

ビッグデータビジネスと データサイエンティスト人材の状況

2014年2月7日

サムライクラウドサポーター
国際大学GLOCOM 客員研究員
林 雅之

目次

- ・ビッグデータビジネスの概観
- ・データサイエンティスト人材の状況
- ・まとめ

ビッグデータ、オープンデータに関わる諸活動

- ASPIC オープンデータ研究会 有識者委員
- IT協会 データサイエンティスト育成委員会
- JST(学技術振興機構)
データサイエンス・アドベンチャー杯 審査委員
- 書籍「オープンデータ（インプレスR&D）」 2014.2.22発刊予定
- ビッグデータ、オープンデータに関する講演（年間20回くらい）

ビッグデータビジネスの概観

第3のプラットフォーム

- ・クラウド
- ・ビッグデータ
- ・モバイル
- ・ソーシャルメディア

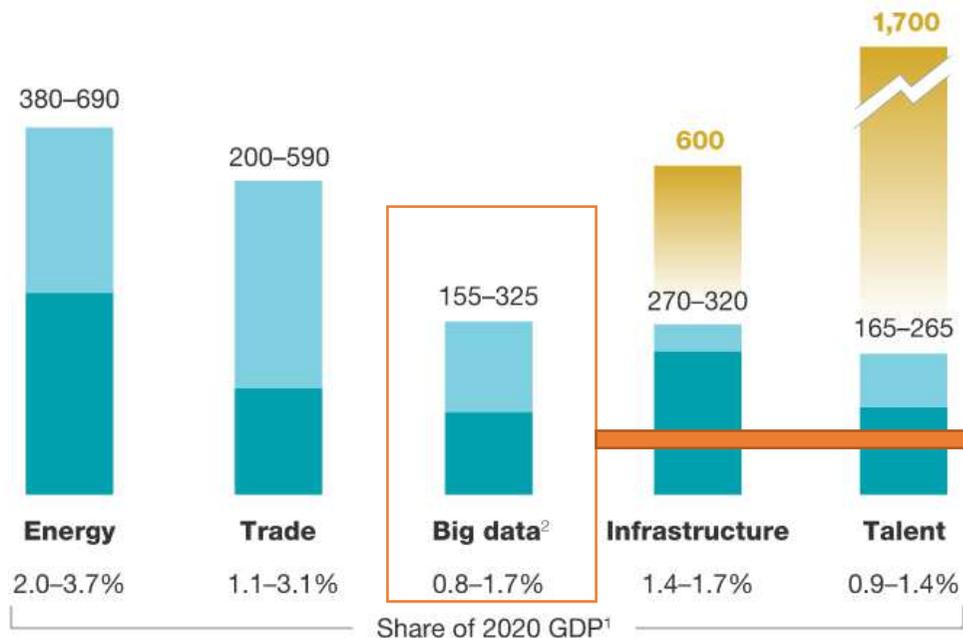
ビッグデータによるGame Change

アメリカは構造的な問題の打開には成長志向の政策と投資の活発化、雇用の促進が必要とし、シェールガスと石油産業、貿易促進、インフラ投資、ビッグデータの分析と教育改革の5つの分野における大規模な投資が必要と指摘

Incremental annual GDP,¹ \$ billion

All five game changers promise solid GDP gains by 2020 ...

... while two offer even greater impact by 2030.

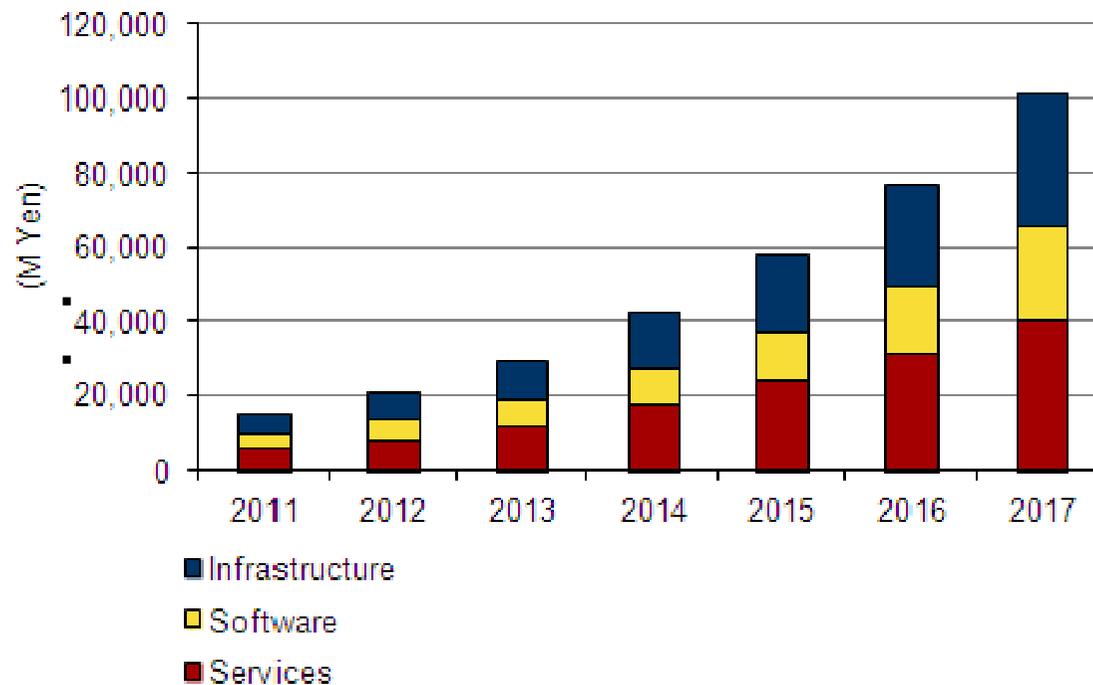


小売りやサプライチェーン・マネジメント、製造業の合理化の推進、政府の医療コスト負担を抑制など、経済効果は、1550億ドルから3250億ドルと推計

出所 : Game Changers: Five Opportunities for U.S. Growth and Renewal 2013.7

ビッグデータビジネスの成長性

・ 2012年の国内ビッグデータテクノロジー／サービス市場は、206億7,000万円。2012年～2017年は年間平均成長率37.5%で拡大、2017年には1,015億6,000万円に達すると予測



【普及の鍵】

- ・ 普及を阻害する要因として、投資対効果が見えにくいとの回答
- ・ ビッグデータの活用を促進するには、構造化、非構造化データに関わらず、アナリティクスソリューションの普及が不可欠

【適用分野】

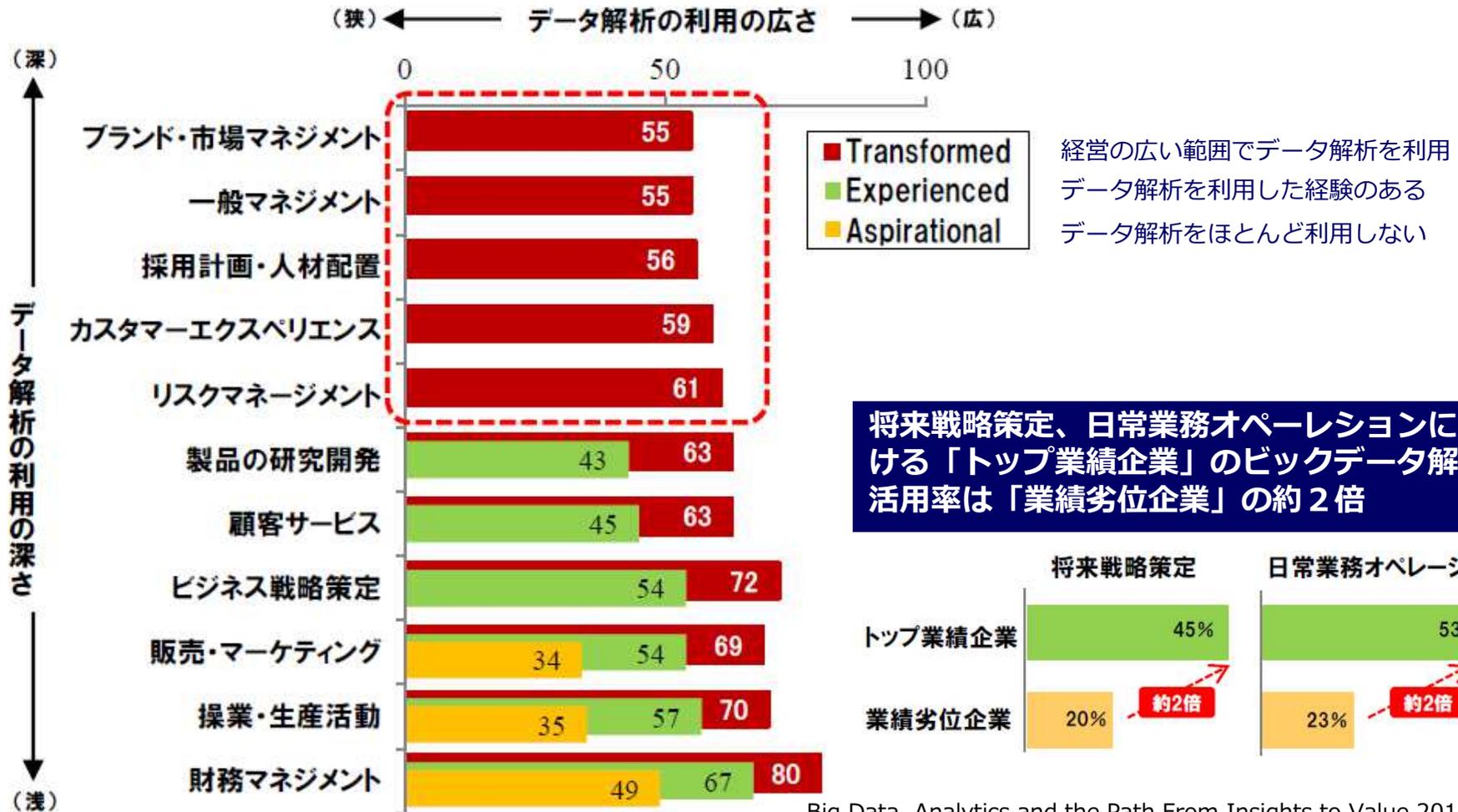
- ・ 2015年度までは企業内でのデータ活用
 - ・ 2017年度を過ぎるとスマートシティなど社会インフラを支えるシステムへ
- <矢野総合研究所 2012.4>

