

株式会社数理システム ユーザーコンファレンス2012 基調講演

ビッグデータ時代の
企業内におけるデータ解析とは
～一現場分析者の考察～

株式会社リクルート テクノロジーズ
ビッグデータグループ

西郷彰



◆本日のアジェンダ

リクルートグループのご紹介

ビッグデータ活用と分析取組事例

データ分析者の方に向けて

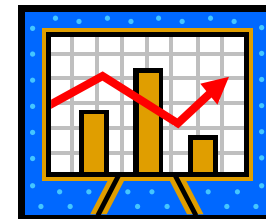
西郷 彰

株式会社リクルートテクノロジーズ
ITソリューション部 ビッグデータグループ
シニアアナリスト



演者紹介

氏名	西郷 彰
所属	株式会社リクルートテクノロジーズ ITソリューション部 ビッグデータグループ
略歴	大手飲料メーカー、マーケティングリサーチ会社を経て、分析サービス会社の設立 その後、リクルートに入社 全社インターネット部署にて分析全般やマーケティング施策の開発、分析組織提案等に従事
学歴	大阪大学工学部 筑波大学大学院ビジネス科学研究科（MBA）
趣味etc	データ解析コンペティション -レコメンデーションアルゴリズム -ベイジアンアプローチ スノーボード（インストラクター資格有） 最近の週末は、一日中子供と遊んでいます



