

2013年12月

A large, stylized version of the SJI logo, consisting of a central dark blue circle and four curved, wing-like shapes in dark blue and yellow extending outwards.

ビッグデータとは？

先端技術研究センター

李 水平

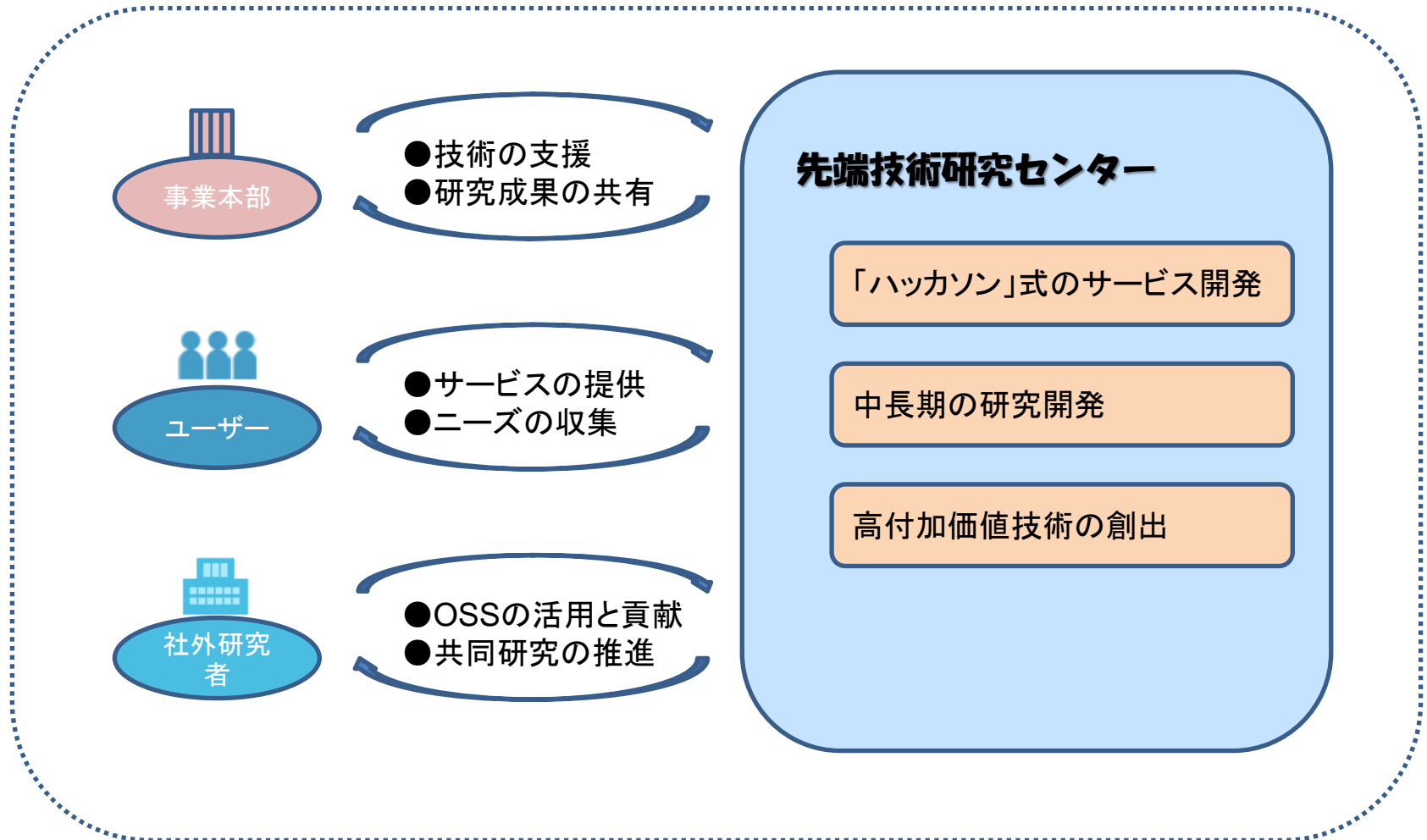
株式会社 **SJI**

- ◆ **ビッグデータとは、何か？**
- ◆ **なぜ、このタイミングで「ビッグデータ」なのか？**
- ◆ **ビッグデータ活用の事例**
- ◆ **今後注視すべき動向**
- ◆ **ビッグデータの活用に関する課題**
- ◆ **ビッグデータの活用に関する主な技術**

目指すもの:

顧客のために最新技術に革新的な
アイデアを加えて違う何かを生み出し
新たな価値創造につなげていく

研究開発方針:



1、「ビッグデータ」とは

ビッグデータ活用を経営視点で考える



ビッグデータ活用は重要な国家戦略

— 米国など“先進国”に後れをとるビッグデータ活用でどう巻き返していきますか。

山本：まず政府としてビッグデータの活用が世界の潮流となっていることを認識しています。

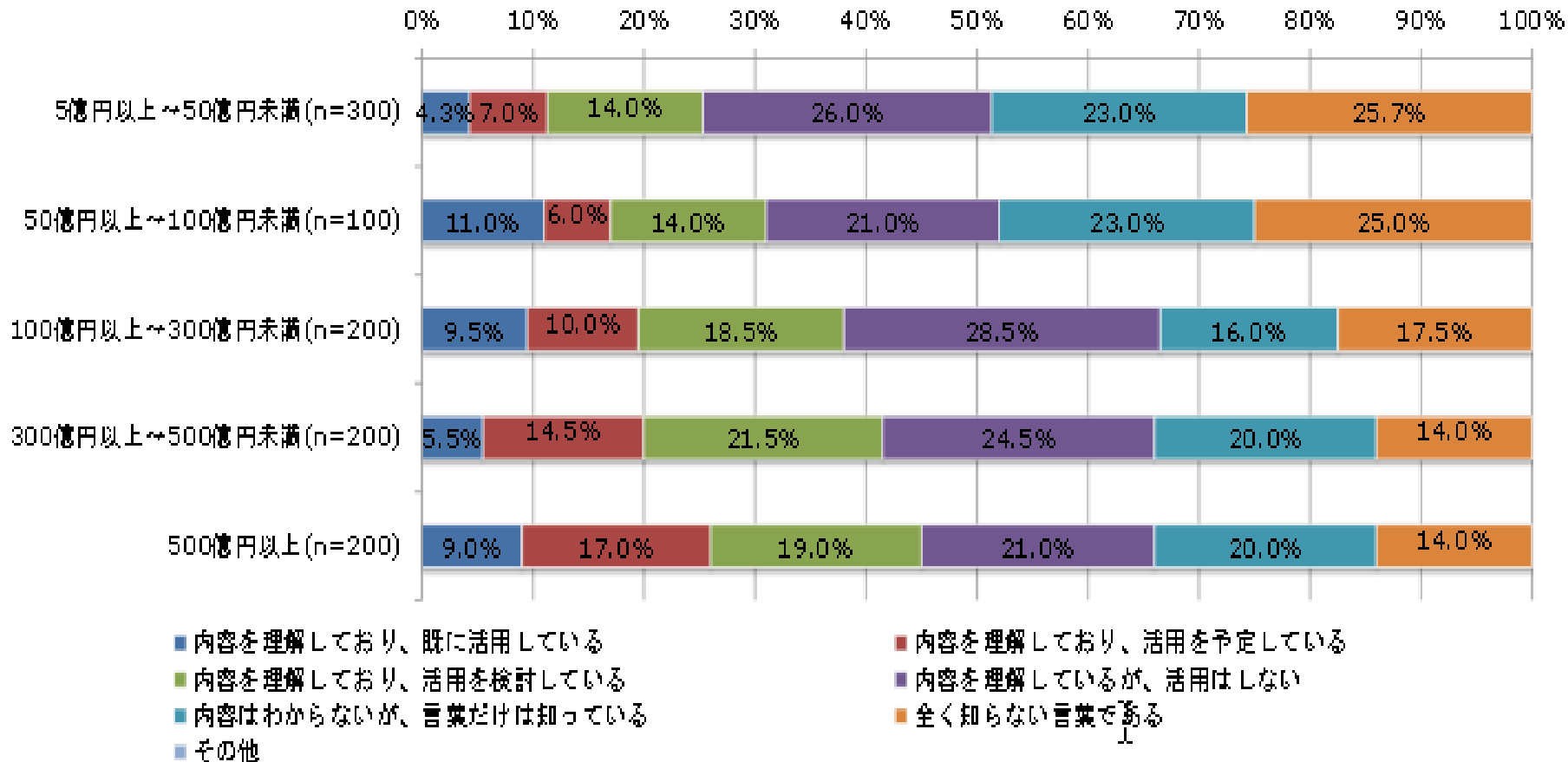
例えば、2013年9月に世界経済フォーラムが中国・北京で開いた「サマー・ダボス会議」に参加しました。その中で「Big Data or Big Hype?」と題して、ビッグデータがしっかりと世界のトレンドを変えていくものか、それとも一時の流行かというテーマでパネリストとして登壇しました。4人のパネリストが2グループに分かれて、我々は「一時の流行ではない」という立場で論陣を張りました。



山本 一太氏
内閣府特命担当大臣・情報通信技術（IT）政策担当。1958年群馬県生まれ。82年中央大学法

1、「ビッグデータ」とは

D3-2-1.「ビッグデータ」という言葉の認知



対象企業： 日本全国／全業種の年商5億円以上～500億円未満の中堅・中小企業
および年商500億円以上の大企業

調査会社： ノークリサーチ 2013年1月～2月

「ビッグデータ」とは何か？

**ビッグデータ
= データ量???**

ビッグデータ
= 3つのV

1つ目のV

Volume

(データ量)

1、ビッグデータとは



ビッグデータは、これまでのデータベースとはデータの規模が根本的に異なる

Volume



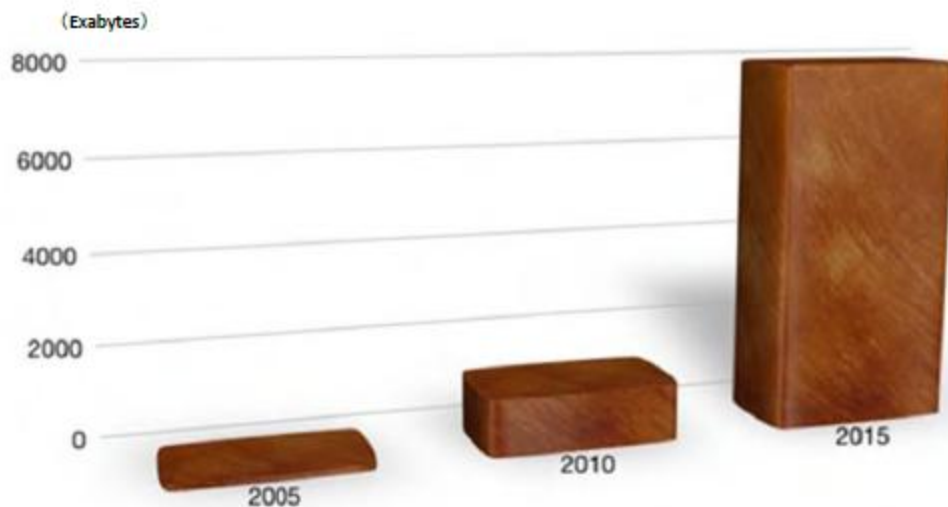
のビジネス変革
あらゆる企業に

1、ビッグデータとは



- 国際的なデジタルデータの量は、2011年の約2ゼタバイト(2兆ギガバイト = 2千エクサバイト)から約4倍増加し、2016年には約8ゼタバイトへ拡大する見込み。
- ビッグデータの活用により、例えば、米国ヘルスケアで年間3千億ドル、EU公共セクターで年間2.5千億ユーロ、位置情報データの活用により年間6千億ドルの消費者価値創出等が期待。

