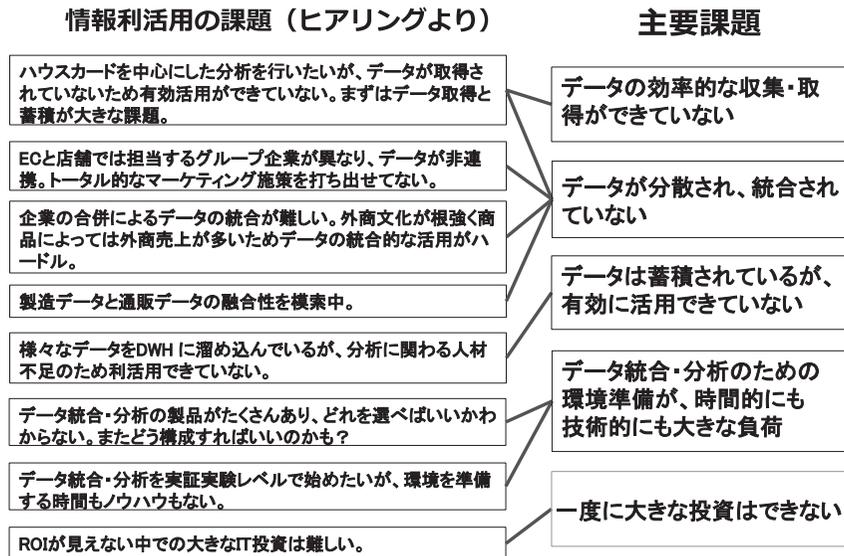


図1 情報利活用の課題

これらの課題の解決に必要な条件は、Speed/Scalability/Cost/Enterprise という四つのキーワードにまとめられる。日本ユニシスは、長年にわたる情報利活用領域でのシステム構築経験に裏付けられたナレッジに加え、強力なアライアンスパートナーとタッグを組むことにより、この四つのキーワードを満足する基盤を提供している。世界最速の Hadoop デイストリビューション「MapR」、リアルタイム分析が可能な DWH ソフトウェア「Vertica」などを組み合わせ、エンタープライズ・クラウド「Microsoft® Azure」などのパブリック・クラウド上からプライベート型 PaaS として提供する「データ統合・分析共通 PaaS」である。主要な構成製品を以下 2.3 節までで説明し、その後「データ統合・分析共通 PaaS」と四つのキーワードの対応を示す。

## 2.1 Hadoop と MapR

Hadoop は、Apache Software Foundation (ASF) が開発・公開している、ビッグデータに適したオープンソースのデータストアであり、データの種類を問わず、あらゆる形式の大規模データを効率的に分散処理・管理することができる。分散ファイルシステムの HDFS (Hadoop Distributed File System)、大量データの並列分散処理フレームワーク Hadoop MapReduce、構造化された大規模なデータを管理する分散データベース HBase などから構成されている。

マップアールテクノロジー株式会社提供の MapR は、企業システムに必要な不可欠な信頼性や運用性、パフォーマンスを考慮し、Apache Hadoop と 100% の互換性を保ちながら、

アーキテクチャを設計・再実装した Hadoop ディストリビューションである。世界最速のパフォーマンスはもとより、高可用性や障害回復、セキュリティ、データ保護などの企業での利用に求められる要件を付加した製品である。また、読み書き可能な NAS (Network Attached Storage) として Hadoop に容易にアクセスできることも特長のひとつである。

## 2.2 Vertica

日本ヒューレット・パッカード株式会社の Vertica は、大量データの分析や検索を超高速に処理することができる列指向型データベースである。汎用的な PC サーバを組み合わせるシステムを構築でき、超並列処理 (MPP: Massively Parallel Processing) アーキテクチャにより、サーバを必要に応じてクラスタに追加していくことで容易にスケールアウトを実現する。かつ、サーバ間でデータを共有することのないシェアードナッシング方式のため、I/O の競合が発生せず、サーバの追加によりリニアにパフォーマンスを向上させることが可能である。また、Database Designer (DBD) と呼ばれる自動チューニングツールにより、手間を掛けずに短時間で最適なデータベースチューニングができ、ユーザの運用負荷を大幅に軽減することができる。

