

IoT時代の分析開発

ビッグデータを活用するための開発スタイルと環境

ネットワークの普及とデバイス・テクノロジーの進歩により、さまざまなデバイスが直接インターネットに接続できるようになってきました。それに伴い、多種多様なデータが、より早く手に入るようになってきました。IoT (Internet of Things) 技術の発展・向上に伴い、今後もこの傾向はますます加速していくことでしょう。このような大量かつ多種多様なデータ、すなわち、ビッグデータをいかにしてビジネスに生かすか、そのためのシステムをどのように開発するかが、これからの時代の課題となります。

本稿では、ビッグデータの活用に向けたシステム開発のスタイルや環境について解説します。

▶▶ 1. ビッグデータ活用における課題

集めた大量のデータをどう管理するか、そのデータからどんな価値を見いだすかということは、ビッグデータを活用する上で担当者を悩ませる課題の一つです。システム開発において、この課題にどう取り組むべきなのでしょう。

一般的なウォーターフォールの開発スタイルで臨むとすれば、まずは管理対象のデータがどのような性質を持ち、それらをどのように分析するかをシステム化の要件としてまとめる必要があります。では、ビッグデータと

はどのようなデータで、どのように分析すればビジネス価値につなげられるのでしょうか。ビッグデータとは、「市販されているデータベース管理ツールや従来のデータ処理アプリケーションで処理することが困難なほど巨大で複雑なデータ集合の集積物」と定義されています[1]。つまり、大量であってもRDBMS(リレーショナルデータベース管理システム)で扱うことのできるきちんとした構造のデータはビッグデータではなく、構造の複雑性や多様性といった特徴を持ち、事前に構造や種類、精度などのデータ定義が難しいものがビッグデータだということです。

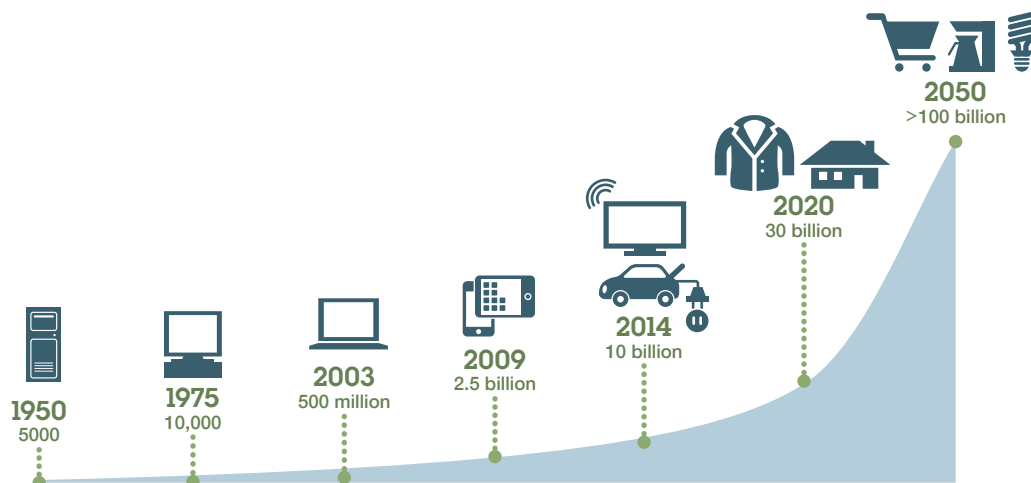


図1. IoTデバイス数の変遷と予測 [2]

IoT技術の発展により、インターネットに接続されるデバイスの数は今後急速に増加し(図1)、ガートナー社の2014年の予測[3]では「2020年に260億に達する」とされています。

IoT時代には、PCやスマートフォンのみならず、多くのセンサーが直接インターネットに接続されるようになります。それらのデータは、構造自体はシンプルであっても、取得データの種類や精度がセンサーごとに異なるだけでなく、連続的に生成されるデータを時系列に扱う必要があります。データの多様性と合わせてインターネットに接続されるセンサーの数を考えれば、これはまさにビッグデータと言えるものです。

自動車もIoTデバイスの一つです。安全・安心な運転環境をドライバーに提供するため、車両の予防的保全のため、渋滞を緩和したスマートな街作りのためなど、さまざまな目的で車両から収集した情報が使われ始めています。

近年の自動車に目を向けてみると、1台の車両には100を超えるコントロール・ユニットが搭載されており、車内ネットワークを流れるデータの項目数は数百にもなります。それらを10ミリ秒(0.01秒)から100ミリ秒(0.1秒)という周期で収集すると、1台の車両で生成されるデータは1時間当たり数万~数十万といったレコード数になります。さらにこれらのデータはデバイスごとにデータ項目が異なるだけでなく、収集の設定によって取得するデータ項目や精度、さらにはデータ項目間の関係が変わるため、事前にきちんとしたテーブル設計がしにくいといった特徴があります。