国内ビッグデータテクノロジー／サービス市場 エンドユーザー売上額予測、2011年～2017年

出所：IDC Japan 国内ビッグデータテクノロジー／サービス市場予測を発表2013.8.26

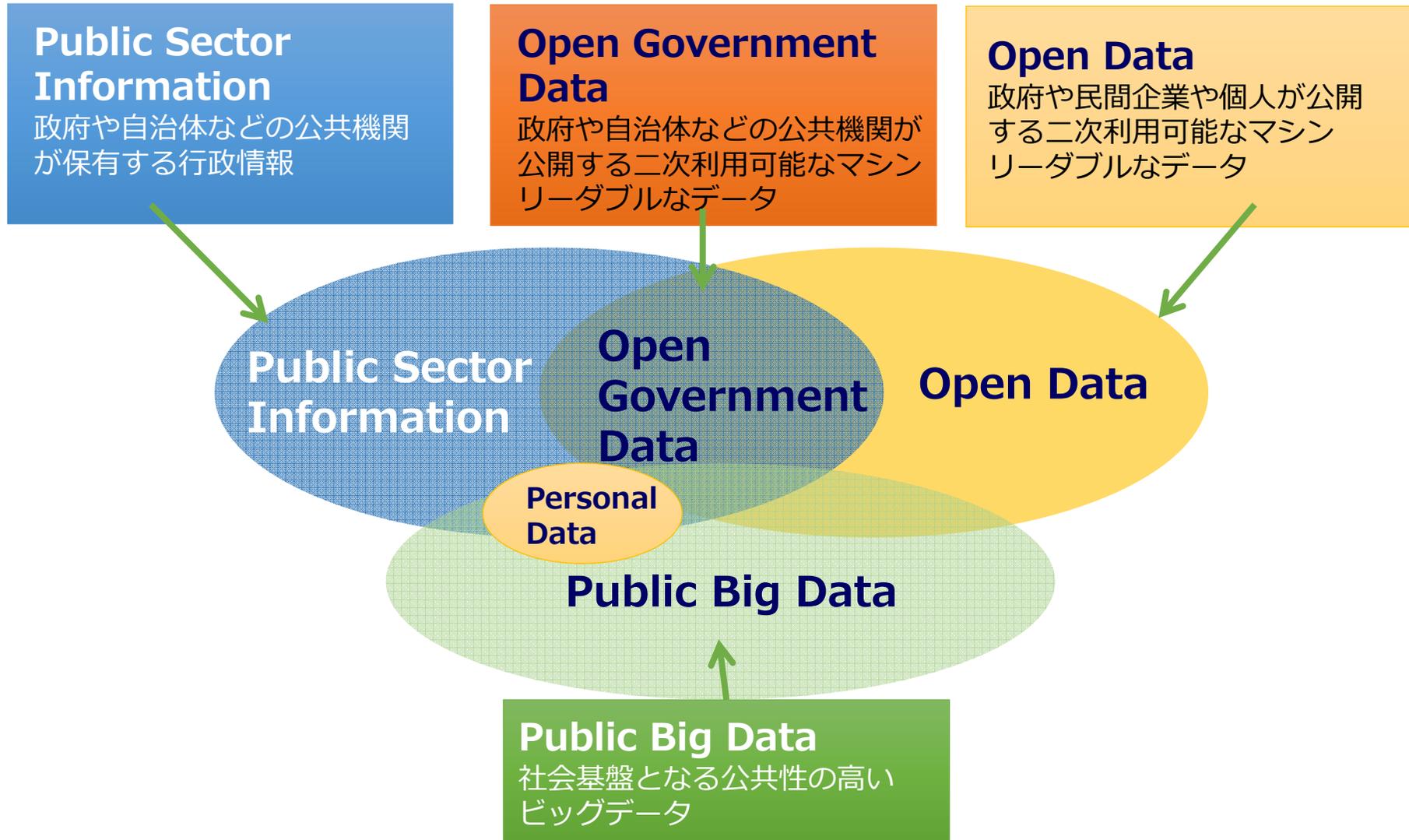
ビッグデータ解析の活用領域

データをよく利用するTransformed企業は下記すべての項目において分析を利用



Big Data, Analytics and the Path From Insights to Value 2011.11

公共系ビッグデータとオープンデータ



事例：ビッグデータとクラウド（NTTコム：日本ラッド）

Hadoop・Erlang・Riakなどのビッグデータ分野で活用されている最新技術を採用し、エヌ・ティ・ティ・コミュニケーションズ株式会社が提供する「Cloudⁿ」上にビッグデータ解析基盤を構築するサービスを提供。自治体のオープンデータ基盤などに採用



※管理画面は実際の管理画面とは異なります。

出所：日本ラッドホームページ

事例：ビッグデータとクラウド（NTTコム：インテリジェンス）

店舗のPOSシステム端末をモバイル（iOS・Android）、クラウドで低コスト化
POSシステムは『レジ機能の提供』から『データ分析ツール』へ
NTTコミュニケーションズのBizホスティング Enterprise Cloudを採用（マレーシア）



適切な時間軸で、タイムリーに店舗や業態のコンディションを把握可能で、将来の打ち手として、具体的なアクションを促す仕組みを構築

⇒アジア展開（マレーシア拠点）

店舗で1日の予算達成状況と将来見込みを可視化

現状：3時間後に未達見込 ※夕方迄の貯金消化

Action！（From店長） ポップアップ作業指示

- ✓ 店外呼び込み、アップセルトーク、クロスセルトーク、会員へDM、バイトシフト見直し

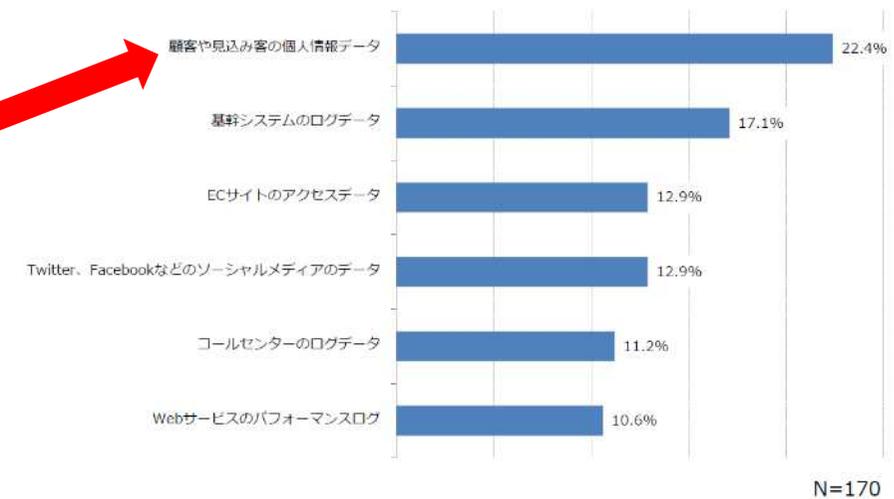
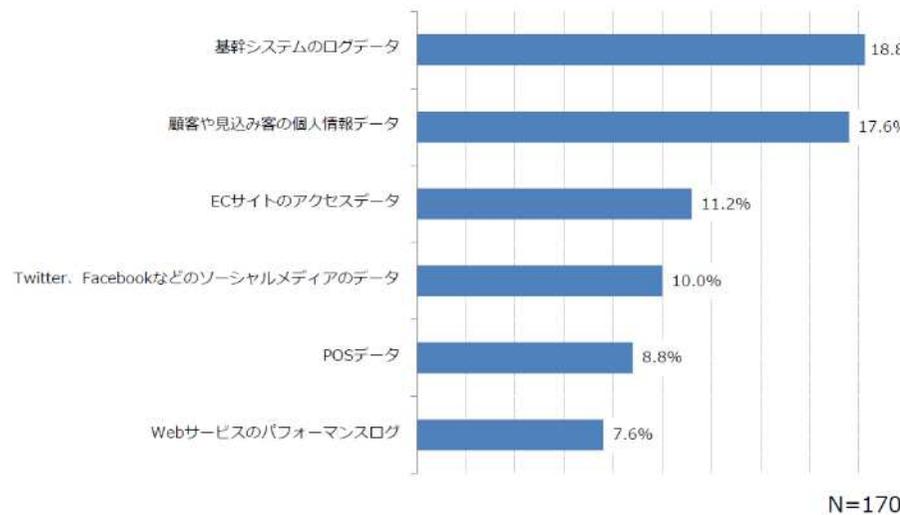
ex：人時売上が5,000円以下でアラームを発し、改善OPのリコメンド