ネットマーケのデータ分析

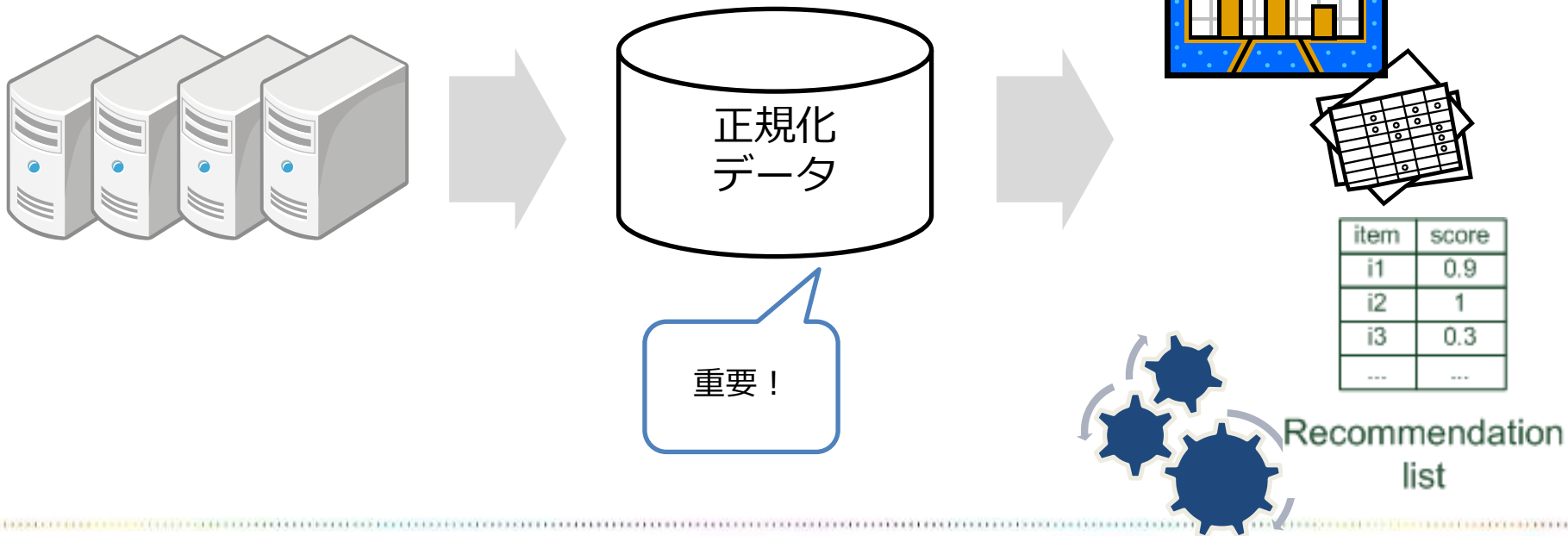
行動ログが手に入る

URLから正規化して使うことができる

さらにデータ加工によりいかようにもなる

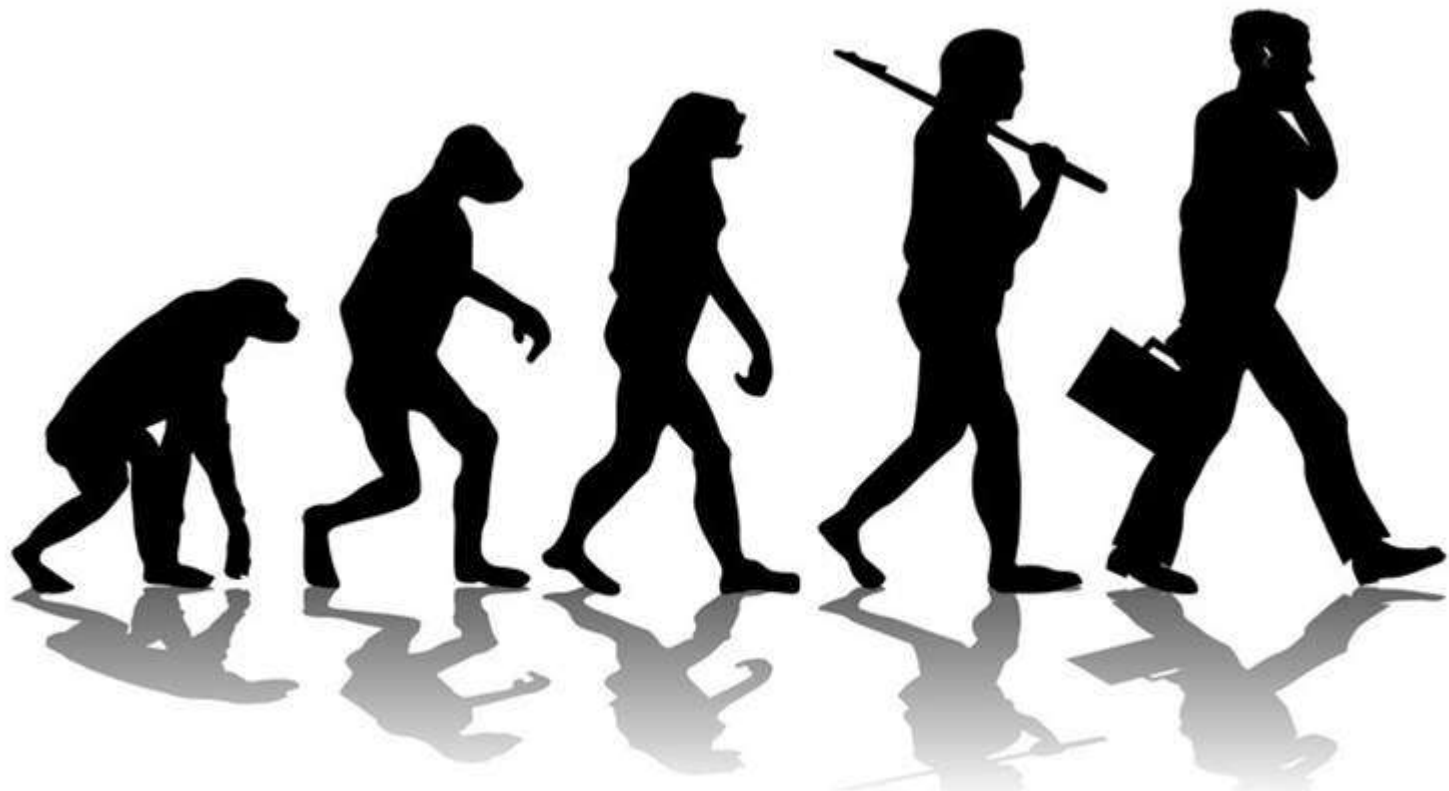
非構造のデータを使い出す企業も増えてきている

マーケティング施策に直接接続できる



ネットマーケティングのデータ分析

まさにビッグデータ
時代になった



**自分と同じような思いを
持った人に伝えたい**

リクルートグループの紹介

リクルートグループのご紹介

～リクルートについて

ライフイベント領域

車購入 

住宅購入 

転職 

出産/育児 

結婚 

就職 

進学

ライフスタイル領域

旅行



お稽古



時事

R25



L25

ファッション

eruca.