10年間のデジタルデータの成長



いわゆる「ビッグデータ」の定量的価値(例)

50億台の携帯電話が使用(2010年)

300億のコンテンツが毎月Facebook上で共有

IT費用の5%増加で、年間40%増のデータ創出

米国のヘルスケアでは年間3000億ドルの価値創出が期待(スペインの年間ヘルスケアコストの2倍)

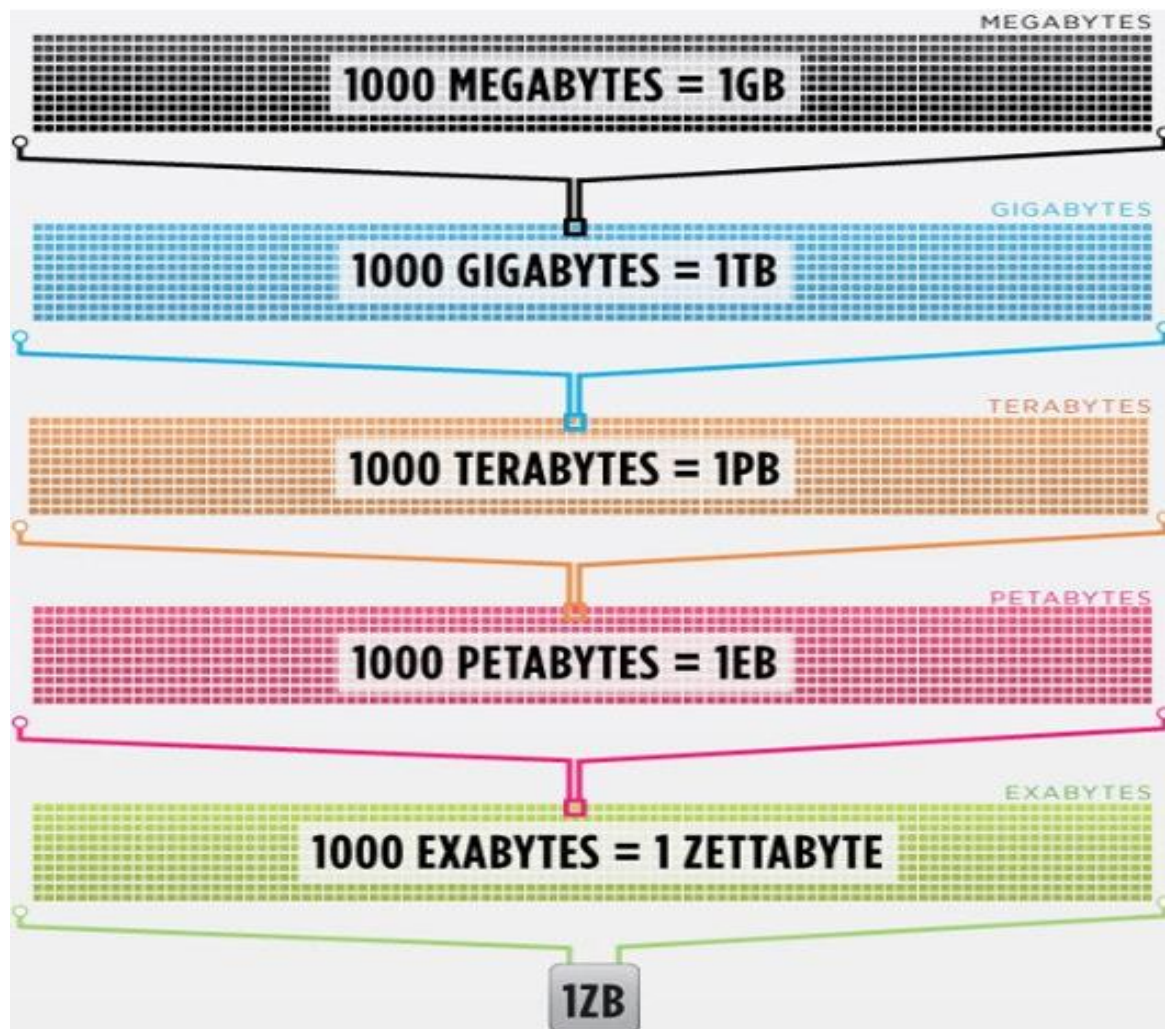
EUの公共セクターでは年間2500億ユーロの価値創出が期待(ギリシアのGDPを超える)

個人の位置情報データを活用することで年間6000億ドルの消費者価値創出が期待

1、ビッグデータとは



参考：そもそもゼタバイトってどれくらい？



ゼタバイトは 10^{21}

「DVDで2500億枚」
に相当

2つ目のV Velocity (データ到達速度)

1、ビッグデータとは



Velocity

無線ICタグ
13億個→300億個

膨大な
センサー情報

巨大な情報ストリーム
バッチ処理からリアルタイム処理へ

The graphic features the word "Velocity" in a large, white, sans-serif font against a dark grey background. Below the word, there are several text elements and a small image. On the left, there is a blue curved bar. In the center, there is a small image of a white circular sensor with a black antenna, with the text "無線ICタグ" (Wireless IC Tag) and "13億個→300億個" (1.3 billion to 3 billion) next to it. Below this image is the text "膨大なセンサー情報" (Massive sensor information). On the right, there is the text "巨大な情報ストリーム" (Massive information stream) and "バッチ処理からリアルタイム処理へ" (From batch processing to real-time processing).

3つ目のV Variety (データの種類)

1、ビッグデータとは

ソーシャルメディアデータ

- ☞ ソーシャルメディアにおいて参加者が書き込むプロフィール、コメント等



マルチメディアデータ

- ☞ ウェブ上の配信サイト等において提供等される音声、動画等



ウェブサイトデータ

- ☞ ECサイトやブログ等において蓄積等される購入履歴、ブログエントリー等



カスタマーデータ

- ☞ CRMシステムにおいて管理等されるDM等販促データ、会員カードデータ等



ビッグデータ

ICT (情報通信技術) の進展により生成・収集・蓄積等が可能・容易になる多種多量のデータ (ビッグデータ) を活用することにより、異変の察知や近未来の予測等を通じ、利用者個々のニーズに即したサービスの提供、業務運営の効率化や新産業の創出等が可能。

センサーデータ

- ☞ GPS、ICカードやRFID等において検知等される位置、乗車履歴、温度、加速度等



オフィスデータ

- ☞ オフィスのPC等において作成等されるオフィス文書、Eメール等



ログデータ

- ☞ ウェブサーバ等において自動的に生成等されるアクセスログ、エラーログ等



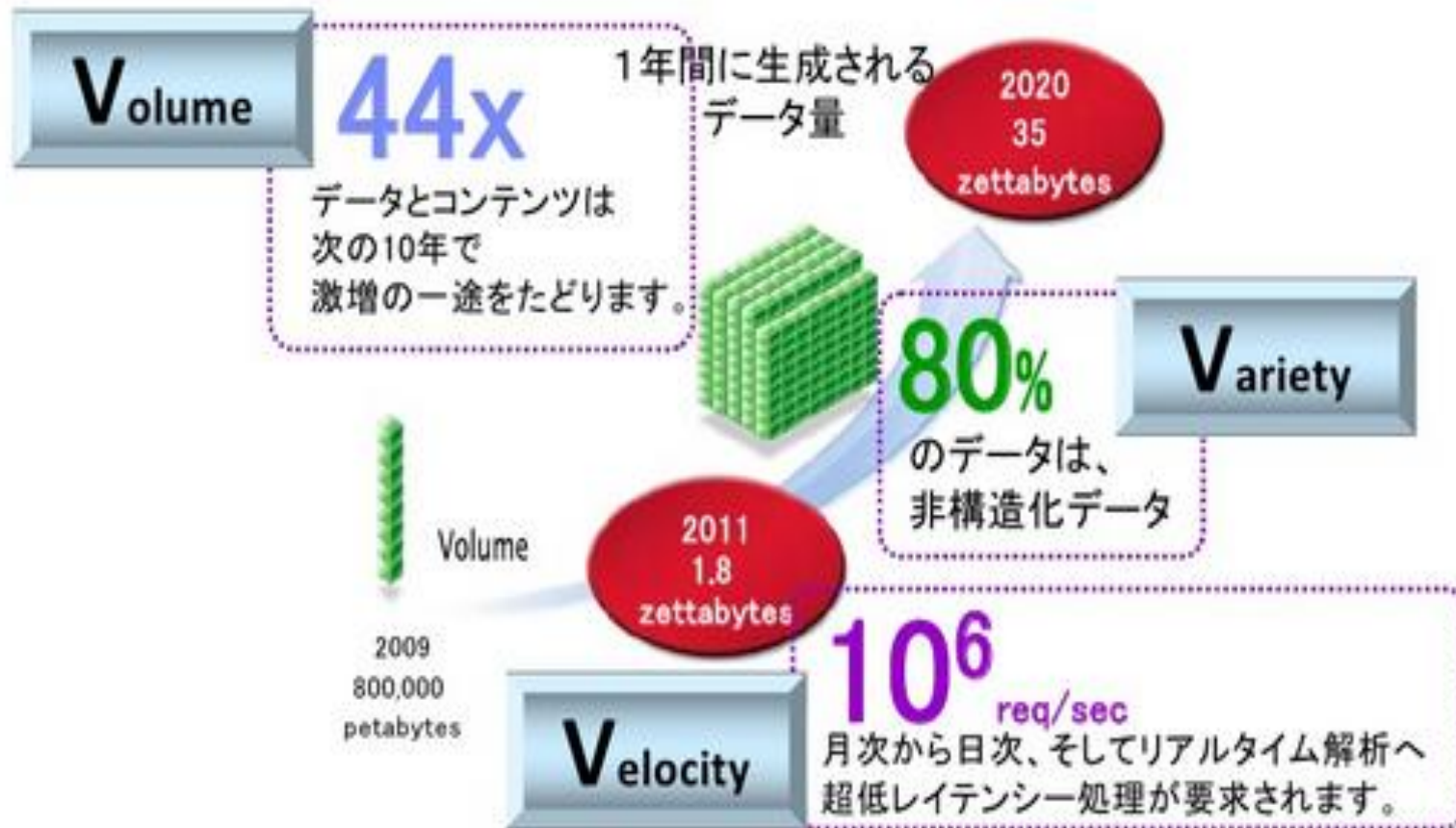
オペレーションデータ

- ☞ 販売管理等の業務システムにおいて生成等されるPOSデータ、取引明細データ等



1、ビッグデータとは

インターネットの登場以来、人間社会で生成されるデータ量は加速度的に増えている。これらデータには、単に量が多い(Volume)だけでなく、種類の増加(Variety)、速度の高速化(Velocity)という「三つのV」という特徴がある。