データ分析の活用状況と今後の活用領域

「個人情報データ（17.6%→22.4%）」を分析したい企業が、増加傾向

現在、どのようなデータを分析していますか（複数選択可）

今後、どのようなデータを分析したいですか（複数選択可）

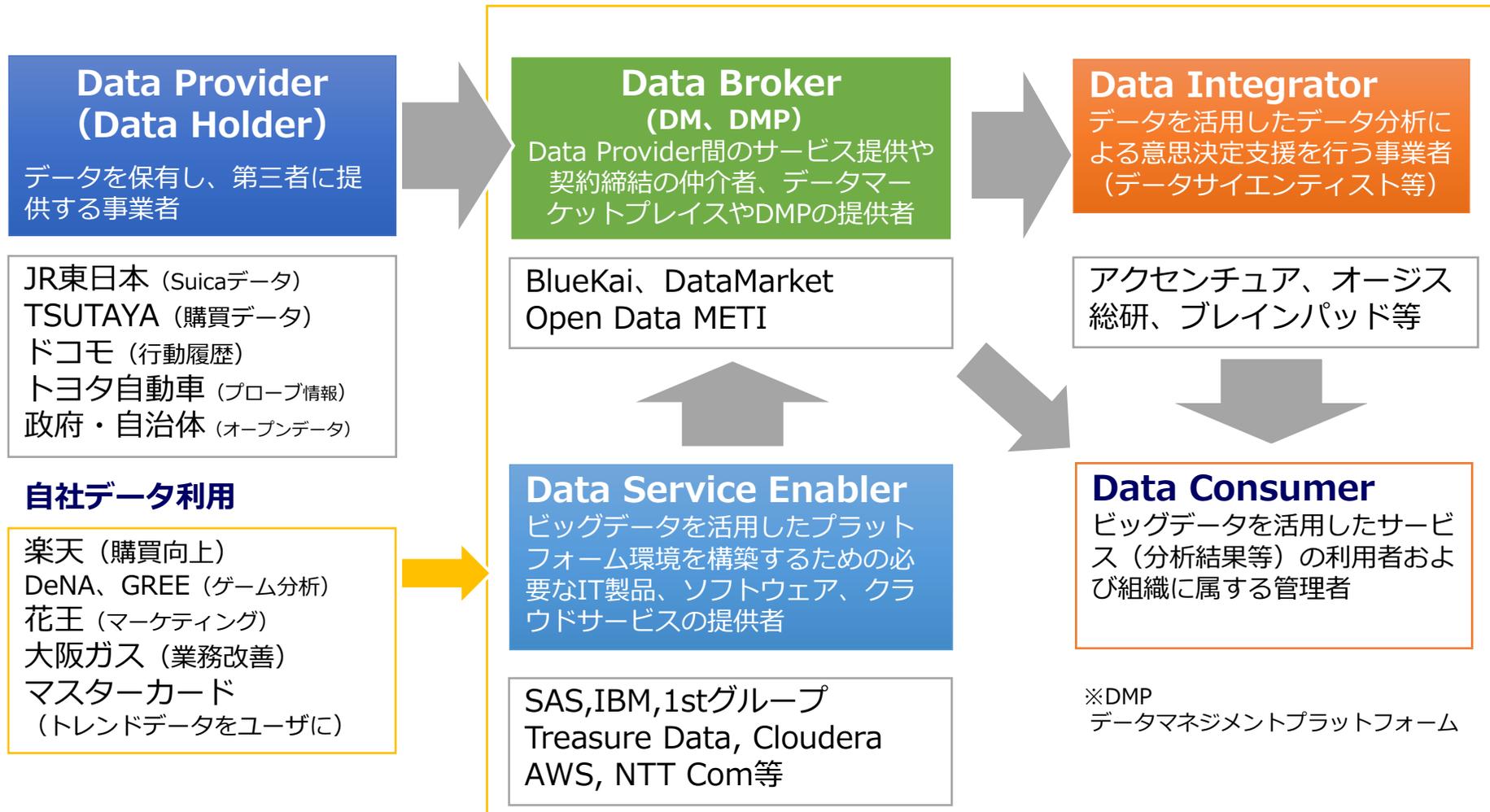


出所：TechTarget マーケティングのデータ活用に関するアンケート調査（2013年9月9日～9月25日）

ビッグデータビジネスの事業者のマッピング

Data Consumer	ビッグデータを活用したサービス（分析結果等）の利用者および組織に属する管理者
Data Provider (Data Holder)	ビッグデータを保有し、第三者に提供する事業者 （例：JR東日本、TSUTAYA、ドコモ、政府・自治体等）
Data Broker	Data Provider間のサービス提供や契約締結の仲介者、データマーケットプレイスの提供者（BlueKai等）
Data Integrator	データを活用したデータ分析による意思決定支援を行う事業者 （ブレインパッド<データサイエンティスト>等）
Data Service Enabler	ビッグデータを活用したプラットフォーム環境を構築するための必要なIT製品、ソフトウェア、クラウドサービスの提供者
Data Auditor	データのプライバシー、著作権、個別法などのアセスメント（評価）を行う第三者機関となる組織