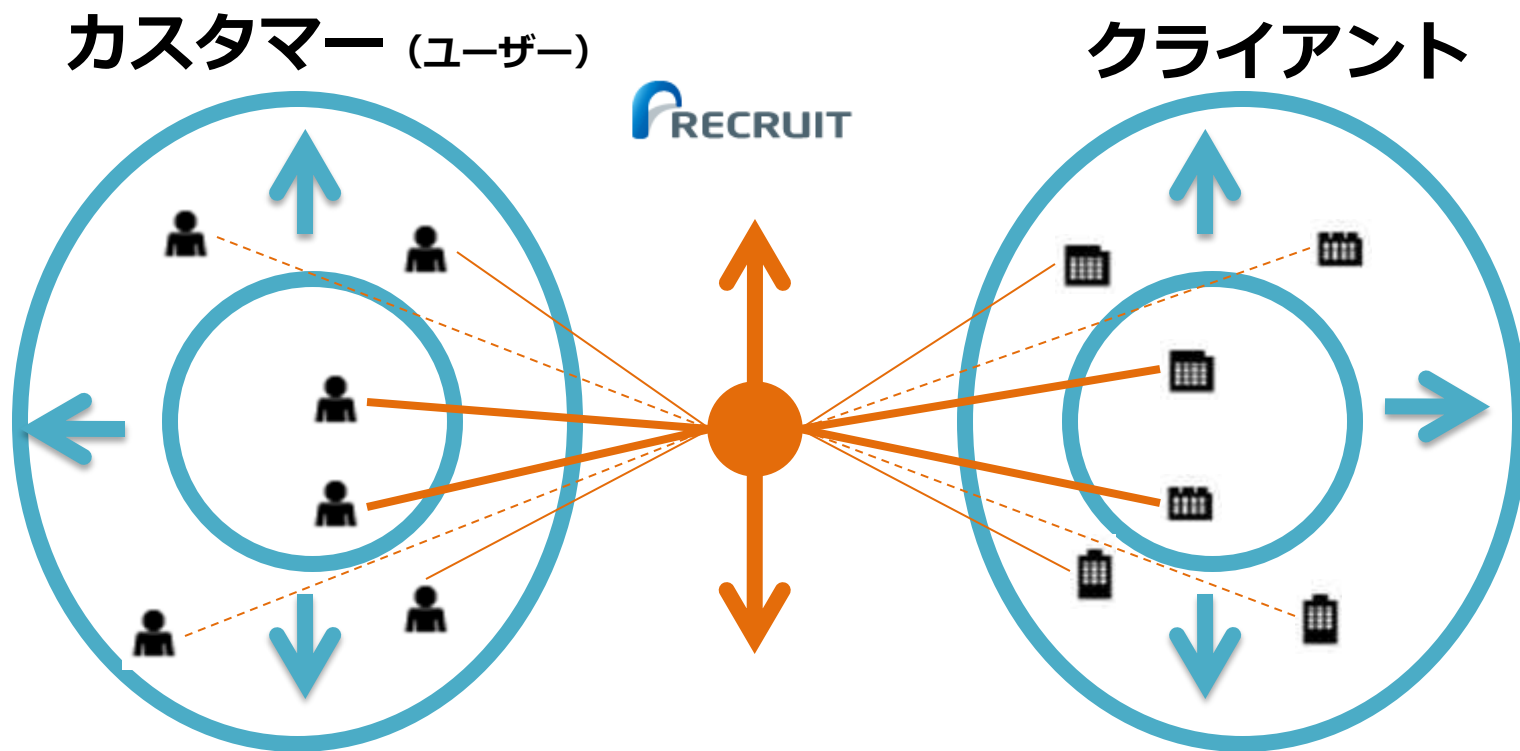
飲食

HOT PEPPER

「選択・意思決定・行動」を支援する
情報サービスの提供

リクルートグループのご紹介

カスタマーとクライアントを新しい接点で結び、「まだ、ここにはない、出会い」の場を創造する。



リクルートグループのご紹介

～リクルートグループについて

創 業 : 1960年3月31日

資本金 : 30億264万円

売上高 : 3,720億57百万円 (2011年4月1日～2012年3月31日)

連結売上高 : 8,066億61百万円 (2011年4月1日～2012年3月31日)

グループ従業員数 : 22,368名 (2012年3月末時点)

代表者 : 代表取締役社長 峰岸 真澄

2012年10月、会社分割を行い、

7事業会社 + 3機能会社のグループ体制となりました

リクルートグループのご紹介

事業会社

(株) リクルート
ホールディングス

(株) リクルートキャリア

(株) リクルートジョブズ

(株) リクルートスタッフィング

(株) リクルート住まいカンパニー

(株) リクルートライフスタイル

(株) リクルートマーケティングパートナーズ

(株) スタッフサービス・ホールディングス

機能会社

(株) リクルートアドミニストレーション

(株) リクルートコミュニケーションズ

(株) リクルートテクノロジーズ

リクルートグループのご紹介



大きいだけでなく多種で複雑なデータ

Webサイト

pageview

セッション

アイテム

営業情報

エリア UU

クライアント

会員

店舗情報

ビッグデータ活用と分析取組事例

ビッグデータ活用と分析取組事例

ビッグデータに対峙する2種類のアナリスト



コンサル型 データアナリスト

【役割】

事業の抱える課題解決に向けた仮説を立て、大量データをマイニングし具体的な解決案を提案する。現場に近いところで超具体的な方針を示す

【スキル・知識】

分析力、論理的思考力、仮説力、統計/マーケティング知識、SQL、Rなど



エンジニア型 データアナリスト

【役割】

データマイニングや機械学習を行った結果をもとに、ユーザーの行動特性など一定の規則性を見出し、提供サービスの品質向上に努める

【スキル・知識】

Hadoop、MapReduce、Mahout、Java、R、統計解析、時系列解析、データマイニング、機械学習、自然言語処理など



出展: <http://engineer.typemag.jp/knowhow/2012/02/-13-2.php>

ビッグデータ活用と分析取組事例

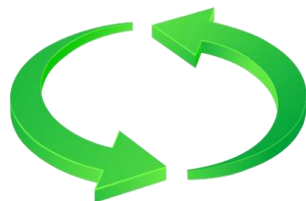
「エンジニア型」アナリスト

ビッグデータ関連技術の活用方法を、技術力・インフラ基盤と共に提供し、新たな施策を事業とともに考え実装していく

▶ アルゴリズムを「実装する」・「組み合わせる」



事業担当者



エンジニア型アナリスト

事業の状況を背景とした、新たな施策の検討、期待する成果・目的の設定

技術で実現できることを背景としたソリューションの紹介、技術力・インフラ基盤の提供、活用方法の**事例展開**や新たな**用途開発**など

ビッグデータ活用と分析取組事例

「コンサル型」アナリスト

ツールを駆使してデータを分析する。どんなマーケティング施策を打てばビジネスにプラスになるか事業と共に考え、意思決定の手助けをする

▶ 最適なアルゴリズム・ロジックを「見つける」



例えば

...

需要予測

何にどのタイミングで広告宣伝費をかけるのが一番効率的か？
(某メディアでは年間の広告宣伝費を3億削減)