1、ビッグデータとは



参考：誰がビッグデータと言いだめたのか？

1997

John R. Mashey
Chief Scientist, SGI

2008

Nature ビッグデータ特集

2010

英Economist誌
大量データ活用特集

なぜ、このタイミングで
ビッグデータなのか？

2、なぜ、このビッグデータで？



**理由1： この10年間の「電子化・自動化」の進展により
使えるデータがたまってきたため。**

- 2001年以降の10年間に於いてどのような変化があったのか？
- 2002年のITバブル崩壊の印象が強すぎるが、確実に電子化・自動化は進展していた。

Suica

2001年交通系サービス開始
2004年ショッピングサービス開始

FOMA
サービス

FOMAサービス開始

GPSの標準搭載は2004年～
当時は500円玉サイズ、
1ユニット17万円。
(今は数百円)

iPod

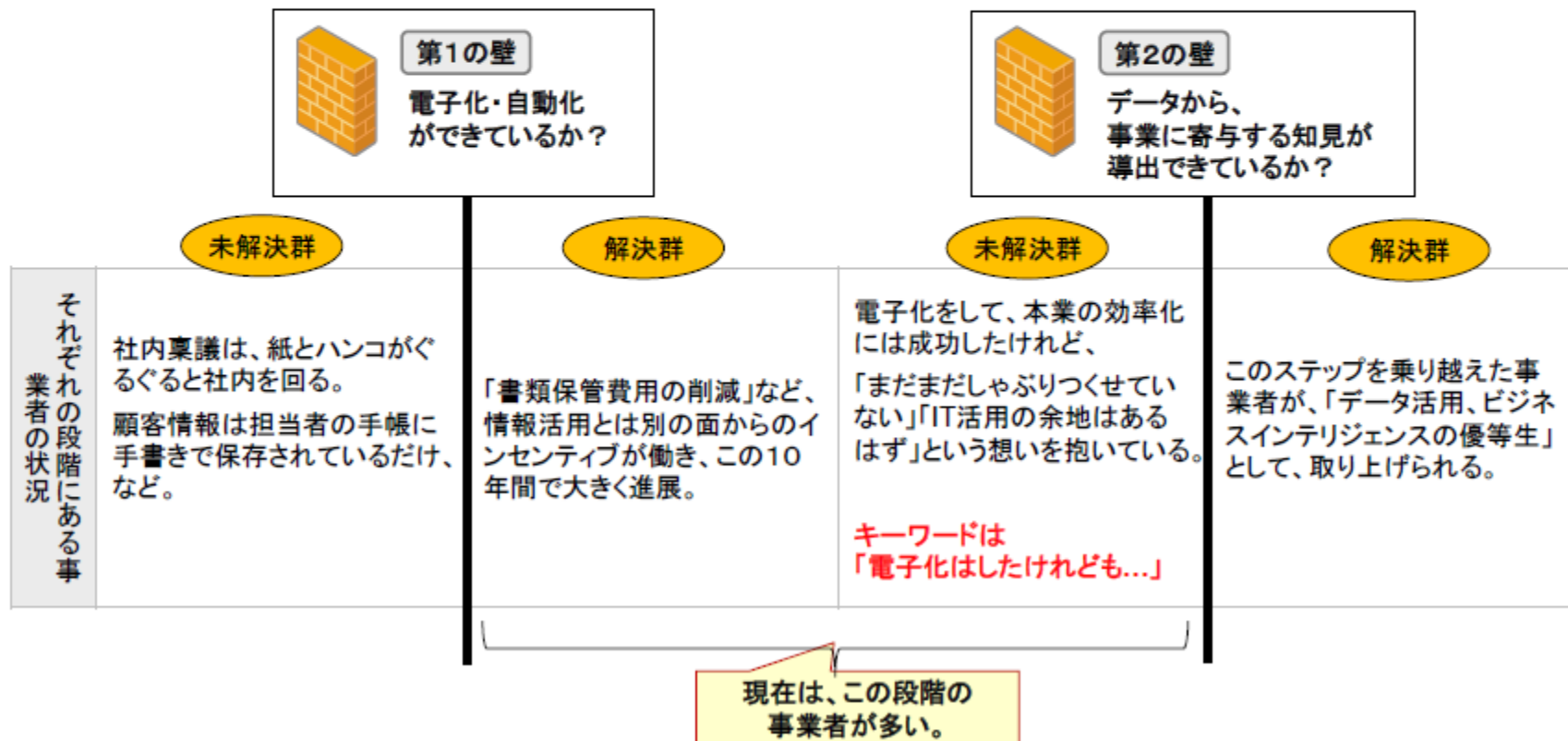
iPod 第1世代
登場

2、なぜ、このビッグデータで？



理由2: 利用サイド事業者にて、IT活用の段階を進めざるを得ない競争状態に入ったため。

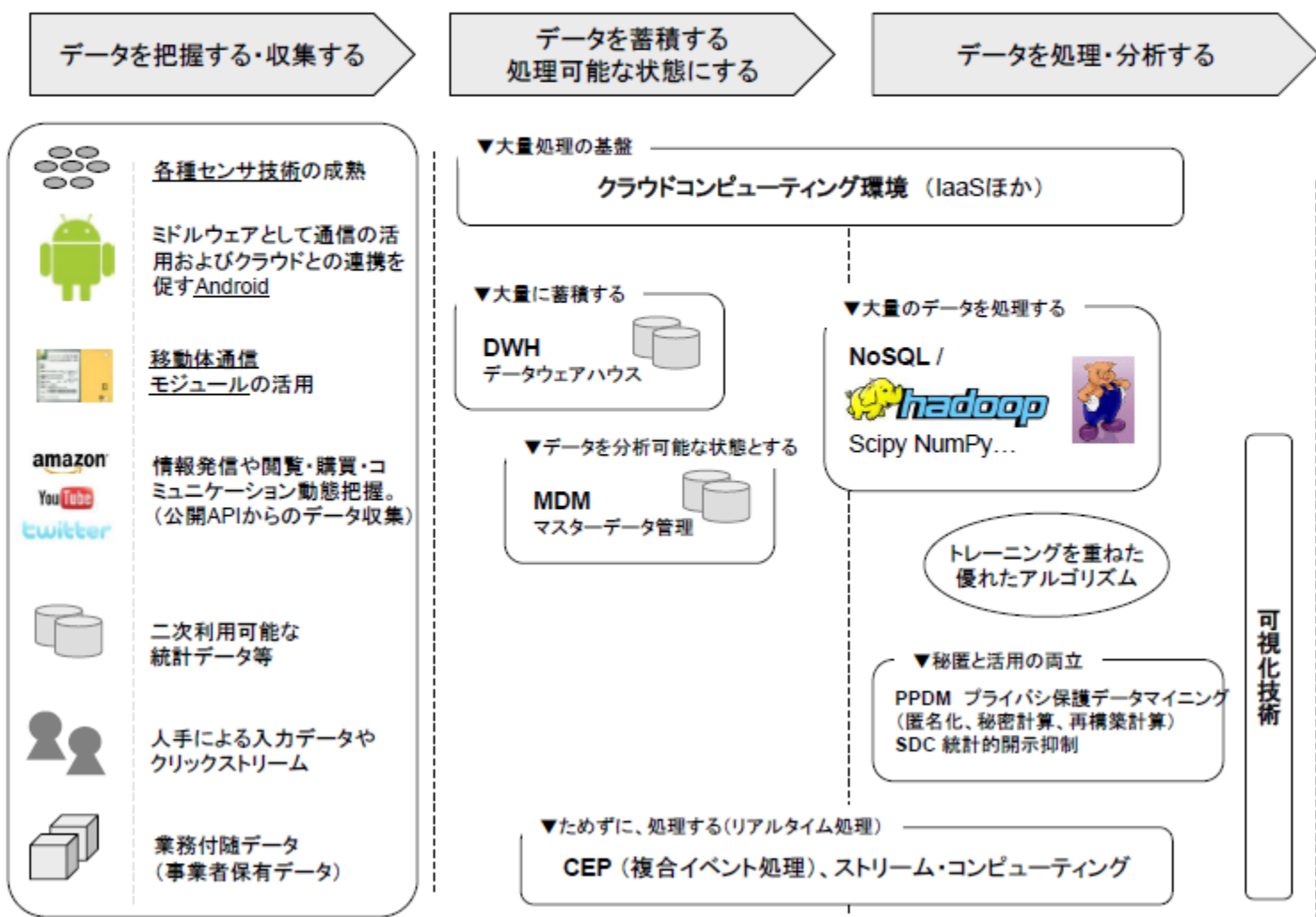
- 「IT活用」「情報化」は一つの意味ではない。
- 「電子化・自動化」ができていないことと、「事業に寄与する知見導出」ができていないことには大きな隔りがある。
- 同業種の中で「第二の壁」を越えた事業者と越えてない事業者が勝負しようとしたとき、競争力に大きな差が生ずる。



2、なぜ、このビッグデータで？



理由3: ビッグデータを「取得・生成」「蓄積」「処理・分析」するためのツールが成熟してきたため。



ビッグデータ活用の事例

ビッグデータの活用は、事業者に対してどのような付加価値を与えるか？

- 将来的にどのような施策を講ずれば良いのかがわかり、事業の効率的な実施を行うことが可能となる。
- 具体的には次のような業務プロセスにおいて効用を得ることができる。

- 「製品開発」において、どのような製品を開発することが消費者に対して訴求するのか、ということが分かる。
- 「販売促進」において、誰に、何を、いつ売れば良いのかが分かる。
- 「保守・メンテナンス・サポート」において、いつ、どのようなメンテナンスを行えばよいか分かる。(総コストの低減も)
- 「コンプライアンス」において、不正の予兆や、特に注視すべき事象が何であるかがわかる。
- 「業務基盤・社会インフラの運用」において、全般的な性能向上・コスト削減が実現される。