## 2.3 Microsoft Azure

日本マイクロソフト株式会社の Microsoft Azure はオープンで柔軟な、エンタープライズレベルのクラウドコンピューティングプラットフォームである。他社クラウドやオンプレミスを含めたマルチクラウド環境を共通の UI で管理できる。ISO/IEC27018 に準拠し、堅牢かつ厳格なデータ保護と、コンプライアンスに関する情報公開という透明性を併せ持っている。また、Windows ソフトウェアだけでなく、Linux などのオープンソーステクノロジーのサポートを早くから進めており、昨今では Red Hat などの Linux ディストリビューターとの関係も強化し、マルチプラットフォームの環境を提供できることも特長のひとつとなっている。

なお、「データ統合・分析共通 PaaS」は、Microsoft Azure の他、Amazon Web Service 上からの提供も可能となっている。

## 2.4 データ統合・分析共通 PaaS

「データ統合・分析共通 PaaS」により、Speed/Scalability/Cost/Enterprise という四つのキーワードは以下のように満足される。さらに五つ目のキーワード「Managed」を加え、日本ユニシスグループのマネージドサービス<sup>\*1</sup>として提供する。

### 1) Speed

MapR、Vertica が市場の競合製品と比較しても圧倒的なパフォーマンスを発揮することは、その卓越したアーキテクチャの採用のみならず、各種ベンチマークの結果で明らかになっている。

### 2) Scalability

コモディティサーバやパブリック・クラウド上のサーバノードを並列に増やしていくだけで、処理性能がリニアに、正比例にあがるような先進的アーキテクチャを持っている。

### 3) Cost

従来の DWH アプライアンスに代表されるような専用ハードウェアが必要な製品と違い、コモディティサーバやパブリック・クラウド上で稼働させることができ、またクラウド環境

の利用により、夜間など必要なときだけ稼働させることで、投資コストを大幅に抑えられる。またデータ容量としては1TBからのスモールスタートも可能である。

#### 4) Enterprise

MapR、Vertica、Microsoft Azure は、企業での利用に不可欠な高可用性、耐障害性を実現する各種管理機能の実装に加え、セキュリティにも十分に配慮された製品、サービスであり、エンタープライズな用途に安心して適用可能である。

#### 5) Managed

サービスの利用に必要なシステムの運用や管理、導入時に必要なシステム設定などもサービスメニューとして日本ユニシスグループが提供することにより、ユーザ負担を軽減しながら安心・安全なシステム構築・運用を提供可能である。

### 3. データ統合・分析共通 PaaS の特徴

「データ統合・分析共通 PaaS」の基本的な特徴は以下の四つである。本章で一部について詳細を述べる。

#### 1) ベスト・オブ・ブリードな技術・製品を提供

世界で最も付加価値の高い技術、製品を日本ユニシスが目利き、選択し、最適な形で組み合わせ提供。利用者は製品選択、組み合わせ精査、稼働検証が不要となる。

#### 2) 手間をかけずに素早くサービス利用

一連のデータ統合・分析処理に必要な機能を汎用データ処理ツールに仕立て、製品とセットでクラウド上から月額課金で提供する。利用者はシステム開発が不要で、すぐ使い始められ、スモールスタートが可能である。

#### 3) 安心、安全なシステム運用

エンタープライズ用途に必要な運用監視、ジョブ管理、セキュリティ対策も含め、標準的なサービスメニューを提供する。利用者は、迅速なシステム構築、安心、安全なシステム運用が得られる。

#### 4) 業務要求に応じた柔軟なシステム利用環境

クラウド上で利用している PaaS の環境を、そのままオンプレミスの環境に容易に移行することが可能である。利用者はスモールスタートで ROI を見極めたあと、オンプレミスで本格的な情報系システムを構築できる。

#### 3.1 データ統合・分析共通 PaaS の処理プロセス

「データ統合・分析共通 PaaS」では、以下のプロセスにより、利用者が収集したデータの蓄積・加工、統合・分析を行う（図2）。

- 1) 利用者データ、その他データを収集し、生データのまま Hadoop (MapR) に格納する。
- 2) Hadoop (MapR) にて「データ加工・蓄積ツール」を使い、データの形式変換やクレンジングを実施する。
- 3) DWH (Vertica) にて「データ統合・分析ツール」を使って、マスター更新、統合テーブル更新、データ統合、協調フィルタリングによるデータ分析などを実施する。