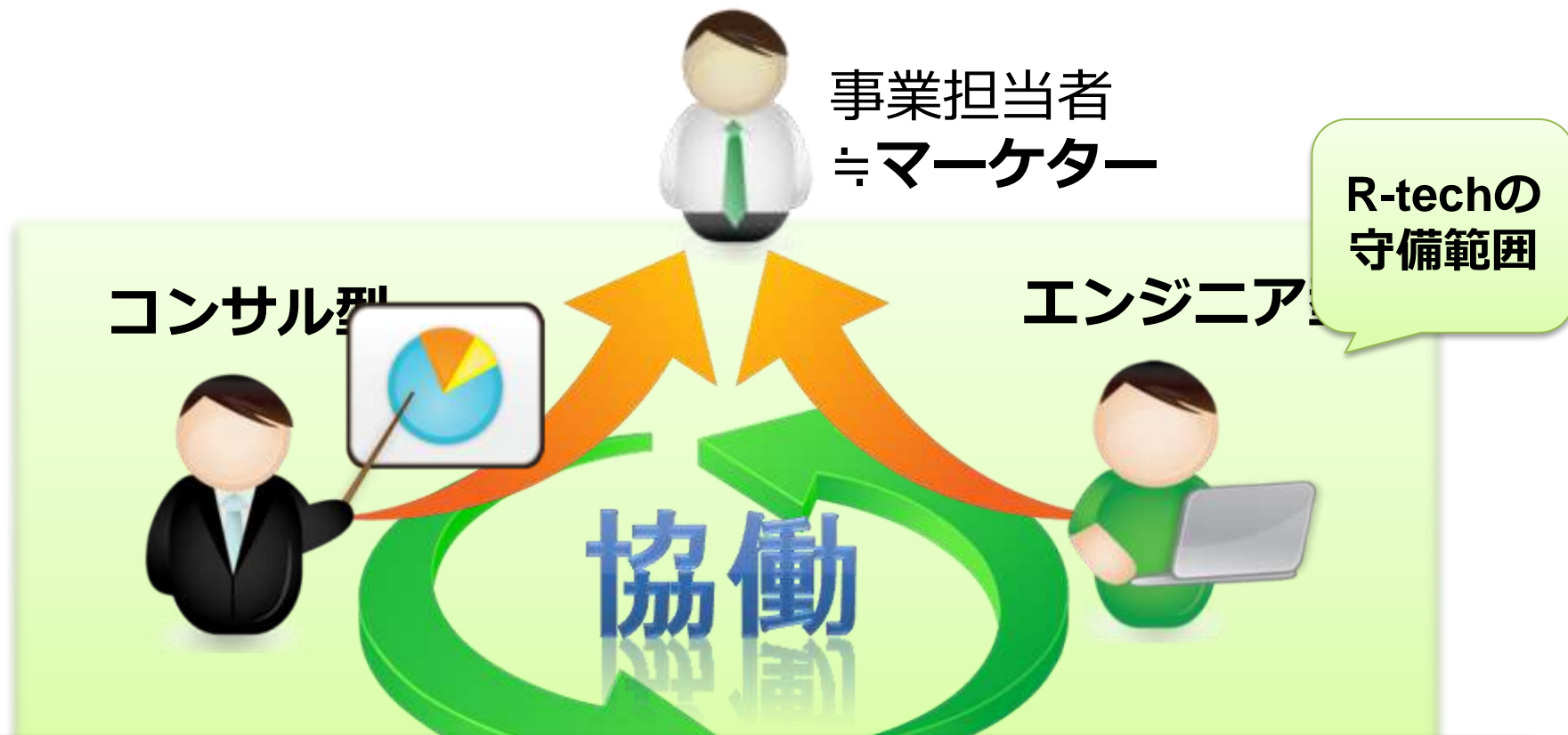
レコメンド

どのようなロジックにするとアクション率が向上するか？



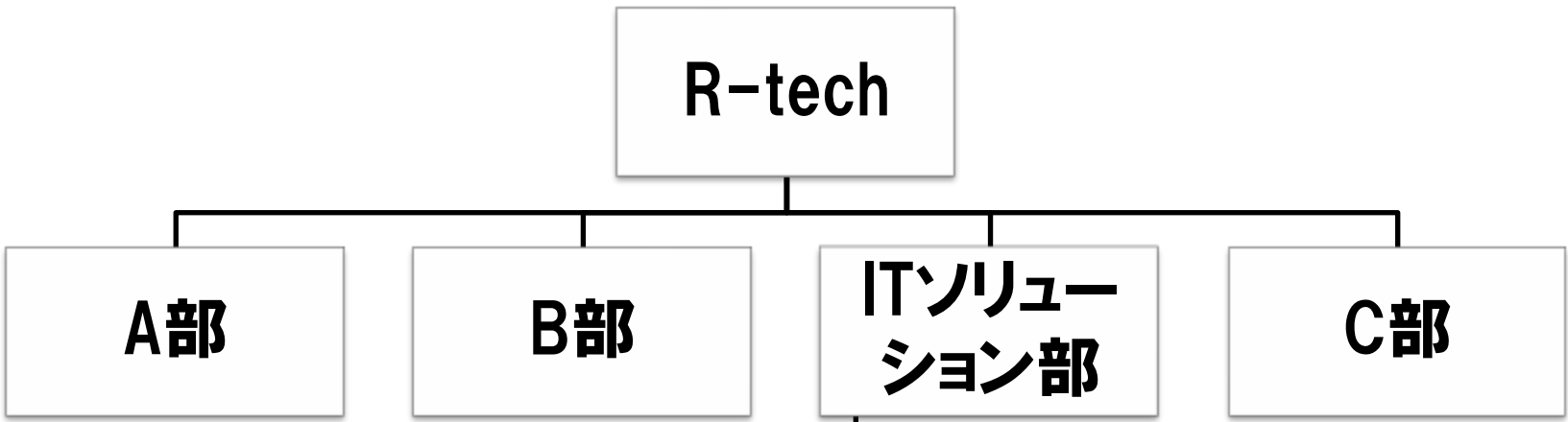
ビッグデータ活用と分析取組事例

(「コンサル型」 + 「エンジニア型」) × マーケター



「コンサル型」 + 「エンジニア型」の異なるアナリストと
マーケターの**三位一体**で、最適なデータ活用を実現

ビッグデータ活用と分析取組事例

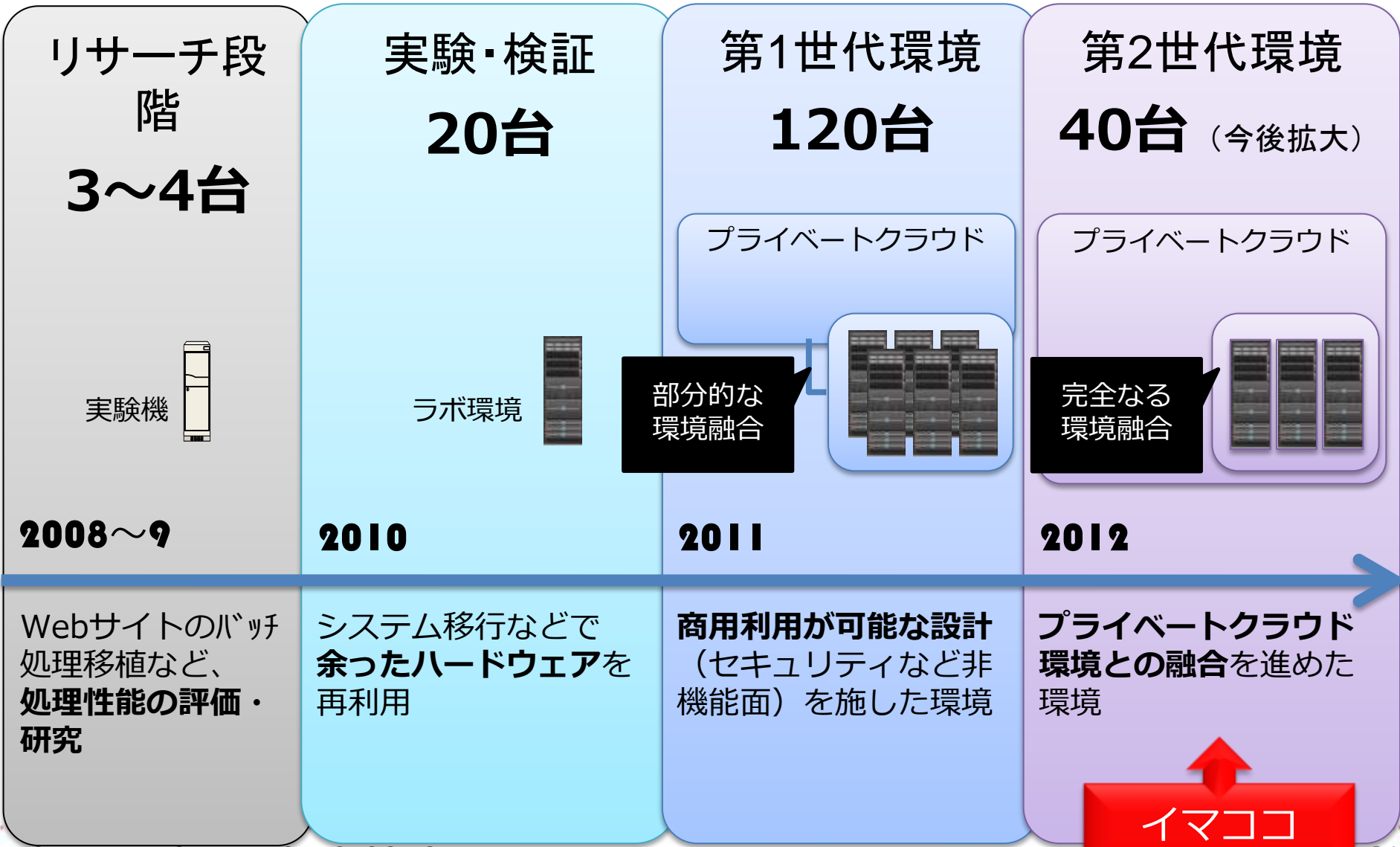


リクルートでも
ビッグデータ活用がバブルに
見えますが、

実は、、、

ビッグデータ活用と分析取組事例

システム構成概要



多くの失敗と
少しの成功の
積み重ねだった

ビッグデータ活用と分析取組事例

- 集計・モニタリングのお手伝い
- CRM分析
- 予測モデル
- レコメンデーションのアルゴリズム開発
- ・ ・ ・ etc

- ある調査会社の説明では8割の会社が情報活用を失敗する
- リクルートも同じく失敗
- でも、自分をはじめとする現場の人間は向き合っていた
- **ボトムアップ型**のビッグデータ活用

● 投資が通らないこともあった。



2009年
↓
2012年

- ビッグデータの相場は未成熟
- ボトムアップでやり続けることによって相場観があう&言葉があう。
- ROIがわかるようになり、投資が可能になってくる。（いくらくらいかければどのくらいできるかがわかってくる。）

そうこうしている間に・・・

成果が出てきた。

ビッグデータ活用と分析取組事例

具体的な事例…

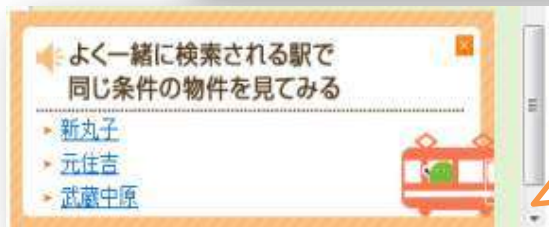
あっ、ここに住もう!



スーモ



アソシエーションルールによる
レコメンドエリアの算出



下まで閲覧すると
レコメンドバナー表出



1週間分のログをデータマイニングツールでレコメンド計算



約2万人に
レコメンド



1年半分



でレコメンド計算

約20万人に
レコメンド



処理が10倍
CVRは
1.6倍に

アルゴリズム & 実装 & FS & 運用

Industry Report



Recommendation algorithms can help Amazon list their top-selling products. Each time a customer browses the site, they are presented with a list of recommendations. Many applications can only do this if the customer purchase and explicitly rate to improve their interests, but they can also use other techniques, including item-based, user-based, and

G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing, 7(1):76-80, January 2003.