3、ビッグデータ活用の事例



SoftBank | SoftBank ホーム

▶ プラチナバンド特設サイト トップ | ▶ プラチナバンドの特長 | ▶ プラチナバンドの広がり | ▶ プラチナバンド対応製品 | ▶ CMギャラリー | ▶ つながりやすさNo.1へ

つながりやすさ No.1へ

No.1
号外篇
CM公開中▶

「錯覚かもしれない。調査の失敗かもしれない。
通話発信をして「つながったか」を調べる調査で
ソフトバンクが No.1 という結果がでたのです。」
(2013.01.31 ソフトバンク決算説明会にて)

3、ビッグデータ活用の事例



収集データ元アプリケーションのご紹介



チェッカーシリーズに実装

アプリの通信可否を位置情報と共に蓄積



全キャリア合わせて
330万DL



アプリ通信ログ数
7.5億件/月
2,500万件/日

リリース済みアプリケーション

iPhone

Android



パケット接続率調査 (特許出願済)



月間**7.5億件**の
アプリ通信ログ



※ユーザー許諾取得済

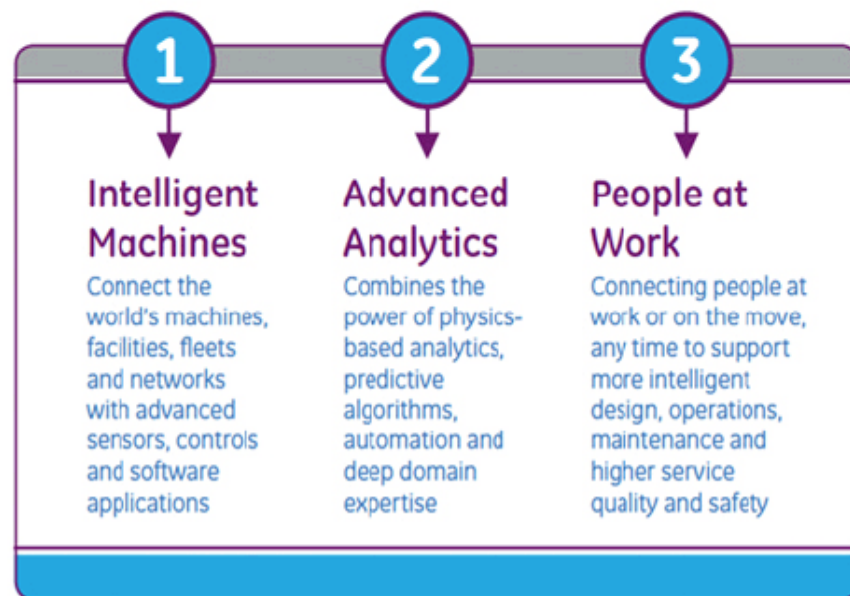
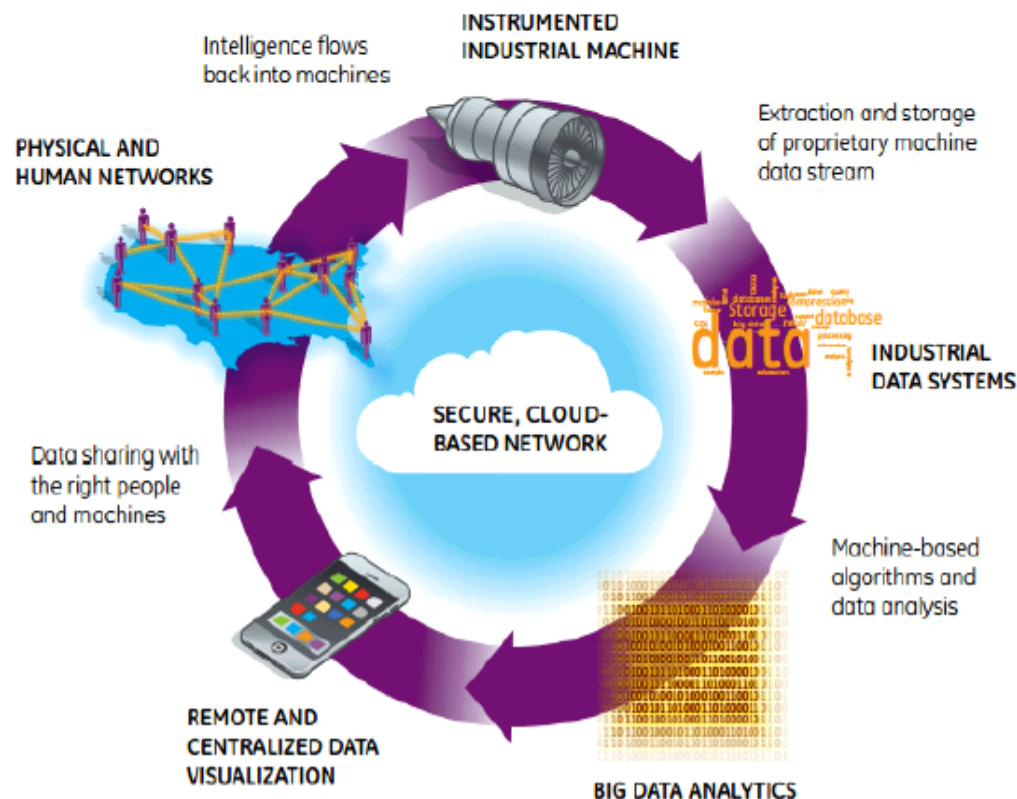
This is the Big Data

3、ビッグデータ活用の事例



製造業 GE「Industrial Internet」構想を発表(2012年11月)

- GEの製造する各種機械に多数のセンサーを取り付け、ネットワーク経由で状態を常に監視し、センサーデータの分析により故障等を事前に予測する仕組み。
- たとえば、ある飛行機のジェットエンジンの場合、出発遅延に繋がるようなメンテナンスが必要な状態を70%の確率で特定可能なアルゴリズムを開発



3、ビッグデータ活用の事例

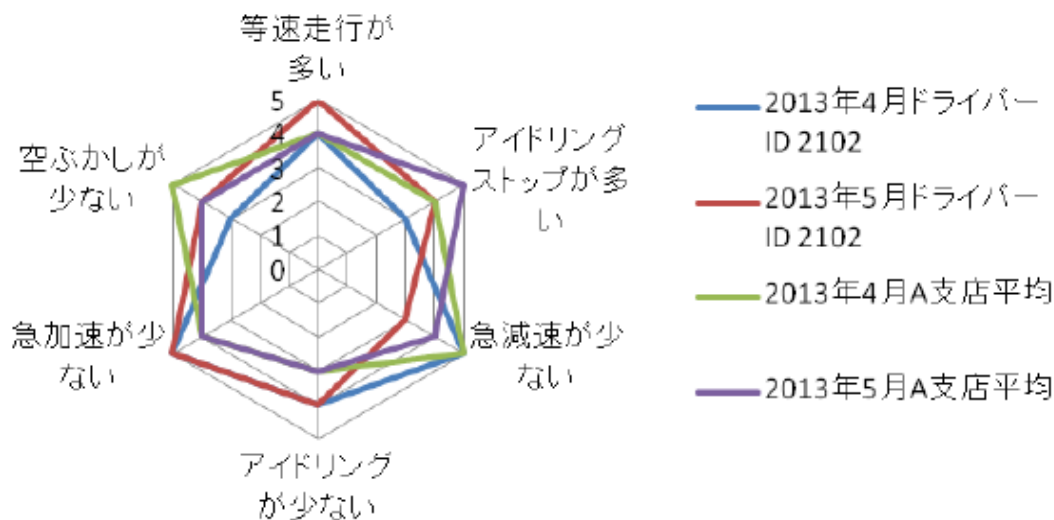


運送業 U.S. Xpress フリート管理システム

- 22,000台のトレーラー、8,000台のトラクターのタイヤ、エンジン、ブレーキなどに取り付けられた数万のセンサーデータ(約900項目)をHadoopデータベースに集約し、リアルタイム分析
- 年間1千億件以上にのぼる膨大なデータ分析を分析し、**年間約600万ドルのコスト削減に成功**
 - ドライバーのアイドリング時間を測定し、個別指導することで、初年度に2000万ドルの燃料コストの削減に成功



<フリート管理のイメージ
(エコドライブレーダーチャート)>

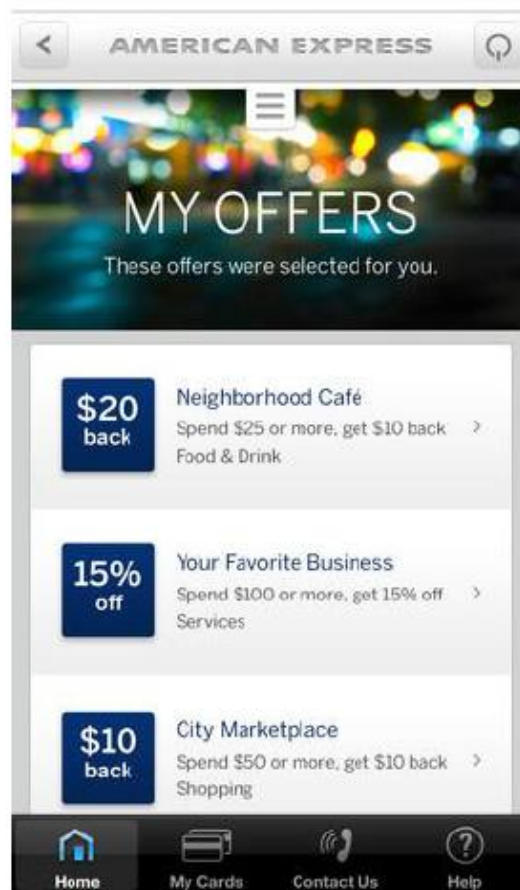


3、ビッグデータ活用の事例



金融業 Amex MyOffers(2012年5月～)

- American ExpressのiPhone用アプリ。アプリをダウンロードした**顧客の決済履歴と現在の位置情報に応じて、加盟店からのオファー**がランク付けされて提示される
- 加盟店には、モバイル用のオファーの作成ツールと効果の測定レポートが提供される