ビッグデータに関するビジネスエコシステム例

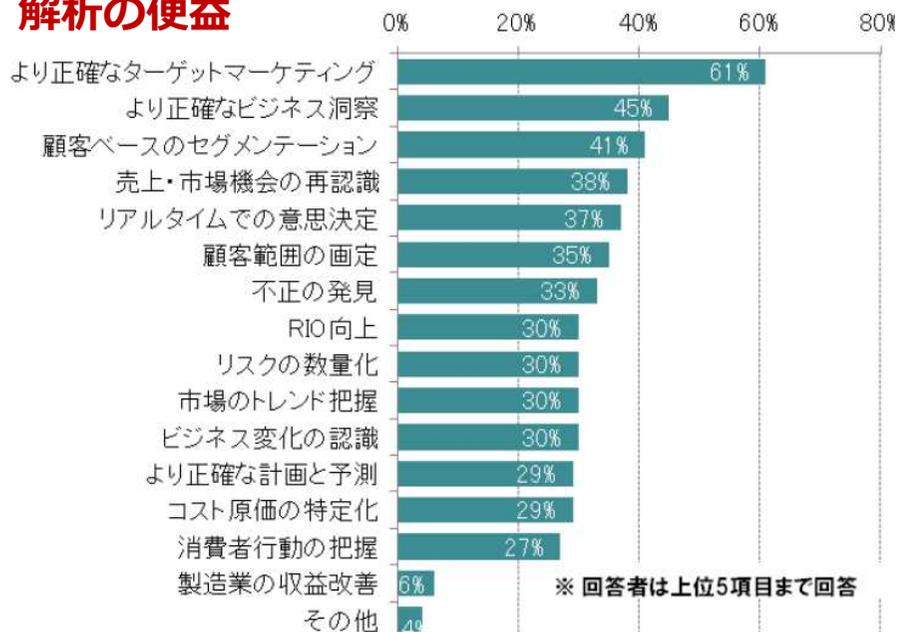


データサイエンティスト人材の状況

ビッグデータ解析の便益および障壁

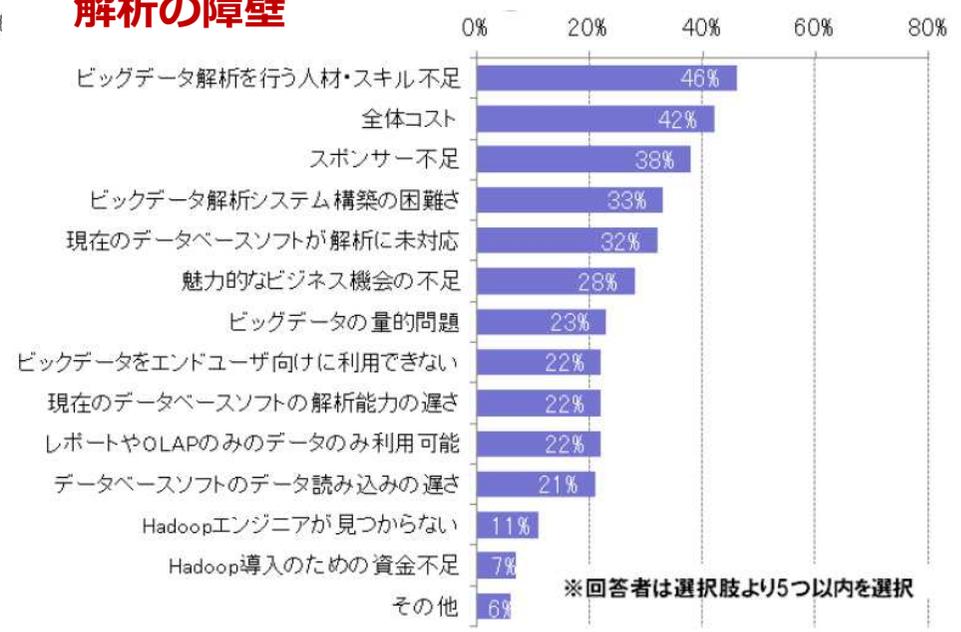
- ・ 便益においては、「より正確なターゲットマーケティング（61%）」が最も高く、「より正確なビジネス洞察（45%）」のビジネス・インテリジェンス系が続く
- ・ 障壁においては、「ビッグデータ解析を行う人材・スキル不足（46%）」が最も高く、全体コスト（42%）」が続く

解析の便益



より正確なターゲットマーケティング 61%
 より正確なビジネス洞察 45%
 顧客ベースのセグメンテーション 41%

解析の障壁



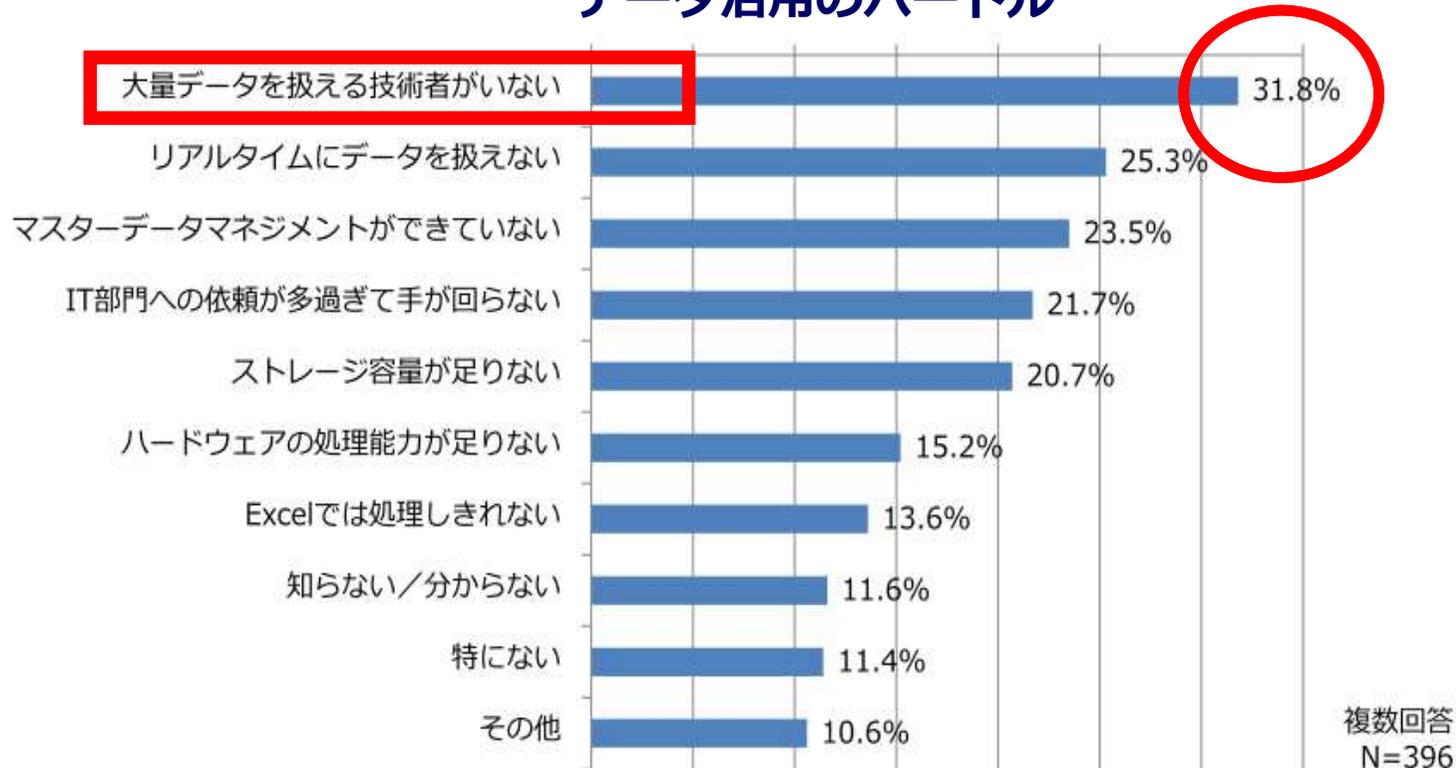
ビッグデータ解析を行う人材・スキル不足 46%
 全体コスト 42% スポンサー不足 41%
 ビッグデータ解析システム構築の困難さ 33%

BIG DATA ANALYTICS TDWI RESEARCH 2011 4Q

データ活用の課題

データ活用における課題は、「大量データを扱える技術者がいない」が最多で32%。「リアルタイムにデータを扱えない」(25%)、「マスターデータマネジメントができていない」(23%)などがこれに続く

データ活用のハードル

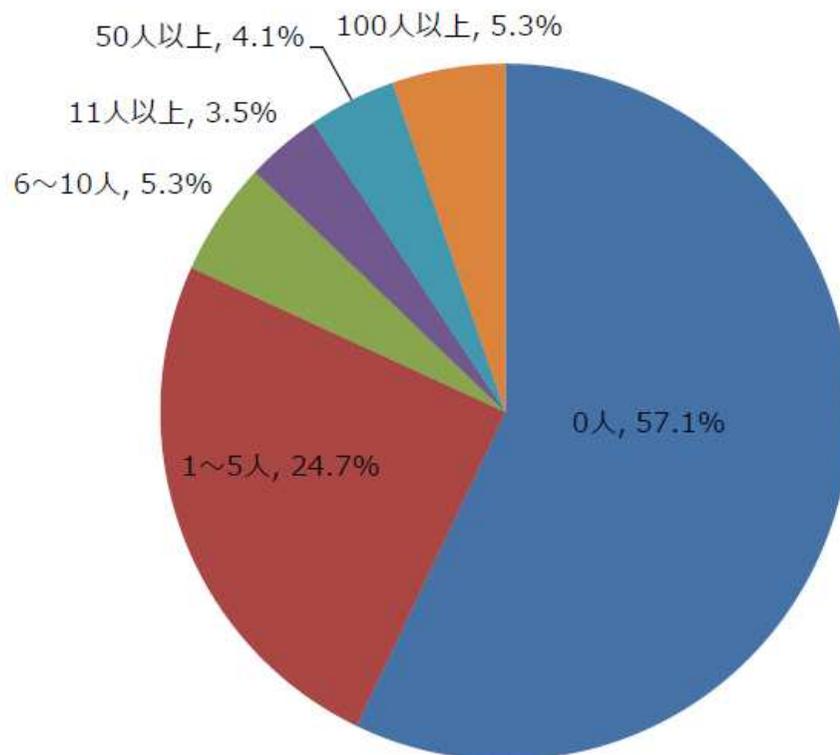


出所：TechTarget 「大規模データ活用動向についての調査」 2012年1月

各企業におけるデータサイエンティストの状況

57.1%の企業でデータサイエンティストが不在。100人以上が5.3%??