Item to Item
アイテムベース
フィルタリング

User to Item
協調フィルタリング

アルゴリズム

モノとモノの類似度を計算して推薦
・アソシエーション
・ピアソン相関係数

ヒトとヒトの類似度を計算して似ている人の購入アイテムを推薦
・コサイン類似度
・ピアソン相関係数

適用

・これを買った人は、これも買っています (チェックも同様)

・パーソナライズ
・あなたにおすすめの商品



要件定義
アルゴリズム
の作成



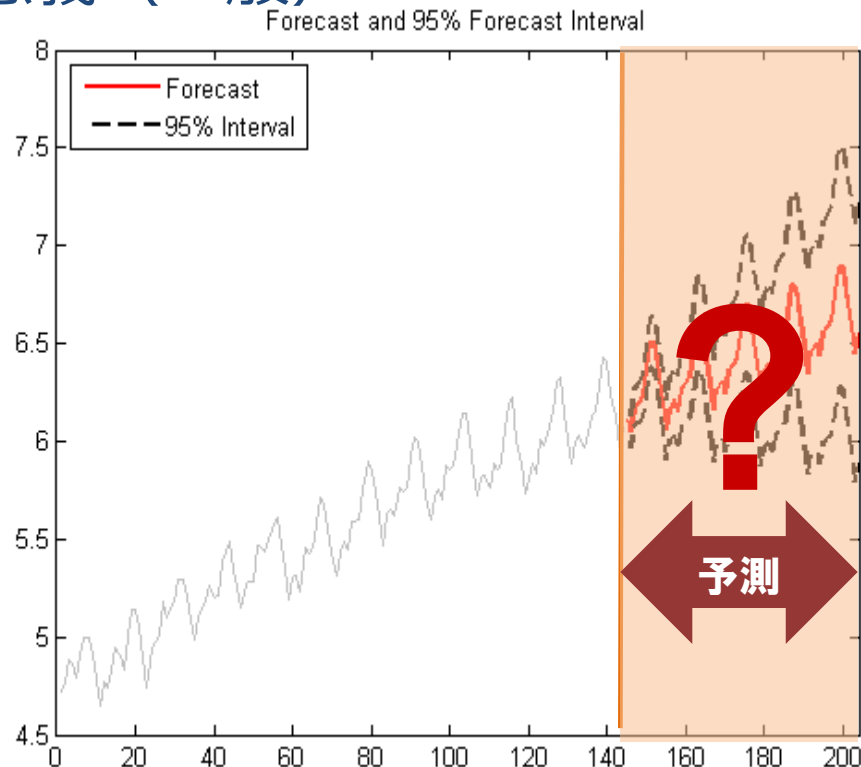
実装



サイトでの
FS→運用

ビッグデータ活用と分析取組事例

需要予測をやる意義（一般）



将来の需要量をできるだけ精度高く見立てることで、
適切な投資（ヒト・モノ・カネ）を行い、最大の利益
を上げたい

● 「統計モデル」による集客構造の数式化

ごく単純な例)

$$\begin{aligned} \text{CV数} = & \overset{\text{寄与度}}{\text{0.25}} \times \text{Listing Cost} + \text{0.63} \times \text{Banner Cost} + \\ & \text{0.50} \times \text{Display Cost} + \text{Error} \\ & \text{リスティングによるCV} \quad \text{バナーによるCV} \\ & \text{ディスプレイ広告によるCV} \quad \text{誤差項} \end{aligned}$$

- 寄与度：各コストのCV数に対する影響度を表します
 - データから計算（統計的に推定）します
- 統計モデルなので、「誤差項（予測できない部分）」を含みます

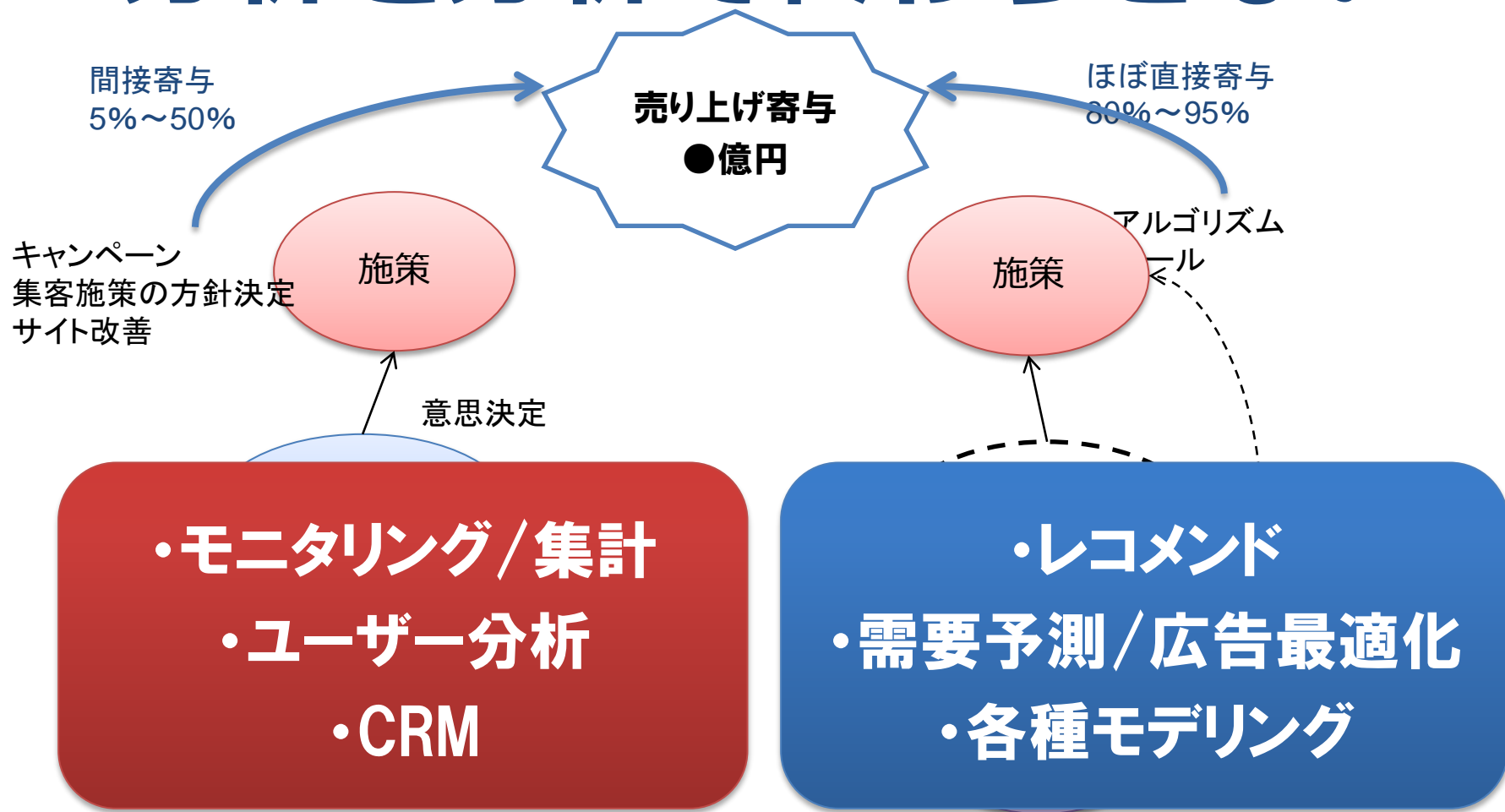
- 事業A
- 事業B
- 事業C
- 事業D
- 事業E
- 事業F
- 事業G
- 事業H
- 事業I
- 事業J
- 事業K
- 事業L
- 事業M