＜モバイル用オファーの作成ツール＞



＜オファー結果の測定レポート＞

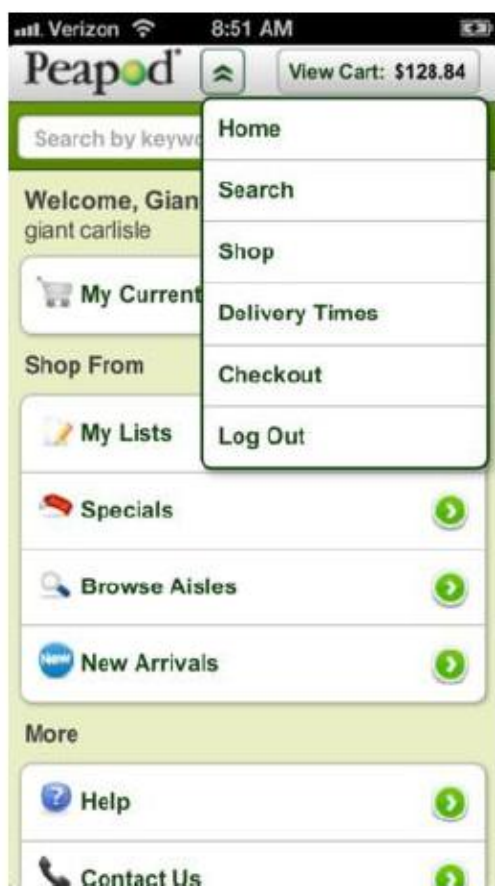


3、ビッグデータ活用の事例



小売業 ショッピングアプリ オンライン食品店Peapod(1)

- 年商240億ドルの食品小売大手Ahold社傘下のオンライン食品店Peapodは、モバイルアプリ「PeapodMobile」を2012年9月にリリース
- 買い物リストの作成、商品の原材料やカロリーの確認、配達日時指定等の機能を持つ

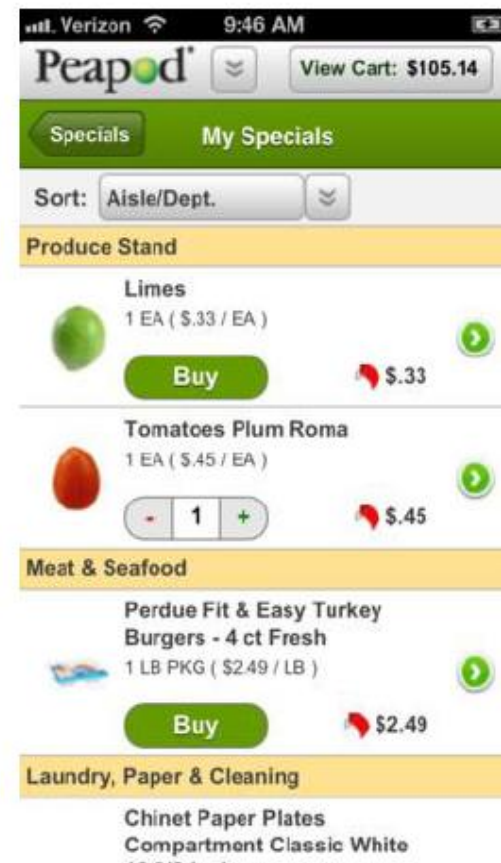
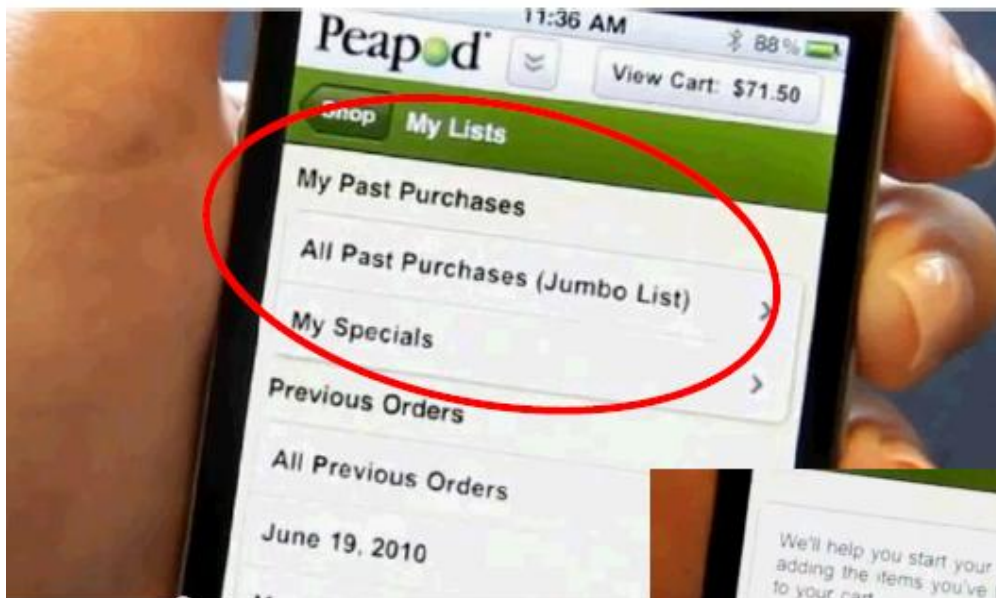


3、ビッグデータ活用の事例



ショッピングアプリ オンライン食品店Peapod(2)

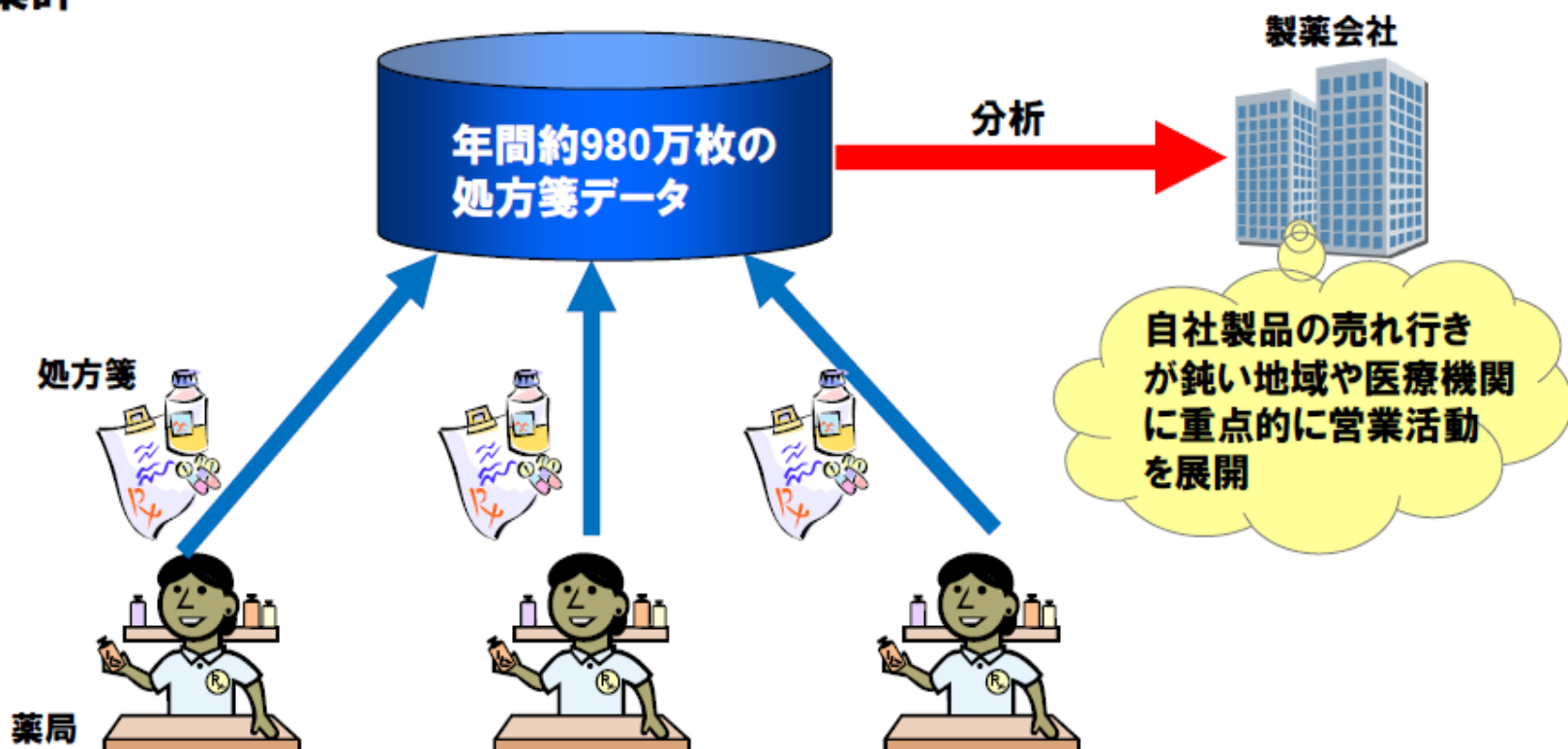
- 過去の購買履歴を分析し、本日のセール商品をお薦めしたり、頻繁に購入する商品から、顧客が必要としている商品を提示する「Guess My Order」など多彩な機能を持つ
- ⇒顧客の利便性とブランド・ロイヤリティの向上が狙い



3、ビッグデータ活用の事例

薬局業界 日本調剤

- 調剤薬局大手の日本調剤は全国約430店舗で扱う年間約980万枚の処方箋の情報を分析、「医療機関や診療科ごとに、どの医薬品をどの期間処方したか」レポートにまとめ、製薬会社などに有料で提供
- 処方箋に記載された情報は全て電子化し、一定期間の処方箋データを匿名性を確保したうえで集計



3、ビッグデータ活用の事例



販売促進に関する事例

南アフリカにて開発の“Dynamic Discount Solution”(DDS)は
通信設備の混雑状況に基づき、リアルタイムでの割引施策を行い混雑を平準化。

- トラフィックが集中する時間帯においては、料金を高くし、脆弱なインフラにおいても耐えられるよう使用を平準化する。
 - 類似のサービスとして、“Billing and Revenue management”サービスの一環として、“Location Based Charging”の提供も行っている。
- これは、2008年に開設した現地研究所(Ericsson Innovation Center in South Africa)から登場したもの。

▼ Ericssonによる、“Dynamic Discount Solution”(DDS)

時間帯や場所により設定される
通話料金の割引状況を示す。

3、ビッグデータ活用の事例



企業概要

- 電気製品の購入支援情報サイト「Decide.com」
- 2010年設立の米国ベンチャー企業



提供するサービス

- 単なる価格情報のみでなく、ビッグデータ解析による予測をベースに、製品のプライスの変動、ニューモデル発売時期の予測を行い、それをもとに、今「買う(Buy)」べきか「待つ(Wait)」べきかの推奨を行う機能を持つ
- 数百のオンラインショップから毎日10万以上のデジタル関連製品の価格データを収集する他、関連するブログ、ニュース記事を検索、テキストマイニングや機械学習の技術を用いて自動で行われている模様
- 収集したデータはAWSに送られ、AWS上でEMR(Hadoop)を使った集計・分析を行っている(1日で25GB、全体で100TBのデータが分析対象)