あなたの会社にデータ分析専任スタッフ（データサイエンティスト）は何人いますか。



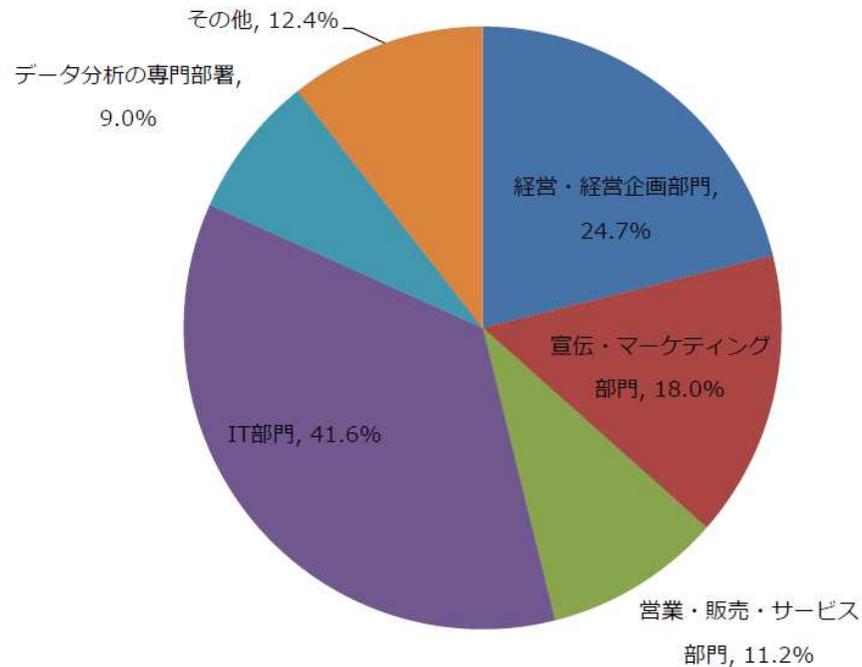
N=170

出所：TechTarget マーケティングのデータ活用に関するアンケート調査（2013年9月9日～9月25日）

データサイエンティストが所属する部門

IT部門(41.6%) 部門がもっとも多く、経営企画部門 (24.7%)、マーケティング部門 (18.0) と続いている。データ分析専門部署は9.0%にとどまる

データ分析専任スタッフがいると答えた方にお聞きします。データサイエンティストが所属する部門を教えてください。



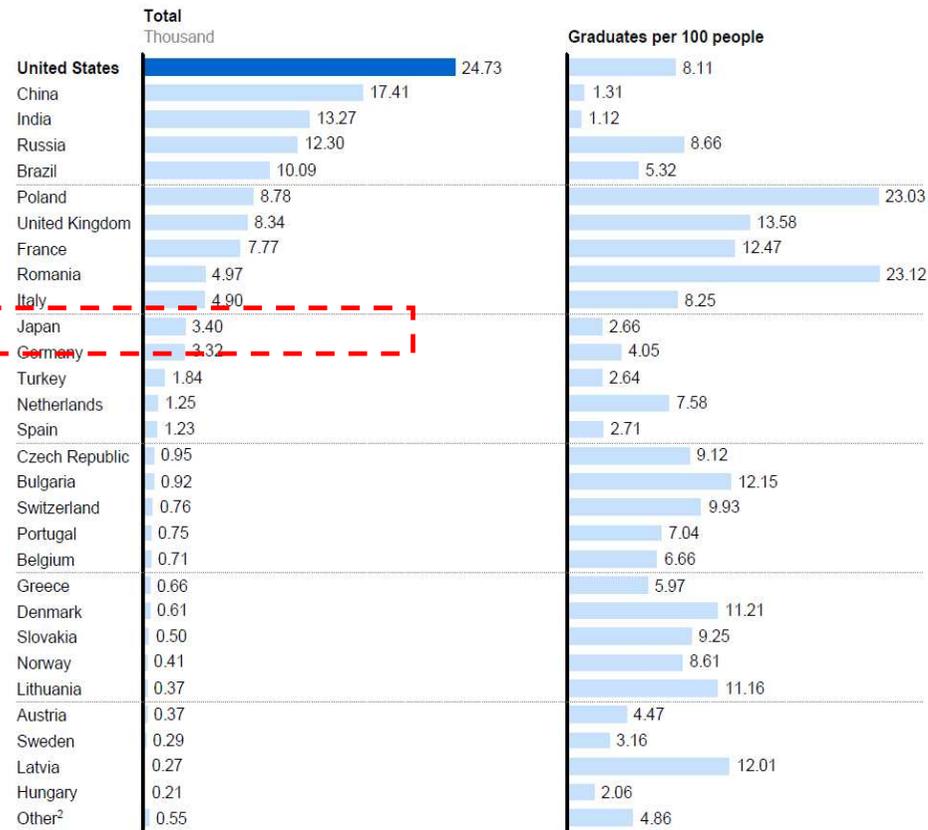
データサイエンティストがいる企業
N=89

出所：TechTarget マーケティングのデータ活用に関するアンケート調査（2013年9月9日～9月25日）

世界におけるデータ分析のスキルを有する卒業生数

- 統計学や機械学習に関する高等訓練の経験を有し、データ分析に係る才能を有する大学卒業生数は、米国の2万4,730人、中国の1万7,410人、インドの1万3,270人、日本は3,400人。

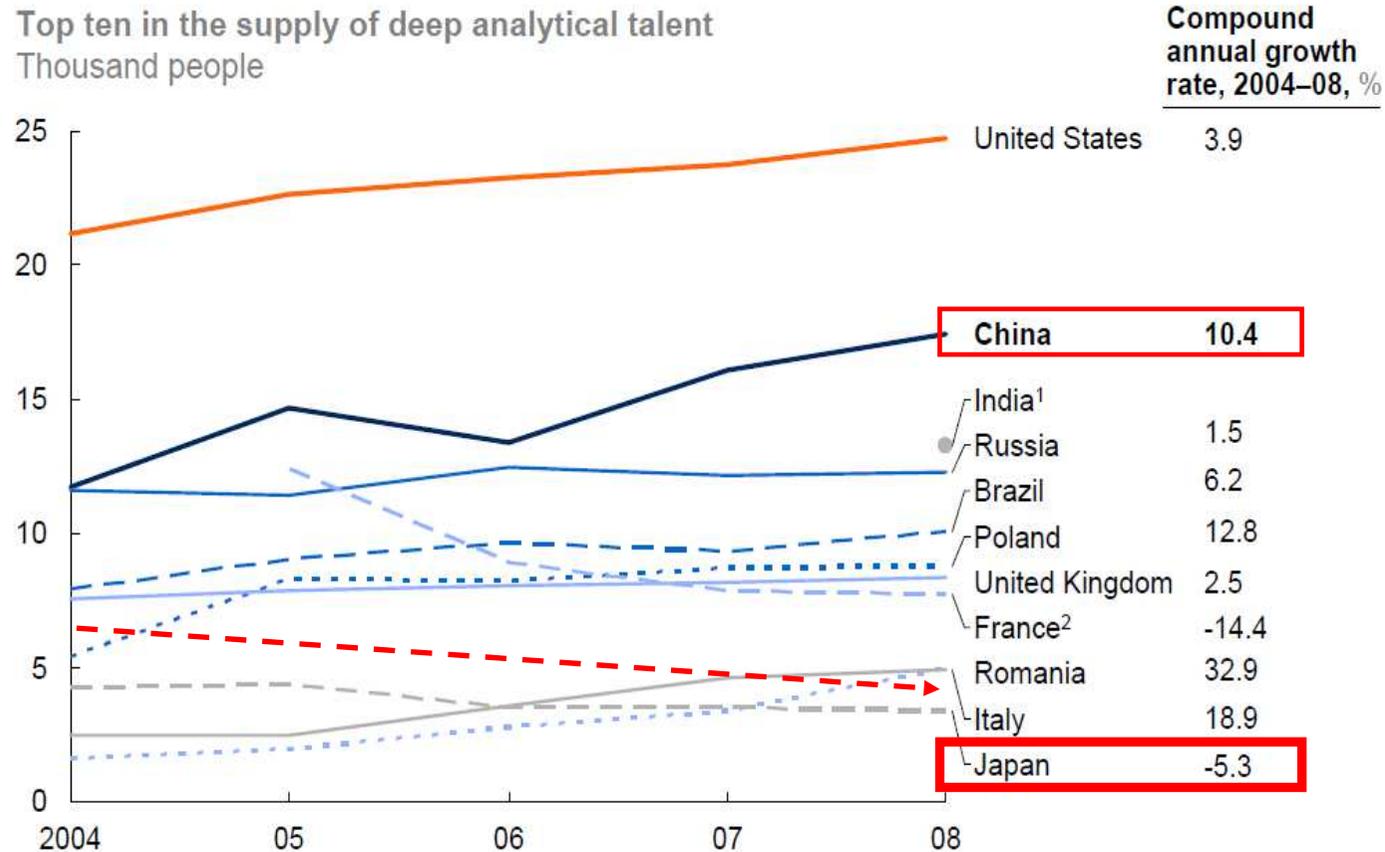
China	17,410
India	13,270
Russia	12,300
Brazil	10,090
Poland	8,780
UK	8,340
France	7,770
Romania	4,970
Italy	4,900
Japan	3,400
Germany	3,320



出所 : McKinsey Global Institute 「Big data: The next frontier for innovation, competition, and productivity」 2011.5