1 3 事業に対し、

年間1000件超の データ利活用

を展開中

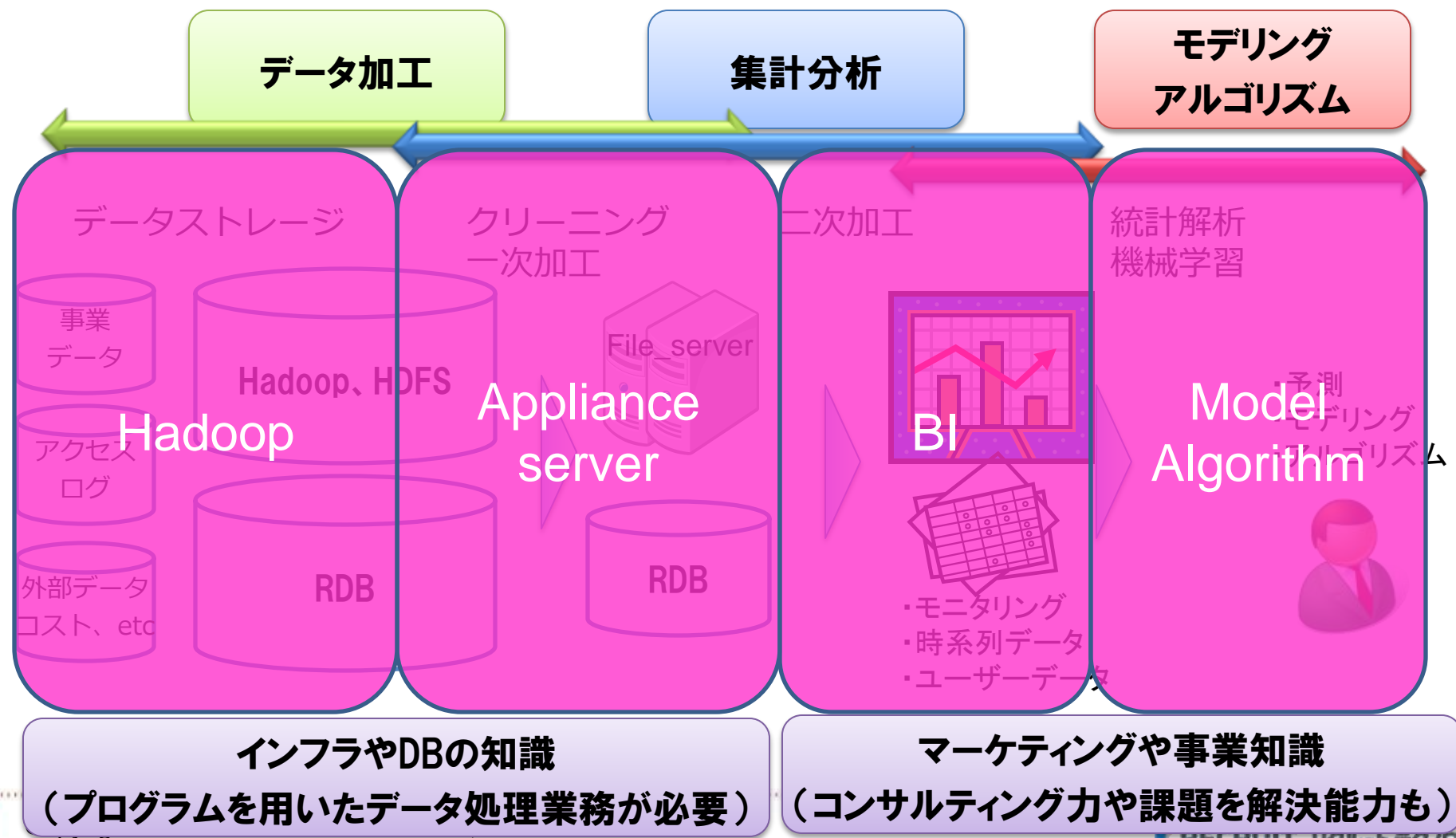
大切にしていること “分析を分析で終わらせない”