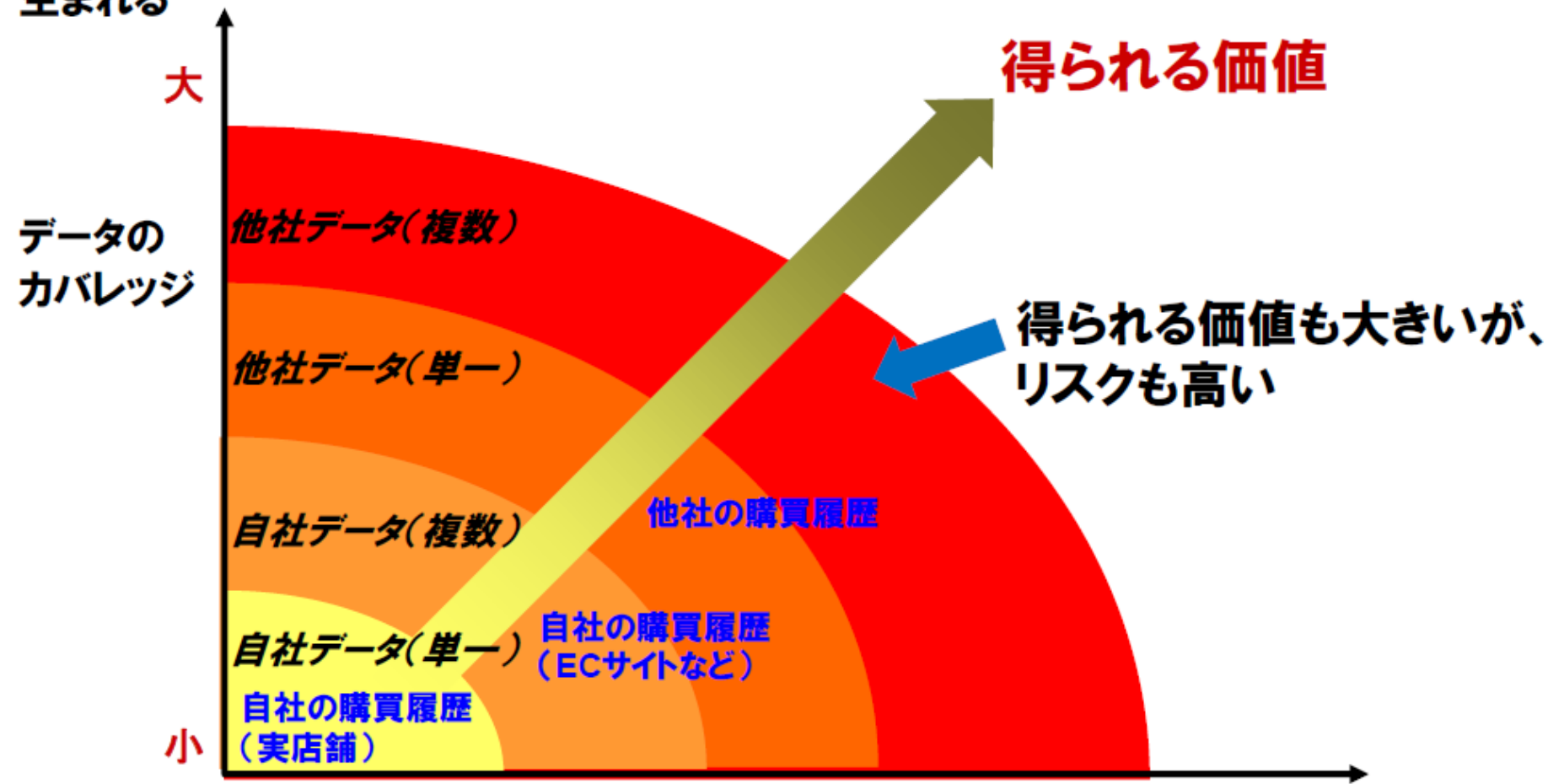


今後注視すべき動向

4、今後注視すべき動向

ビッグデータ第二章は、他社データ・外部データの活用

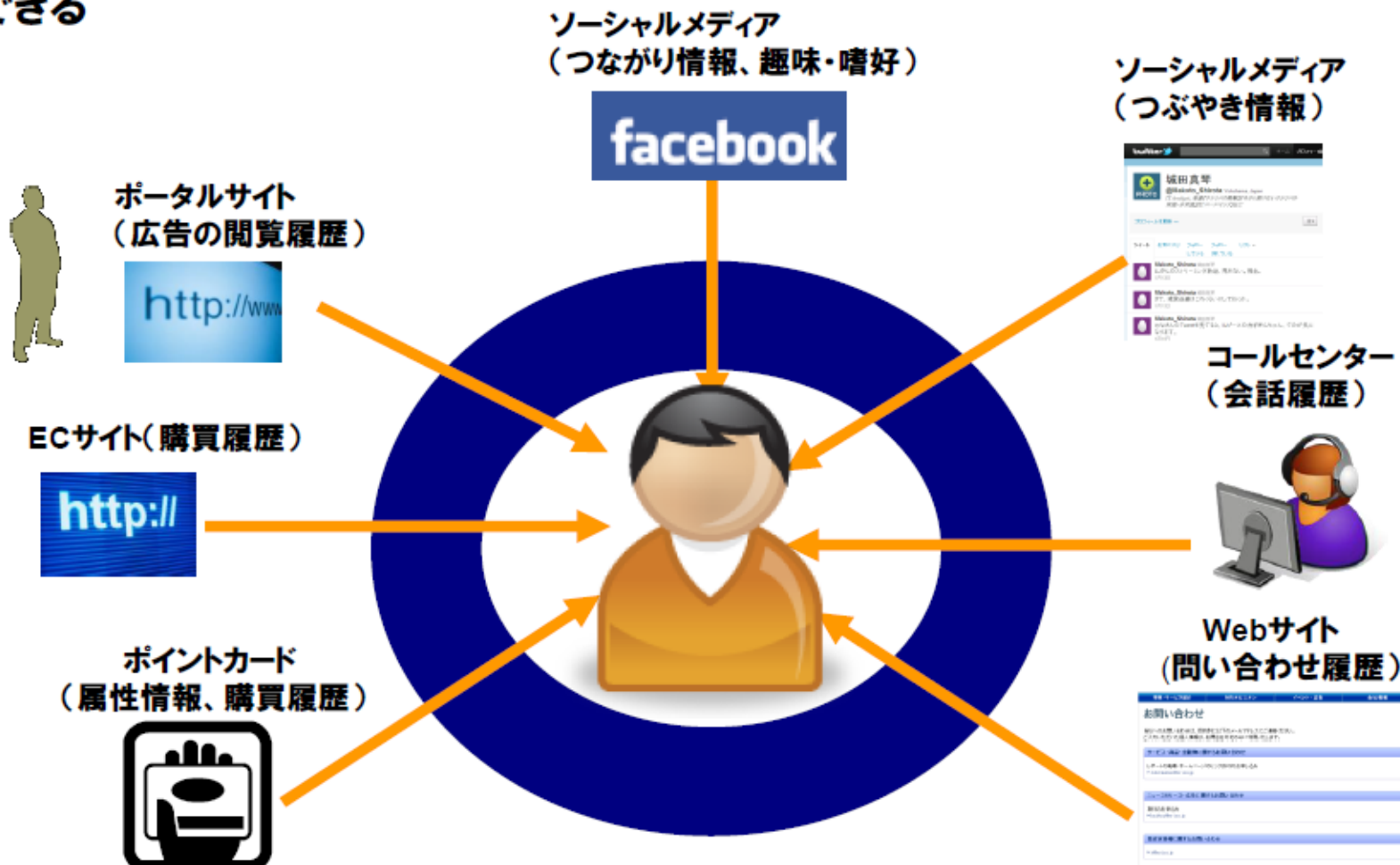
- センサーデータの活用による機械の故障予測など、自社データのみでも十分な価値が得られるケースもあるが、マーケティング分野を中心に外部データの活用で大きな価値が生まれる



4、今後注視すべき動向

目指すは、あらゆるチャネルのデータを活用した真のCustomer 360° View

- 顧客の実店舗における購買データや属性データに加えて、ECサイトやソーシャルメディアでの行動履歴やコールセンターへの問合せ内容を分析することで、顧客をより深く知ることができる

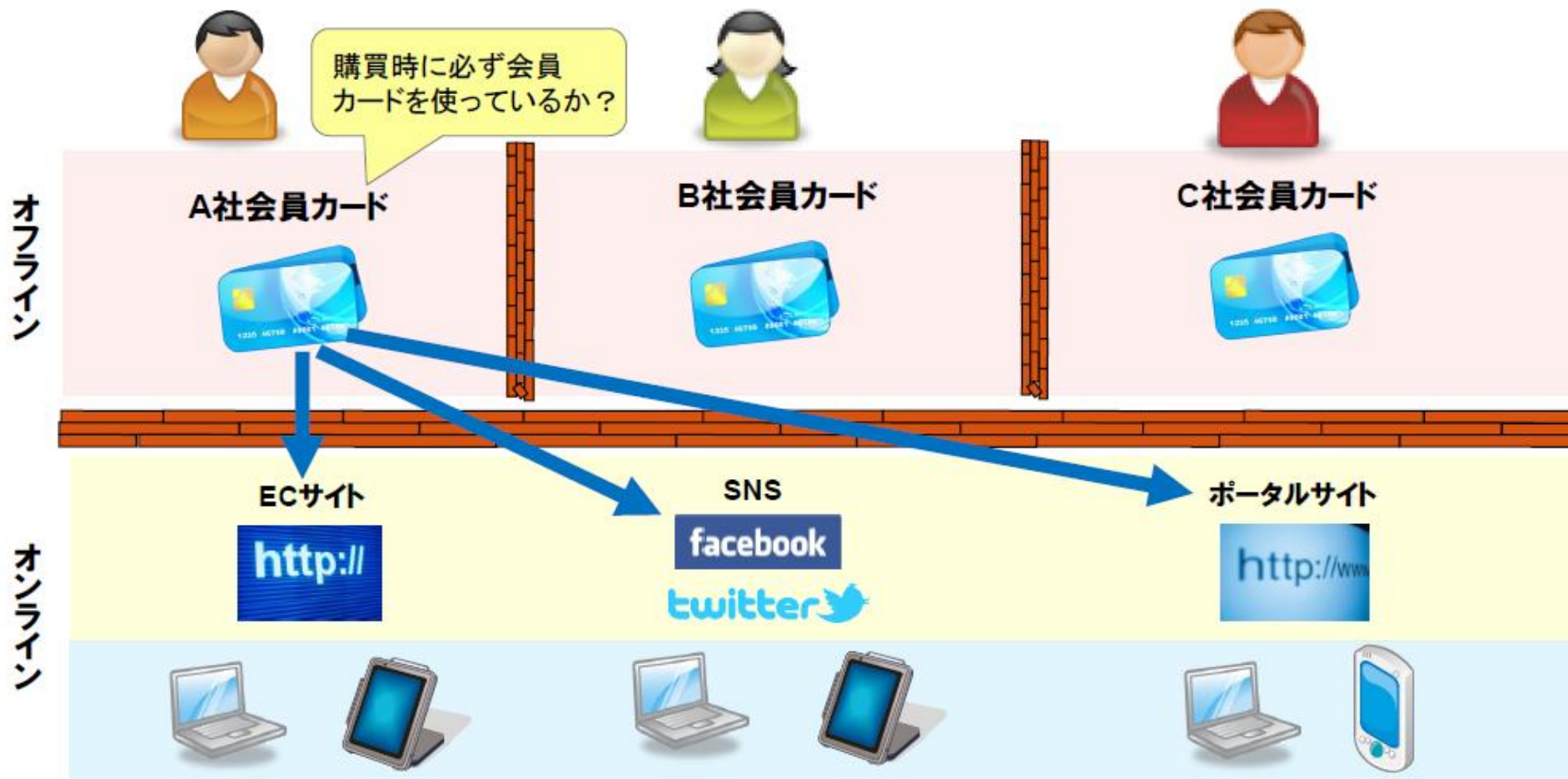


4、今後注視すべき動向



困難を極めるマルチデバイス時代の消費者行動の全体像の把握

- 自社のデータだけでは、消費者行動の全体像を把握することは、もはや不可能
- タブレットやスマートフォンの急激な普及による、「**クロススクリーン・マーケティング**」も課題