世界における分析のスキルを有する人材の推移

- 中国は年率平均10.4%の成長率でデータ分析の才能を有する人材が増加するなど諸外国が増加傾向にあるのに比べ、日本は年々減少(平均-5.3%)



出所 : McKinsey Global Institute 「Big data: The next frontier for innovation, competition, and productivity」 2011.5

日本におけるビッグデータ関連の雇用予測

- ・ 企業の情報システム部門が受け持つIT予算は次第に減少し、2015年にはIT支出企業のIT支出の35%がIT部門の予算外へ
- ・ 2017年にはマーケティング部門が行使できるIT予算が情報システム部門を上回る

米国では2018年までに、高度なアナリティクス・スキルを持つ人材が**14万～19万人不足**。大規模なデータセットのアナリティクスを活用し意思決定のできるマネージャーやアナリストが**150万人不足**

日本では、ビッグデータ関連の雇用が**36万5000人分** 増える見込み

25万人不足

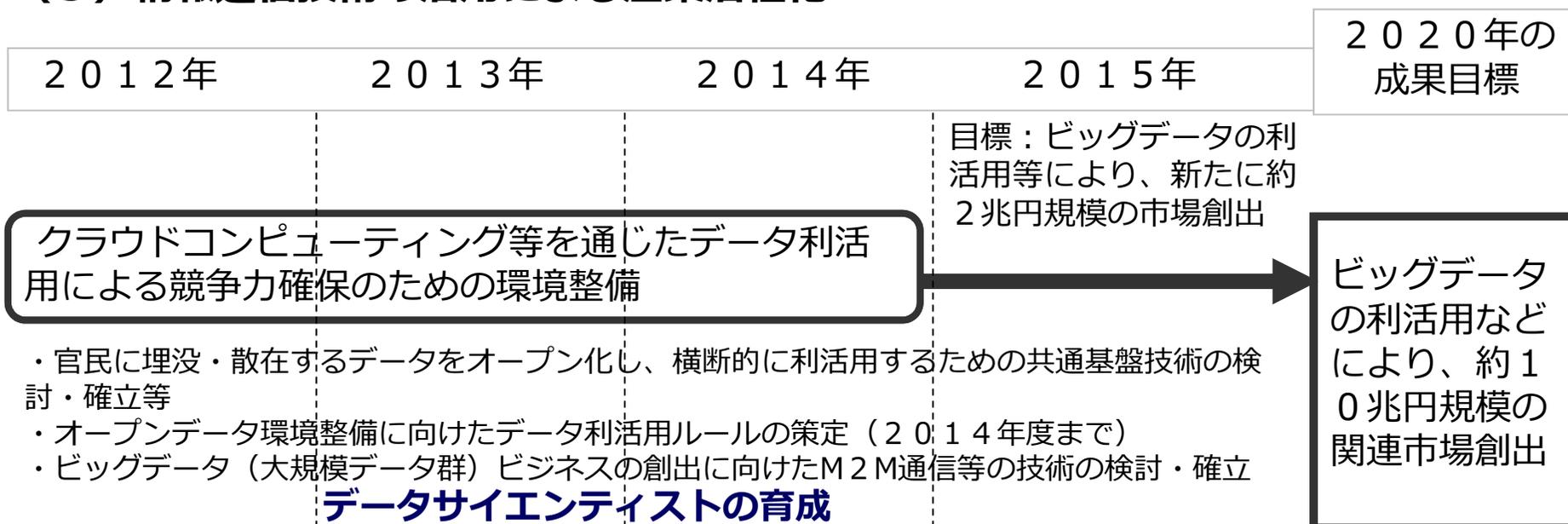
実際に雇用条件を満たせる人材は**11万人程度**

出所 : Gartner Symposium Report : 201x年に情報システム部門はどうするべきか?
<http://www.itmedia.co.jp/enterprise/articles/1210/04/news117.html>

政府におけるデータサイエンティストの育成

・政府は、日本の再生戦略、クラウド等を通じたデータ利活用による競争力確保のための環境整備において、2013年度までに実施すべき事項として、データサイエンティストの育成をあげている

(3) 情報通信技術の活用による産業活性化



〔出所〕 日本再生戦略 科学技術イノベーション・情報通信戦略 2012.7

政府におけるデータサイエンティストの育成施策（文科省）

事業名	課題名	課題概要	事業代表者
ビッグデータ利活用によるイノベーション人材育成ネットワークの形成	データサイエンティスト育成ネットワークの形成	情報科学，統計・機械学習の研究教育機関，それにデータを持つ現場としての研究機関・企業を広くネットワークし，認知度の向上，人材の流動化，ベスト・プラクティスの調査と共有，教育コンテンツの開発，それに国際連携を行う。本事業終了後も，人材育成が継続的に行われるよう，人材の出口戦略に特に留意する。	統計数理研究所 樋口 知之所長
ビッグデータ利活用によるイノベーション人材育成ネットワークの形成	スキルと実践を重視したビッグデータ・イノベーション人材育成プログラム	新規開講する「ビッグデータ技術演習」によりビッグデータスキルを体得させ，「ビッグデータ実践」により，その分析結果の関係機関による評価を受けながら成果をまとめ，実戦経験を積む。選択必修科目群「ビッグデータ技術論」，「ビッグデータシステム論」，「ビッグデータ活用論」の各科目群から1科目以上履修させ，スキルと実戦経験と関連知識を修得。	慶應義塾大学 理工学部 山口 高平教授

出所：文部科学省次世代IT基盤構築のための研究開発「ビッグデータ利活用のためのシステム研究等」採択課題

施策名	参加大学・機関	取組概要	達成目標・成果
文科省大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」	東京大学、大阪大学等8大学、6学会（応用統計学会、日本計算機統計学会）、8団体（大学入試センター、日本科学技術連盟等）	具体的な分析の手順・方法とそれらの数理的基礎、そしてそれらを適用する際の注意事項等を講述。講義に出席し必要な課題を提出し合格した受講生へは単位認定を実施	課題解決力を持つ人材を育成するために、大学における統計教育の標準的カリキュラム体系を策定し、その体系に基づく標準的な達成度評価制度を整備し、統計教育の質保証を行う。連携ネットワークによる認証に基づく共通単位互換制度を設ける

出所：文部科学省大学間連携共同教育推進事業

ステアリング・コミッティ宣言

ICTの飛躍的な進化により従来は可能だったことが可能になった。その代表例がビッグデータである。ビッグデータの活用を推進することで、企業の成長に貢献するだけでなく、社会全体にイノベーションを起こし、豊かな未来を創造しよう。

データの戦略的活用を推進するために

ユーザとベンダー、ユーザ同士、ベンダー同士が協力し合って、異なる分野・業界を横断的につなげていくことで、データをインテリジェンスに変え、新たな価値を生み出していく。

ICTの知識と統計の素養を備えデータを分析できるとともに、経営視点を持ち現場の業務にも精通するデータサイエンティストの育成を急がねばならない

パーソナルデータについても、説明責任を果たしプライバシーを保護することを前提に、個人や社会にとって利便性が高まることを理解してもらい、積極的に活用すべきである

企業内だけでなく農業や医療、交通などの分野、さらには道路や都市、エネルギーといった社会インフラにも適用範囲を拡げていく

「ITpro EXPO ステアリング・コミッティ」の宣言文 2013.8.21

データサイエンティストと人材マップ例

- ・データサイエンティストは、高度で幅広いスキルや知識、そして、コミュニケーション能力が求められる

導入のための目標設定、
企画（戦略部門）

データサイエンティスト

統計学、機会学習、プログラミングや可視化といったデータ分析に高度なスキルを有し、分析結果を新たな知に結びつけ、ビジネスの課題を創造的に解決する人材

1. データマイニング、機械学習、人口知能、自然言語処理
2. ビジネス・インテリジェンス（BI）、競合情報分析
3. 分析、統計。特にウェブ解析、A/Bテストや統計分析

データキュレーター/ データアナリスト

大量で、多種・多様なデータから、埋もれた知識を引き出し、異分野間をつなぐ専門家

- ・経営者×データサイエンス
- ・管理職×データサイエンス
- ・営業×データサイエンス
- ・ marketer×データサイエンス
- ・運用保守×データサイエンス

バイオインフォマティシャン

特定のビッグデータや業界に専門的に精通した人材
(ソーシャルメディア、顧客データ等、環境、医療等)

システム開発者

プログラミング言語の多様化・高度化に対応した人材
(開発言語、データマイニング用の専用言語)
(分散コンピューティング、分散処理システム、分散ファイルシステム)
(Hadoop、HDFS、MapReduce、RDB、NoSQL、KVS..)