※ 3) の分析結果は、業務アプリケーションや BI ツール、統計解析ツール等で参照や分析が可能となる。

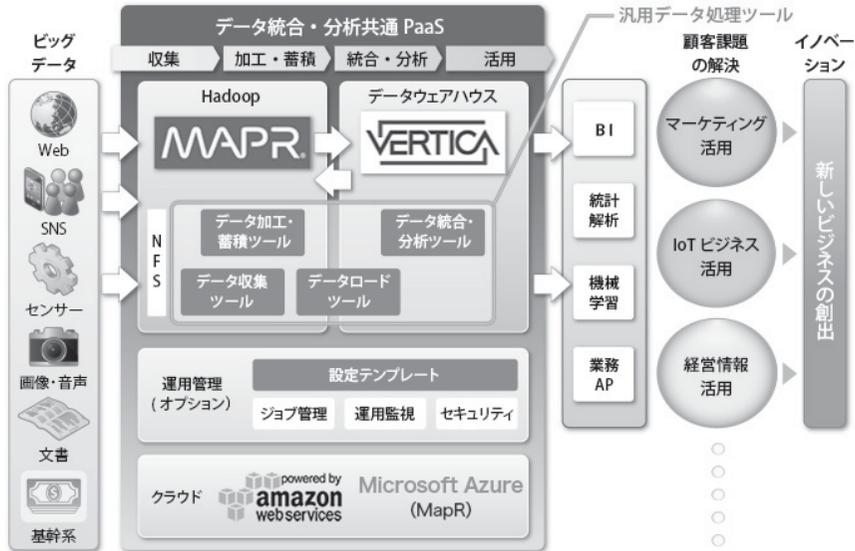


図2 データ統合・分析共通 PaaS の構成と処理プロセス

### 3.2 汎用データ処理ツール

製品を組み合わせただけで前節の処理を行うためには、「MapR」「Vertica」の機能を理解した上でのプログラミングが本来は必要である。それを避けるために、図2の中央に示す「データ加工・蓄積ツール」「データ統合・分析ツール」「データ収集ツール」「データロードツール」をあらかじめ日本ユニシスが「汎用データ処理ツール」として開発し、提供している。このツール群を使うことによって、上記の一連の作業を早期に実現することが可能である。「汎用データ処理ツール」は表1に示すモジュールから構成されている。

表1 汎用データ処理ツールの各モジュールと機能

モジュール名	機能
データ取得	ユーザ企業がアップロードした各種ファイルを取得し、「MapR」サーバの所定フォルダに転送する
データ加工	「データ取得」で取得したファイルに対して形式加工を行い、標準のCSV形式に変換する処理を行う（区分コード設定、変更、データクレンジング処理を含む）
蓄積	整形済ファイルを「MapR」サーバ内で蓄積（バックアップ）する
データロード	整形済ファイルを所定の「Vertica」テーブルにロードする
マスター更新	マスター関連テーブルについて、連携されたデータ内容で全件更新する
統合テーブル更新	連携されたデータ内容で統合テーブルを更新する
データ統合	紐付情報をもとに、マスターデータおよびトランザクションデータを統合する
データ分析	協調フィルタリングを行う。購買履歴などをもとに、行動・購買が類似した他のユーザの情報を用いて推論を行い、訴求の対象とする商品や顧客のリストを作成する

### 3.3 運用管理機能の提供

「データ統合・分析共通 PaaS」は標準的な運用管理機能をオプションサービスとして提供している。データ統合処理などの各種バッチのジョブ管理機能やデータベース等のバックアップ管理機能、各サーバの死活監視やリソース管理といった運用監視機能、またウイルス対策やログ管理などのセキュリティ機能について、推奨製品・サービスと標準的な設定テンプレートをセットで提供し、安心、安全なシステム運用環境を素早く構築することを可能としている。

### 3.4 マルチクラウド対応

「データ統合・分析共通 PaaS」の仕組みは、クラウドからの提供だけではない。クラウドと相性のよい業務には、例えばマーケティングや IoT の領域がある。また、クラウド上で情報系システムをスモールスタートで構築したいというニーズもある。まずはそれらを PaaS 上でスモールスタートして PDCA を回した後、投資対効果が望め、本格的な情報系システムの構築が必要になった場合は、PaaS の環境をそのままオンプレミスやプライベートクラウドの環境、またハイブリッドの環境などに容易に移行することができる。