データ分析者の方に向けて

データ分析者の方に向けて

専門性が高く業務範囲が広い



データ分析者の方に向けて

Facebookのデータサイエンティスト募集要項

- 定量的アプローチを用いて分析的な課題を解決した豊富な経験
- 多様なソースからの複雑で、大量かつ高次元のデータを容易に操作し、分析できること
- Python, PHPなど、少なくとも1つのスクリプト言語を自由に扱えること
- RDBとSQLに精通していること
- R, MATLAB, SASのような分析ツールについて専門知識を有すること
- 大量のデータセットを扱った経験、MapReduce, Hadoop, Hiveなどの分散コンピューティングツールを使用した経験

データサイエンティストに必要な資質

- コミュニケーション能力: データ分析の知識に乏しいビジネス部門側のスタッフや経営層にも、結果を効果的に伝えられる。
- アントレプレナーシップ: データ中心の新しいサービスを生み出そうとする起業家精神も。
- 好奇心: 芸術、技術、医療、自然科学などの特定の分野にとらわれずあらゆる分野に好奇心が旺盛。

城田真琴 “ビッグデータの衝撃”2012, 東洋経済新報社より抜粋

**5年前はBIG DATAについて、
誰も議論してなかった。**

5年後は当たり前になってるかもしれない。

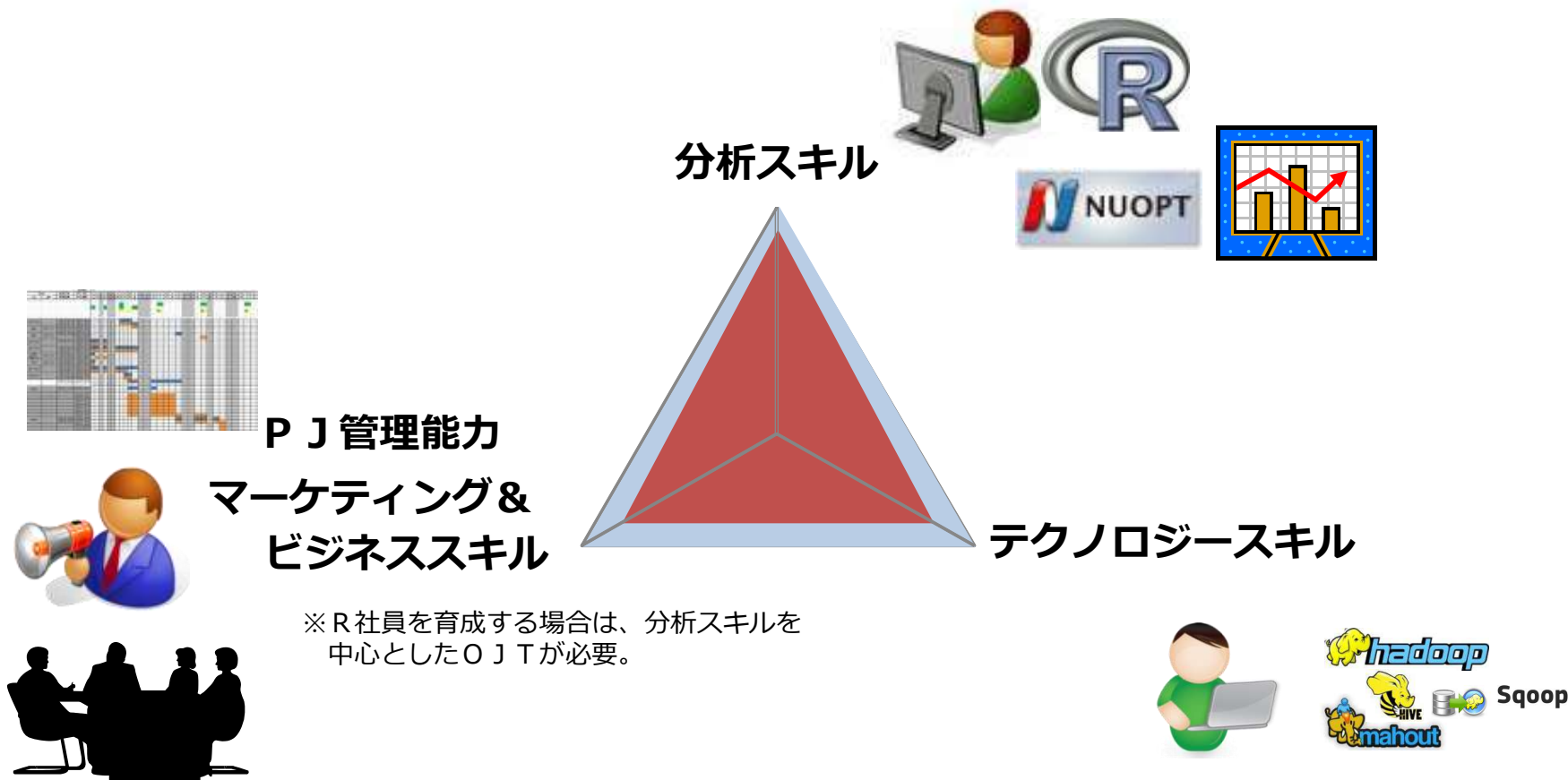
データ分析者の方に向けて

ビッグデータ時代の分析者に必要とされるケイパ

- ビジネス課題から情報活用を**企画する力**
- **ビジネスの特徴をモデルや数式、数理最適化に落とし込む分析設計力**
- **統計解析/データマイニングなど**
- **プログラミング/IT技術**
- 課題解決の具体的な出口に接続する**施策接続力**

データ分析者の方に向けて

● 何か一つの能力の高さではなく面積なのでは？



エンジニア&アナリスト 分析スキルだけでなくテクノロジースキルも必要

また、施策接続や結果の活用（意志決定支援）のため、

マーケティングや事業のビジネスに対する理解・コンサルティング、PJ管理能力も必要

ご清聴ありがとうございました。

**ご質問等はこちらまで
saigo@r.recruit.co.jp**