クラウド（インフラ）エンジニア

ビッグデータを利用する際のクラウド環境の設計・構築・運用
(サーバ、ストレージ、認証・セキュリティ、ネットワーク等)

出所：文部科学省 アカデミッククラウドに関する検討会2011.5 などから作成

データサイエンティストの5つの資質・能力と業務例

データサイエンティスト 5つの資質

「Quantitative」

(物事を数学的命題として把握できる高い能力)

「Technical」

(統計処理に活用できるIT技術に関する知識)

「Curious & Creative」

(データの裏にストーリーを見い出せる能力)

「Skeptical」

(自らを疑うことのできる能力)

「Communicative and Collaborative」

(分析結果を事業につなげるためのコミュニケーション力)

出所：「データサイエンティスト」って何だ EMC Data Science and Big Data Analytics

データサイエンティスト 5つの能力

- データを扱う統計学の知識
- 仮説を基にデータを解析する手順を組み立てる能力
- データを解析するソフトウェアを使いこなす能力
- データを解析した結果をビジネスに適用する
ストーリーを組み立てる能力
- データを解析した結果を可視化し他人に伝える能力

データサイエンティスト 業務例

1. 業務でデータを使って改善できそうなところを探す
2. 改善のための仮説を立てて提案を作る
3. 社内で整理してプロジェクト化する
4. データを収集する・データを受け取る
5. データをクリーニングする
6. データを可視化する
7. データを集計する
8. データマイニングや機械学習を用いる
9. 結果を元に再調整や再実験を繰り返す
10. 結果を可視化してレポートにまとめる
11. レポートを関係者を共有する
12. 結果を元に有効と思われるビジネス
アクションを取る
13. ビジネスアクションの効果を測定して検証する
14. PDCAサイクルをまわす

出所：PFIセミナー データサイエンティストのつくり方 2012.9.13

出所：経産省 融合新産業創出に向けた動向調査事業 2013.3

IT事業者のデータサイエンティストの状況

企業名	データサイエンティストの人員	ビッグデータ関連組織概要
アクセンチュア	データ分析専任部隊 300人のうち100人	グローバルのアクセンチュア アナリティクス直下の組織。戦略の立案から解析、運用まで幅広く提供。ソーシャルメディア分析やSCMの最適化、研修・診断、M2Mなど6つの領域を軸にサービスを展開
日本ユニシス	50人。さらなる増員を検討	CRM/O2O、社会・産業インフラ、コンプライアンス・リスク管理、情報システム基盤構築の4つのシナリオ分類で、データを収集し、仮説検証やシナリオごとのアプリや適用ツール、ビッグデータ分析基盤の構築などを実施
日立製作所	データ分析専任部隊 300人のうち80人	データ・アナリティカル・マイスターは、自社のビッグデータによる価値の創出するための中心的位置づけで、より経営に近い視点で対応
富士通	100人。2015年までに倍増	BI/BAを担当してきた研究者、コンサルタント、プロダクト開発者、SEなどを集めて2011年1月にキュレーター組織を設立。ビッグデータを分析する専門家集団として100人強の技術者。SEなど数百人がビッグデータの関連サービスを提供できる体制構築
NEC	100人。2015年300人までに倍増	2012年2月にR&D部門等から選出されたメンバー計50人から開始。全社横断でHadoopやM2Mをはじめとする先進技術の開発強化や顧客との実証実験への参画、マーケティング・プロモーションなどを推進
ブレインパッド	130人（従業員数）	データマイニング、数理計画（OR）による最適化をユーザ企業のビジネスに展開し、データに基づく合理的な意思決定によるプロジェクト・ゴール（収益拡大、コスト削減など）の達成を支援する総合的なコンサル、ソフトウェアのライセンス提供やシステムを構築
アイズファクトリー	10人。従業員30人	理学博士や環境学の研究者らが経営するデータサイエンティスト集団。ベネッセの販売支援や久光製薬のサロンパスの評判測定などの事例
オージス総研	データサイエスセンター	大量データ高速処理培ってきた技術を組み合わせ、お客様に適したシステムを構築。大量のデータを統計手法を用いて解析し、ビジネスに役立つソリューションを開発等

日本ユニシスのデータサイエンティスト

- データサイエンティストのような能力を一人でカバーすることは難しく、チームとしてデータサイエンティストの機能を持たせる

データサイエンティストに必要なスキル群



出所：日本ユニシスホームページより

データ解析分類、分析方法例（オージス総研様ページ）

分類	分析方法	例
予測型	販売量予測	過去トレンドや気温影響などから販売量を予測
	医療費予測	レセプトデータや健診データから医療費を予測
異常検知型	機器故障予兆分析	故障前に生じる異常状態を検知することで予兆
	サイレント故障分析	正常時の状態値からの逸脱を検知
	サーバログ解析	ログデータもとに異常を予兆
最適化型	車両配置最適化	到着時間をシミュレーションすることで最適配置を探索
	在庫最適化	予想需要と品質期限から最適在庫量を導出
	ワークスタイル分析	P C 操作ログからアプリの使用状況を解析
自動化型	シフトスケジューリング	労働条件を満たす勤務計画を自動生成
	プラントオペレーション	複雑なプラントの操業計画を自動生成
判断型	顧客ターゲティング	ターゲット顧客を選定
	エリアマーケティング	ターゲットエリアを選定
	Webサイトアクセス分析	各サイトにどれだけのアクセスがあるかを分析
発見型	口コミ分析	ブログなどのテキスト情報をマイニング
	アクセスログ分析	自社Webサイトのアクセス情報を分析
	コミュニケーション分析	メールログから人の繋がりを可視化
	商品分析	同時に買う可能性の高い商品群を抽出
リスク計量型	倒産リスク計量	格付と倒産データから倒産リスクを算出
	市場リスク計量	モンテカルロシミュレーションを用いたリスク計量
社外データ活用	気象データ活用	販売量予測、プラント操業計画、売上予測などに活用
	渋滞データ活用	車両配置最適化、配送ルート最適化などに活用

出所：オージス総研ホームページより

ユーザ企業のデータサイエンティストの状況

企業名	組織	取組概要
日本航空	マーケティングG データ分析官	チケット販売サイトで旅の特集ページのPV回数を増加させるために、会員サイトの閲覧動向分析。サイトを訪れても旅特集の閲覧率が低い →6種類のバナーを用意し、数万人の流入増加
大阪ガス	ビジネスアナリシス センター(9人)	数理計画、統計解析、環境、気象、エネルギー、ITなどを専門家集団。データ分析で家庭用ガス機器の修理完了率の改善など、意思決定プロセスを変革
P&G	ビジネスアナリスト	消費行動などの意思決定に必要なデータを見える化し、顧客が持つ潜在的なインサイトを探り、ビジネスアナリストは意思決定に必要なデータとオプションを提供
花王	データサイエンティスト(6人)	マーケティングに特化し、データ分析の結果を商品施策や販促に展開。リピート率を分析しブランド価値向上。他社への流出率も分析し、対抗商品を出し歯止めを実施
東芝	分析官	ライン設計、搬送、生産の3つのデジタルシュミレーションによる分析で、半導体生産の効率化
エステー	分析官	「ムシューダ指数」「オンパックス指数」をウェザーニュースが提供する気温や降水回数などのデータと商品の販売実績を組み合わせ、売れゆきを指数で予測。
ロイヤリティー マーケティング	LM Analytical LAB(20人)	共通ポイント「Ponta」のデータを分析し、提携会社にフィードバック。キャンペーン像による来店者数の増加などの分析し、投資対効果の最大化に寄与。
マスターカード	マスターカード・ アドバイザーズ	650億件の取引データを収集して分析し、ビジネスと消費者トレンドを予測。トレンド情報を外部に販売。高度な分析情報の販売収入がカード会社の利益に。

(参考) GREEの採用事例 (2012.12時点)

事業戦略部門

グローバル戦略

＜アナリティック・エンジニア＞
(事業戦略上重要である数値指標を設定)

マイニングアナリスト

(分析からサービス価値向上を実現する業務)

BIスペシャリスト

(分析要件の定義を担当)

BIエンジニア

(BIシステムの構築)

エンジニア部門：ソフトウェア

自然言語処理・機械学習

(大規模な投稿データやユーザ行動の分析)

分析PF/データマイニング

(数値指標の提案、分析環境の提供)

大規模データ

(大規模データ処理のためのインフラ構築運用)



出所：GREE 採用ページ 2012.12

(参考) GREEの採用事例 事業戦略部門 (2012.12時点)