## 4. まとめと今後の展望

日本ユニシスグループは、2015 年度から始まった中期経営計画「Innovative Challenge Plan」の中で、顧客や社会のデジタルイノベーション、そしてライフイノベーションを支援することを掲げている。これらのイノベーションには、1 章の冒頭で述べた SoE のシステムのように、社内外に散在するビッグデータをいかに早く統合、分析して競争優位をもたらす正解を見つけ出すかが非常に重要となってくる。本稿で紹介した「データ統合・分析共通 PaaS」が、それを支えるビジネス ICT プラットフォームのひとつとなる。2015 年度には、IoT ビジネスプラットフォームにおけるセンサーデータ分析の実証実験、電力業界におけるメーターデータ分析の実証実験、流通業界における大量 POS データ分析の実証実験、決済連動型マーケティングを提供する CLO (Card Linked Offer) サービスにおける顧客情報分析の実証実験などに適用され、様々なイノベーション領域のビジネス創出に貢献している。今後に向けては、その構成要素に、並列分散型統計解析、機械学習、リアルタイムデータ分析、また人工知能関連の

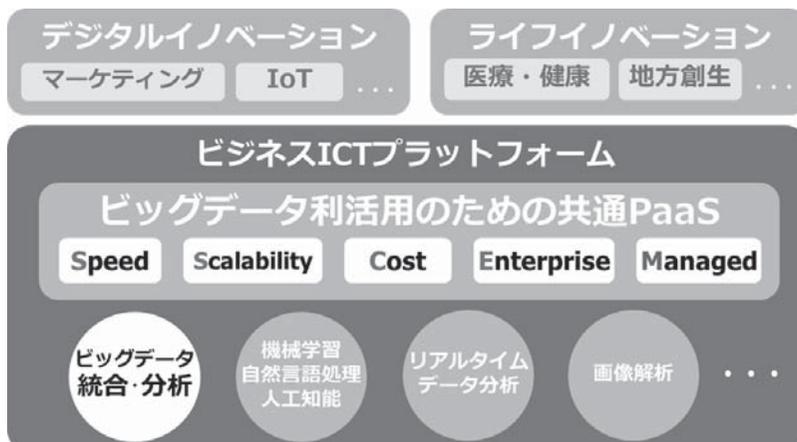


図3 ビジネス ICT プラットフォームに取り込む新しい技術の一例

技術など、次々と出てくる新しい技術を取り込んで（図3）、多様化する顧客のニーズに的確に応えていく所存である。

なお、これらの実証実験は、「ビッグデータ・ラボ」の環境を活用して行われている。ビッグデータ・ラボは、データ統合・分析共通 PaaS などのビッグデータ利活用環境をクラウド上からラボとして提供し、各種新技術の評価や新しいサービスビジネス企画の実証実験などに活用し、新技術の商品化や新ビジネス創出のスピードと品質を向上させることを目的としている。

## 5. おわりに

本稿は、2015年6月4日～5日に東京・赤坂で開催した日本ユニシスグループ主催の総合イベント「BITS2015」で筆者が実施したセッション「データ活用の現実解と未来への展望——もうここまで来た！ビッグデータ統合・分析・活用の最前線」の内容を修正して作成した。2016年の「BITS2016」では「人工知能は電気羊の夢を見るか？——既に起きている未来！アナリティクスが変える暮らしと企業」と題したセッションにて、機械学習や人工知能関連のアナリティクス領域の先進技術への取り組みとそれらの技術がビジネスにもたらすベネフィットについて紹介する予定である。関連するビッグデータ利活用のソリューションも展示ブースにて併せてご覧いただき、日本ユニシスグループの技術力の進歩と継続性をご確認いただければ幸いである。

- 
- \* 1 サービスの利用に必要な機器などの運用や管理、導入時に必要な機器の設置や設定なども一体として提供するサービスのこと。

参考文献 [1] ビッグデータ利活用, データ統合・分析共通 PaaS, 日本ユニシス  
<http://www.unisys.co.jp/solution/biz/bigdata/solution/paas.html>  
(2016年4月22日確認)

### 執筆者紹介 林 直 樹 (Naoki Hayashi)

1986年、日本ユニバック株式会社（現日本ユニシス株式会社）入社。営業部門を経て、CRM、ERP、SCM等の業種横断型業務ソリューション及びDWH、BI、BA等の情報活用ソリューションの企画・販売推進に従事。2010年よりビッグデータ利活用領域の製品/サービス企画・マーケティングを担当。