4、今後注視すべき動向



ポータルサイトとの連携

O2Oが目的とされるID連携、もう一つの目的は顧客プロファイルの補完

- 互いの会員の顧客プロファイル、購買履歴、購買時の位置情報、ページ閲覧履歴などを利用し、**ネットとリアル**の消費者行動を結びつけ、効果的なマーケティングを推進

両社が発行してきたポイントを「Tポイント」に、インターネット上のIDを「Yahoo! JAPAN ID」に統一



出所)CCC

YAHOO!
JAPAN
地域生活圏情報サイト

出所)ヤフージャパン



リアルとインターネットの「Tポイント」提携店舗網および送客の拡大を狙う



セゾンカード・UCカード会員専用のWebサービス
「ネットアンサー・アットユーネット」



SAISON CARD

出所)セゾンカード



4、今後注視すべき動向

ソーシャルメディアとの連携

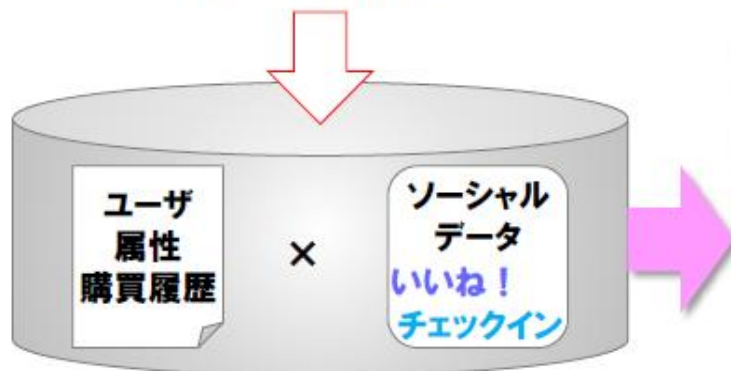
Amex 「Link,Like,Love」プログラム

- Amexのクレジットカード番号と、顧客のソーシャルID(Facebook ID、Foursquare ID)を紐付けるよう促す
- ソーシャルIDを教えてくれた顧客には、Facebookの「いいね」やFoursquareの「チェックイン」の履歴に基づき、「特別なオファー」を提供

ソーシャルメディアのIDと
クレジットカード番号の紐付け



クレジットカードを利用する
だけで自動的に値引きや
キャッシュバックが適用される



ユーザの関心・行動を分析

Recommended For You

Sort By Relevance

Columbia Sportswear
Get 15% Back Online and In-Store

24 Hour Fitness
Spend \$65 or more, Get \$10 back

Columbia Sportswear
Offer Valid 1/3/2012 - 3/31/2012

ADD TO CARD

Get 15% Back Online and In-Store
When you shop at Columbia Sportswear online at www.Columbia.com or in store through 3/31/2012, you can receive a statement credit equal to 15% of your total

4、今後注視すべき動向

原資なしで会員IDの紐づけを促すクックパッドとアイディーズ

- 両社は提携により、各会員の意思により、IDを互いに紐づけ
 - これにより、「東急ストアで買い物したクックパッド会員に限定レシピを提供する」といったサービスを提供
- クックパッドは、売れ筋の食材を把握、スーパーは購入された食材がどのように使われているかを把握可能



レシピの検索履歴
(月間利用者1575万人)

会員IDを互いに紐づけ



スーパー(462店舗)の
会員カードの購買履歴

株式会社アイディーズ



ID-POS分析に特化したIT企業
(全国33のスーパーマーケットチェーン)

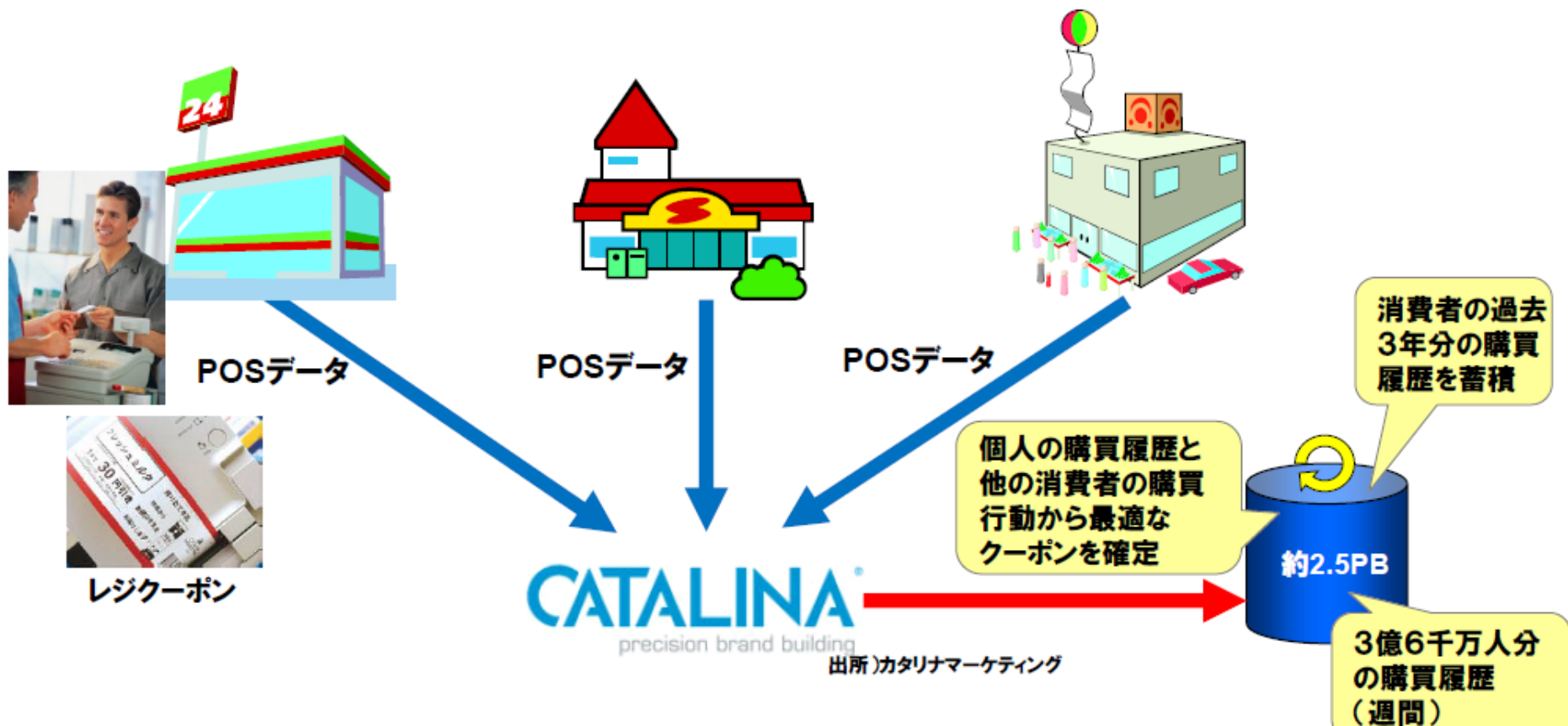
4、今後注視すべき動向



データアグリゲータの例

カタリナ・マーケティング 約2,000万世帯の購買データを蓄積

- スーパーなどの店頭のレジでPOSと連動し、消費者一人一人の購買履歴に応じたクーポンを発行(グローバルで約55,000店舗の小売チェーンに導入)
- 日本では、イトーヨーカドー、イオンリテール、マックスバリュ、相鉄ローゼン、ダイエー、ユニーなど36チェーン、3000店舗に導入(週間約6,500万人の消費者に対してリーチ可能)



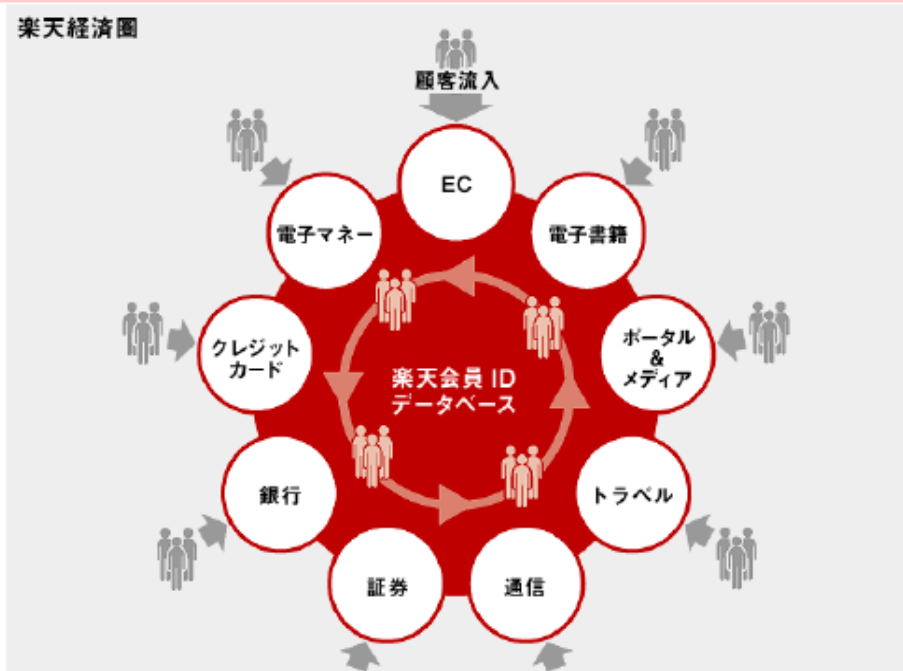
4、今後注視すべき動向

自社グループによるデータのアグリゲーション

楽天 グループ約40のサービスから成る楽天経済圏の構築

- グループ会員DB上には約8000万のIDが存在。各IDの性別、年齢、居住地域、消費志向、楽天会員歴、購買RFM、ポイントRFMなどのデータをDWH上で一元管理
- 横断的なユーザー行動分析を基に、属性に応じたWebページやバナー広告のパーソナライズ化の実現、グループ間でのクロスセルの向上に貢献