このようなビッグデータを扱う際には、RDBMSだけではなく、非構造データを扱いやすいNoSQLデータベースと組み合わせてデータ管理を行うなどの工夫が必要でしょう。

2. ビッグデータの分析

ビッグデータの活用というとき、データマイニングによる知識の獲得を想像することが多いのではないのでしょうか。データマイニングとは、データの中に存在している未知のパターンを見だし解釈することで、ビジネス上で直面している課題に対処するプロセスです[4]。ビッグデータからビジネス価値を創出するためのアプローチとして

期待が高まっているのは当然のことでしょう。

データマイニングを成功に導くためには、分析手法の選択が重要なポイントとなります。しかし、その手法を決めるのは時間のかかる作業です。最適な分析手法を見つけるためにはデータ・サイエンティストとドメイン・エキスパートの協業が不可欠であり、問題定義からデータの調査、データ準備、モデリング評価というサイクルを何度も回すこととなります(図2)。従って対象領域の専門性が高ければ高いほど、複雑で時間のかかる作業となります。

しかし、データマイニングだけがビッグデータ分析への唯一のアプローチなのでしょうか。そうだとすれば、適切な分析手法を見極め、価値ある分析結果を手に入れるまでに、多くの試行錯誤を重ねなくてはならなくなってしまいます。

製造業のあるお客様では、製品の品質向上に向け日々製品のデータを収集し動作状態を分析・調査していました。しかし製品の種類が増え、それらが世界各国で使われるようになるのに伴い、分析はもとより収集データの管理も困難となってきました。そこで、分析作業の省力化を目指し、収集データの一括管理と分析実行を自動的に行うシステムの構築に着手されました。このときシステム化すべき分析手法はすでに開発部門が持っており、すぐにシステム開発を開始することができました。分析ロジックの実装においてはユーザーとシステム開発者で議論を重ね、何をパラメータとして扱えばそのロジックをより広範囲に適用できるかを検討した上でシステムに反映し

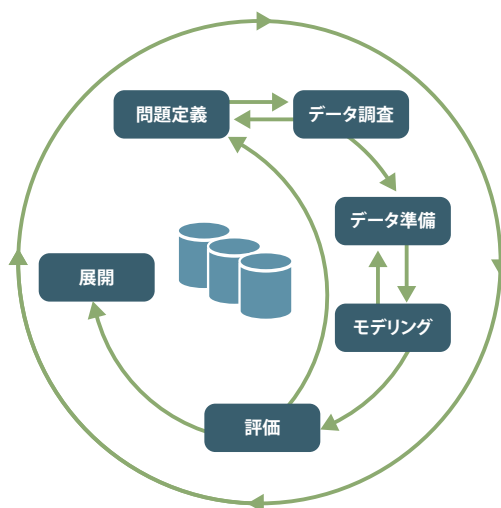


図2. CRISP-DM(Cross-Industry Standard Process for Data Mining) プロセス・モデル

ました。その結果、分析活動の省力化を実現できただけでなく、実績のある分析手法をより多くのデータに適用することができるようになり、得られた結果データからさらに新たなデータ活用方法を発見することもできました。

ここにビッグデータ活用方法のヒントがあります。つまり、大量に集めたデータから価値を発掘するための分析方法を見つけ出すというアプローチだけでなく、手元にある分析手法を元にしてシステムを見直しながら改善を施し、その適用範囲を広げていく、というアプローチもビッグデータ活用における有効な手段と言えるのです。

▶▶ 3. アジャイルな開発スタイルへ

実績のある分析手法をより多くのデータに適用し、効果を確認しながらシステムや分析手法を拡張・改善していくというアプローチは、PDCA(Plan-Do-Check-Action)による継続的な改善のサイクルです。近年、こういった改善を重ねてシステムを拡張していく開発スタイルとしてアジャイル開発が注目・活用されています。アジャイル開発の焦点は、「変化していく要求を認めサポートする一方で、継続的に動くソフトウェアを出荷していくこと」です。まさにIoTのように変化の激しい領域において、改善のサイクルを回しながらシステムを開発するのに適した開発スタイルと言えます。図3はアジャイル開発の一つの方法であるスクラムのプロセスを示しています。スプリントと呼ばれる2~4週間の開発工程を回し、繰り返し機能をリリースしていきます。