『楽天スーパーポイント』を通じ、顧客の流入拡大および経済圏内でのサービス利用や回遊性を促進する「楽天経済圏」



ネットとリアル融合：『楽天カード』『楽天Edy』



ビッグデータ活用の効用を得る
上での課題は何か？

5、ビッグデータの活用に関する課題



人材不足

ボトルネックはビッグデータ取扱いに関するリテラシを持つ人が不足していること。(サービスを受ける側としても、提供する側としても。)

- 「ボトルネックは、ビッグデータの取り扱いができる・理解がある人がいないこと」?
 - 大量のデータを見て、「なにかありそう！」とワクワクできる人はそう多くない。
 - 300万行のデータとか見ると、気持ち悪くなってしまう人が多勢。
 - 外注事業者とコミュニケーションできる程度のリテラシを有した人が少なくとも社内にも必要。
 - システムや、ツール類も重要だが、根幹は数理モデリングに関する人材不足だ。

- 「そもそも発注したことがない。データ分析に金を払う文化が無い。」
 - マーケティングに関しては、電通とか博報堂に頼みやすい風土がある。
「ああ、電博に外注するのね」で稟議が通りやすい。
 - 自社のデータを外出ししましょう、というような文化醸成も必要だろう。

5、ビッグデータの活用に関する課題



人材不足

シリコンバレーにおいても、データ解析人材の争奪戦が始まっている。

「統計学の素養があって、
“Hadoop”が使える、
優秀なやつが入れば、
すぐに年棒10万ドル出すぜ！」

2011年3月 Strata Conference 2011

今後10年間で
最もセクシーな職業は
統計家である。

2009年 Hal Varian, Google Chief Economist

課題としての「プライバシー」

- 一部の悪質なケースを除き、**ユーザーに十分な説明がないまま**、スマホにインストールしているアプリ情報やページの閲覧履歴情報などを収集していたとして批判を受けるケースが大半

事件が明らかになった年月	対象	サービス種別	概要
2011年8月	マニョスクリプト「カレログ」	スマートフォン用アプリケーション	家族や恋人の現在位置をGPS機能を用いて把握する位置情報通知サービス。電話の通話記録やインストールしているアプリケーション、バッテリー残量まで把握できることから、「プライバシー侵害ではないか」「スパイウエアにあたるのではないか」との批判を受ける。
2011年10月	ミログ「App.log」 「app.tv」	スマートフォン用アプリケーション	Android向けアプリケーション分析サービス「AppLog」のソフトウェア開発キット「AppLogSDK」を組み込んだアプリがスマートフォンに導入されると、ユーザーに十分な説明がないまま、端末の固有番号、導入済みのすべてのアプリの名前、使用した時間帯などのデータを1日1回、ミログ社に送信する。スパイウエアではないか、との批判を受け、サービス中止、その後、会社清算に追い込まれた。
2012年1月	電子書籍サービス「ビューン」	電子書籍サービス	ユーザーの同意を得ず、規約にも記載がないままに電子書籍の閲覧履歴や端末識別情報を収集していたとして、批判を浴びる。その後、利用規約等を変更。
2012年3月	はてな「はてなブックマーク」		はてなブックマークボタンがユーザーが閲覧したページの履歴情報などの取得を行い、第三者へ販売していたことが判明。途中から仕様変更した際、事前に十分な告知を行わなかったことも大きな批判を浴びた。結果、はてな社は謝罪。

5、ビッグデータの活用に関する課題



課題としての「プライバシー」

動き出した総務省・経済産業省

- 2012年8月に公表された「スマートフォンを經由した利用者情報の取り扱いに関するWG 最終とりまとめ」では、**同意の取得、契約者・端末固有IDの取り扱い**を以下のように決定

個人情報を含む電話帳、通信内容・履歴、メール内容・送受信履歴等の通信履歴、スマートフォンのアプリケーションの利用履歴(一部の場合を除く)等については、**プライバシー上の懸念が想定されることから、個別の情報を取得することについて同意を取得する。**

契約者・端末固有IDなど、契約や端末に対して一義的に指定・作成され、利用者側で変更が困難であるが、幅広い主体により利用される可能性があるものが取得者において個人識別性を有する情報と結びつきうる形で利用される場合、同一IDの上に様々な情報が時系列的に蓄積し得ること、取得者又は第三者において個人識別性を有する可能性があることから、個人情報保護法への抵触やプライバシー侵害の可能性を考慮し、**個人情報に準じた形で取り扱うことが適切**と考えられる。

具体的には、**取得される項目及び利用目的を明確に記載し、その目的の範囲内で適正に扱うこととする。**

5、ビッグデータの活用に関する課題



誤用・不適切利用

リコメンデーション機能の不適切な推薦がトラブルの原因となったケースもある。

Target(米国の大手総合スーパーマーケットチェーン)

- 従来からユーザーの購買情報を分析し、購買行動に合った品目のクーポンを送付
 - ex. 毎年4月に水着を購入する顧客→7月に日焼け止め、12月にダイエット本
- ある日、「女子高生の娘になぜ妊婦向けのクーポンを送ってくるんだ！妊娠を勧める気か？」とミネアポリスのターゲットの店舗にある男性が怒鳴りこんだ。
- 結局、この娘は本当に妊娠しており、後に父親はターゲットに謝った。



高校生の娘に妊娠を勧めるのか！



出所) <http://www.target.com/>

5、ビッグデータの活用に関する課題



Targetから学ぶべきこと コンプライアンスはあくまで必要条件

- **大切なのは、「消費者の気持ちを押し量ること」。**特にセンシティブな情報の取り扱い。
 - Targetは「プライバシー法などコンプライアンスに保守的な企業」であったが...
- Targetでは、購買履歴に応じて個別に作成している小冊子やクーポンを工夫(ベビー用品とは関係のない商品を混在させる等)
- **法令遵守は必要要件であって、十分条件ではない。**顧客の信頼を得るという基本的要件を満足させなければならない



ビッグデータの活用 に関する主な技術

NoSQL/Not only SQL

- ▶ 表形式によるリレーショナルデータベース管理システム (RDBMS) とは異なる設計によって実装されたデータベースシステム
- ▶ RDBMSが定型データの処理を必要とする業務システムでの利用に適しているのに対し、NoSQLはセンサーやソーシャルメディア等の非定型データを含む多様なデータを大量にデータベース化するために利用

Hadoop

- ▶ 米国NPOのApacheソフトウェア財団のプロジェクトで開発が進められている、大規模データの効率的な分散処理等のためのオープンソースソフトウェアフレームワーク
- ▶ 複数のサーバを通じた並列処理により、柔軟かつ継続的な大規模データの高速処理が可能



クラウドサービス

- ▶ 利用者が必要なコンピュータ資源を「必要な時に、必要な量だけ」利用でき、拡張性、可用性、俊敏性や経済性等の特徴を有するサービス
- ▶ クラウドサービスの利用により、多種多量のデータの蓄積や計算処理のために必要となる多数のマシンについて、自前で用意する必要がなく、低コストで同様の環境の構築が可能
- ▶ また、例えば、「Hadoop」稼働環境を提供するサービスを利用すれば、マシンの調達に加え、ソフトウェアのインストール等の設定作業の省略も可能

DWH (Data Ware House)

- ▶ 定型データ・非定型データを問わず、大量データの蓄積を目的とするデータベースの総称
- ▶ これらの大量データを高速に処理する方法により、
 - ①従来のRDBMSとは異なる設計技術によるNoSQLデータベースと、
 - ②標準的なRDBMSとハードウェアレベルでの高速化技術を組み合わせたDWHアプライアンスの2つに大別

CEP (Complex Event Processing)

- ▶ データをディスクに格納せずに、書込速度がディスクに比べて高速なメモリ上で逐次的に処理することにより、必要な情報をリアルタイムに抽出する技術
- ▶ ディスクにデータを蓄積して分析する手法と比べ、短時間で処理が可能であるため、クレジットカードの不正利用や防犯カメラ映像の異常検知等、短期間での対応が必要となる場合に利用
- ▶ 予め利用者が定義するリアルタイム処理の内容には、単一のデータ属性の閾値による判別に加え、複数の属性を組み合わせた処理の設定も可能

PPDM(Privacy Preserving Data Mining)

- ▶ プライバシーを保護した上で、大規模なデータから特徴や規則性等を抽出する技術
- ▶ 匿名化や秘密計算等により、個々のデータを暗号化したままデータマイニングを実施することで、個人情報流出等のリスクを回避した上で、データの解析等が可能
- ▶ 例えば、どのデータについても同じものがk件以上存在するようにデータの粒度や曖昧さを制御するk-匿名化技術等が研究

MDM(Master Data Management)

- ▶ 業務を遂行する際の基本情報である顧客情報等のマスターデータを管理するためのシステム
- ▶ 多種多様な情報システムに重複・散在し、かつ、多量に生成等されるマスターデータについて、常時の最新状態への更新やシステム全体の整合性の確保等が可能

秘密計算

- ▶ 入力データや演算ロジックを暗号化したままで任意の計算を可能にする技術
- ▶ 複数のコンピュータにデータの断片を送り、断片の部分計算を繰り返し行うことで、データを秘匿したまま統計などの各種計算を行うことが可能
- ▶ PPDMでの利用の他に、民間企業、公共機関、教育現場などにおけるプログラムの不正解析防止、知的財産の侵害防止、情報漏えい防止などの多様な分野への応用が期待

- ◆ **ビッグデータは、Volume、Velocity、Varietyという「3つのV」の特性を持つデータと定義できる**
- ◆ **一般の事業者においては「製品開発」「販売促進」「保守」など、多様な業務プロセスでの効用が期待できる**
- ◆ **ビッグデータ第二章は、他社・外部データの活用。自社で保有していないデータを他社との連携を通じて補完する**
- ◆ **懸念される個人情報保護、プライバシー侵害等について、まずは収集している情報、目的などわかりやすく説明する**
- ◆ **従前より「データ分析の重要性」を説きながらも看板倒れに終わったブームはいくつもあったが、「ビッグデータ」はこれまでと比べて蓋然性の高い潮流**

ご静聴ありがとうございました