ウォーターフォール開発スタイルからアジャイル開発

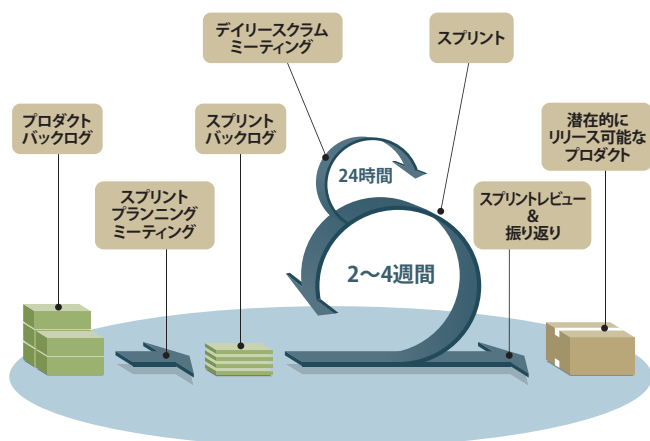


図3. スクラムのプロセス [5]

スタイルへの移行には多くの変化を必要とします。チームは計画に合致していることよりも変更への対応能力を重視し、詳細な要求仕様や設計に時間をかける代わりに、顧客の要求を表現するストーリーと呼ばれる単位で機能群をリリースすることに注力します。そのためには優先順位の高い機能から順に実装し、継続的なインテグレーションを行うこととなります。ユーザーと協業し、スコープではなく常にスケジュールを優先して作業を進めることが重要です[6]。

アジャイル開発により、システムは継続的に改善・成長していくこととなります。従ってこの開発を成功に導くには、単にアジャイル開発手法の適用だけではなく、改善・成長を支える開発および実行環境の利用・活用も重要な要素となります。

▶▶ 4. クラウドによる開発と「IBM Bluemix」

クラウド環境を利用することで、あらゆるリソースを必要な時に入手して利用するというJust In Time(JIT)の開発が可能となります。実行環境や開発環境、ツールのほか、システムに必要とされるさまざまな機能やサービスを得られるため、初期投資を抑えてすぐに動くものを作ることができます。特にデータ管理や分析に必要な機能は、試行錯誤を重ねる中で変更が発生しやすい部分でもあるので、サービスの選択肢が多く、構成が容易に行えるクラウド環境が望まれます(図4)。

「IBM Bluemix」[7]はIBMクラウド上で提供されるPaaS(Platform as a Service)環境です。Webアプリ

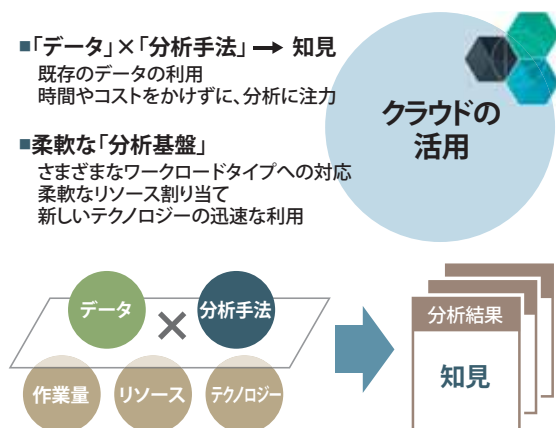


図4. 分析に必要な要素

表1. ビッグデータに関するIBM Bluemixのサービス

サービス名	サービス概要
Biginsights for Hadoop	クラウド上でエンタープライズレベルのHadoopクラスター環境を提供するサービスです。
dashDB	データ・ウェアハウスおよび分析に最適化されたデータストア機能を提供するサービスです。
Geospatial Analytics	デバイスからアップロードされた位置情報を基に特定のエリアへの出入りをイベントとして処理します。
IBM Analytics for Hadoop (β)	Hadoopベースのクラスターを提供します。大規模で複雑な分析を実行するためのエンタープライズ機能を備えたIBM Biginsightsを使用しています。
IBM Insights for Twitter	IBMソーシャル・メディア分析の高度な自然言語処理アルゴリズムに基づいた、感情その他の拡張情報を複数の言語で提供します。Twitterデータ・ストリームのリアルタイムでの処理が完全にサポートされています。
Time Series Database	専用の高速かつ効率的なストレージと時系列データを分析するためのサービスです。

ケーションのためのサービスだけでなく、IoTやビッグデータ分析に関するサービスも提供されており、システム開発者は必要な機能を必要な時に利用できます(表1)。

IBM Bluemixは実行環境の構築も容易で、Webアプリケーションであれば1時間ほどで実行環境を構築することが可能です。環境の保守作業からも開放されるため、開発者はシステム開発に注力することができます。

さらに、DevOpsを支援する機能も提供しています。DevOpsは開発部門と運用部門のメンバーが連携・協力する開発手法であり、継続的なデリバリーを実現するためには欠かせない要素です。IBM Bluemixは、まさにアジャイル開発スタイルとビッグデータ分析システム構築に必要な環境と言えるでしょう(図5)。

5. ビッグデータの活用に向けて

IoTをアジャイル開発スタイルで活用し、ビッグデー

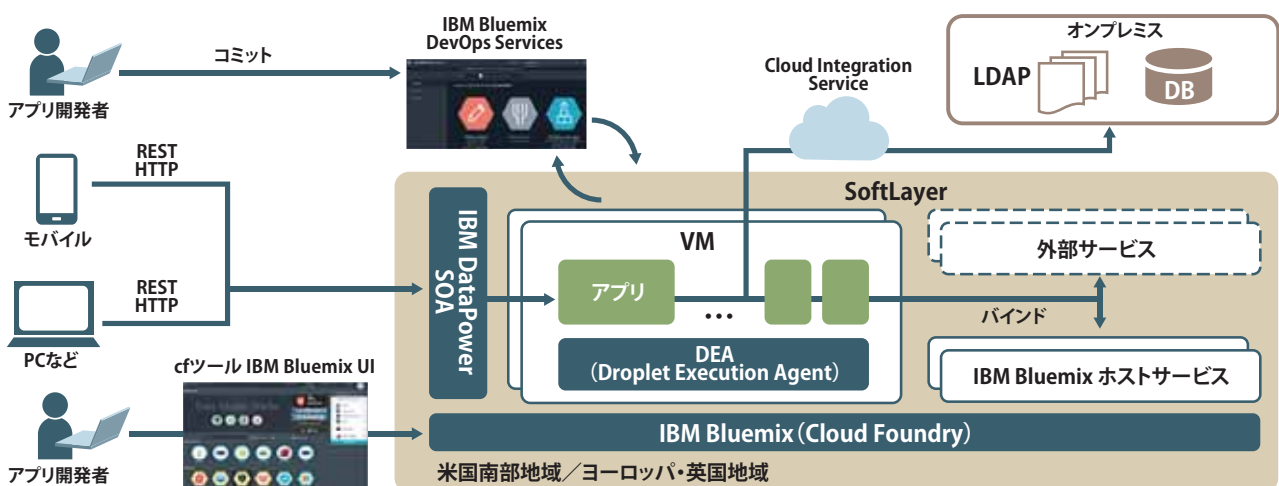


図5. IBM Bluemixのアーキテクチャー

タの分析にぜひチャレンジしてください。IBMは開発手法から開発・実行環境の提供、設計開発の実施まで、幅広くシステム開発をご支援します。

もし、何のデータを集め、どのように分析すればよいのか悩んだ場合には、まずは社内に目を向けてみてください。大量のデータのみならず、実績のある分析手法が皆さんのすぐ近くで見つかるはず。大きく構え過ぎず、まずは手元のデータと分析手法を活用してみることをお勧めします。環境はすでに整っています。小さな一歩を踏み出すだけで、より多くの知見や価値を得ることができることに気付かれるでしょう。

[参考文献]

- [1] Wikipedia,ビッグデータ, <https://ja.wikipedia.org/wiki/%E3%83%93%E3%83%83%E3%82%B0%E3%83%87%E3%83%BC%E3%82%BF>
- [2] The Internet of Things, <http://www-01.ibm.com/software/jp/info/internet-of-things/>
- [3] The Impact of the Internet of Things on Data Centers, Gartner, inc., <http://www.gartner.com/doc/2672920>, 2014
- [4] Arlene Zaima:データマイニング入門, TERADATA, http://jpn.teradata.jp/solution/images/TDMK5022_0710_DataminingIntroductory.pdf
- [5] An Overview of Scrum for Agile Software Development, Mountain Goat Software, <http://www.mountaingoatsoftware.com/scrum/overview>
- [6] ディーン・レフィンゲル:アジャイル開発の本質とスケールアップ, 翔泳社(2010)
- [7] IBM Bluemix, <https://console.ng.bluemix.net/>



日本アイ・ビー・エム株式会社
グローバル・ビジネス・サービス事業
クラウド・サービス
マネージング・コンサルタント

山藤 和人
Kazuto Yamafuji

1989年日本IBM入社。大型システムのソフトウェア調査に従事した後、大和研究所にて組み込み機器のソフトウェア開発に従事。現在はIoT関連のソフトウェア設計開発に関するコンサルタントとして活動。