職種	職種内容	登録資格 (望ましい経験/能力)
グローバル戦略 (アナリティック・エンジニア)	海外におけるモバイルネット事業に関する社内外の膨大なデータから、事業戦略上重要である数値指標を設定し、その数値の可視化及び分析	<ul style="list-style-type: none"> ・ビッグデータの分析経験 ・PHP (もしくはPerl、Rubyなど) の実務経験 3年以上 ・BIツールの利用経験 (Tableauなど) ・ネット業界における実務経験/ネット業界に関する専門知識 ・通信業/メディア業における戦略策定プロジェクトの経験
マイニングアナリスト	社内に大量に蓄積されているデータの分析環境の構築、及び専門的な分析を実施し、グリーというサービスの価値の向上を実現する業務を担当 データの分析から得られた法則や知見を元に、事業の企画・改善のディレクションを実施。 社内のログ管理チームと連携し最適な分析環境を構築していく業務も。	<ul style="list-style-type: none"> ・レコメンデーションエンジンや検索エンジンなどの業務経験 ・大規模分散処理技術の知識または開発のディレクション経験 ・R,SASなどの統計解析用ソフトウェアの利用経験
ビジネスインテリジェンス・スペシャリスト	サービス部門と密にコミュニケーションを保ち、問題点の洗い出しや分析要件の定義を担当。 分析にあたって必要に応じてマイニングアナリストや、ログ管理チームなど関係部署と連携して分析結果をまとめ、サービス部門に知見、示唆を伝達し、実際の施策まで反映させ、ビジネス上の成果を上げる一連の業務をコントロールする。	<ul style="list-style-type: none"> ・WEB業界での業務経験 (分析、企画、開発、数値管理など) ・Python、Ruby、PHP、Perlなどを用いたコーディング能力 ・R言語などの統計解析用ソフトウェアの利用経験
BIエンジニア	<ul style="list-style-type: none"> ・データ分析環境の整備からBIシステムの構築 ・サービスに関わる膨大な蓄積されたデータから、事業戦略上重要である数値指標を設定し、数値の可視化、レポートの作成、特異点検出など。 ・BIシステムを構築してレポートの自動生成や数値分析の自動化を実施 	<ul style="list-style-type: none"> ・ビッグデータの分析経験 ・PHP (もしくはPerl、Rubyなど) の実務経験 3年以上 ・BIツールの利用経験 (Tableauなど) ・分析ソフトを高度に使いこなせる (SAS,Rなど) ・ビジネスレベルの英語の読み書き ・サーバー・DBのセットアップなど分析環境の整備が一人でできる。

出所：GREE 採用ページ 2012.12

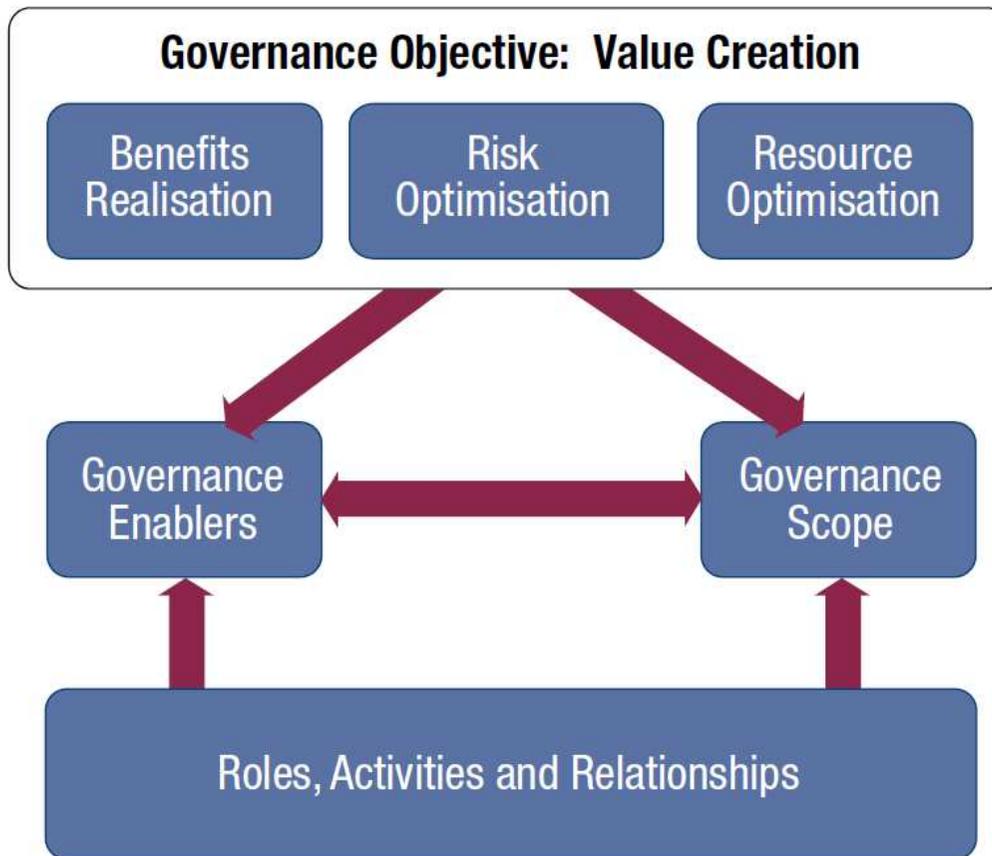
(参考) GREEの採用事例 エンジニア (2012.12時点)

職種	職種内容	登録資格 (望ましい経験/能力)
自然言語処理・機械学習	<ul style="list-style-type: none"> ・大規模な投稿データやユーザ行動の分析 ・不正に対して堅牢なパトロールシステムの構築、運用 ・サービスの健全な利用を促進するインタフェースの提案、構築 	<ul style="list-style-type: none"> ・自然言語処理を用いた開発経験 ・機械学習や統計学に関する知識と経験 ・大規模データシステムの運用経験 ・英語によるコミュニケーション能力
分析プラットフォーム/ データマイニング	<ul style="list-style-type: none"> ・新たな視点や知見を得られるような数値指標の提案 ・各部門との数値指標に関するコンサルティング、分析環境の提供 	<ul style="list-style-type: none"> ・オリジナルのソフトウェア/アプリケーションの設計経験 ・ギガバイト以上の入出力を扱うスケーラブルなアプリケーションの開発、運用経験 ・大規模RDBMS環境の整備、運用経験 ・データ構造やアルゴリズム、計算量などの基礎知識 ・確率分布、統計、検定などの基礎知識 ・基礎的な行列の計算や集合などの数学の知識 ・回帰分析などの統計分析手法の知識 ・パターンマイニング、分類器、クラスタリングなどの機械学習の知識
大規模データ	<ul style="list-style-type: none"> ・ログ、数値指標を扱うサーバ、データベースなどのインフラの整備、運用、管理 ・ログ、数値指標の分析環境の整備、運用、管理 	<ul style="list-style-type: none"> ・テラバイト単位のデータを用いたシステムの設計、運用経験 ・大規模RDBMS環境の整備、運用経験 ・Webアプリケーション開発の知識、経験 ・分散技術・ネットワークプログラミングの知識、経験

出所 : GREE 採用ページ 2012.12

データ活用における組織リスクと対応

- ・企業がデータを基に意思決定するようになる場合、不正確、不完全、不正に操作されたデータは企業にとってリスク。組織によるデータガバナンスへ



データを基に意思決定を行う重要性が高まるほど、意思決定の前提となるデータの品質が求められる。

特に経営等の重要な意思決定に関連するデータについては、そのデータの収集方法の明確化と、データに対するアクセス権限を設定するなどのデータ管理が必要

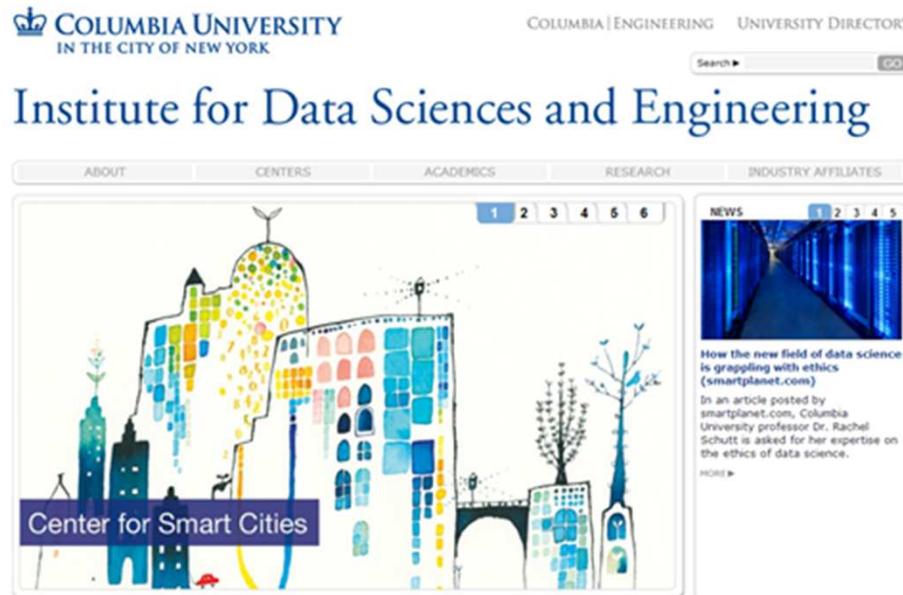
出所：COBIT 5, ISACA, USA, 2012, figure 8

出所：経産省 融合新産業創出に向けた動向調査事業 2013.3

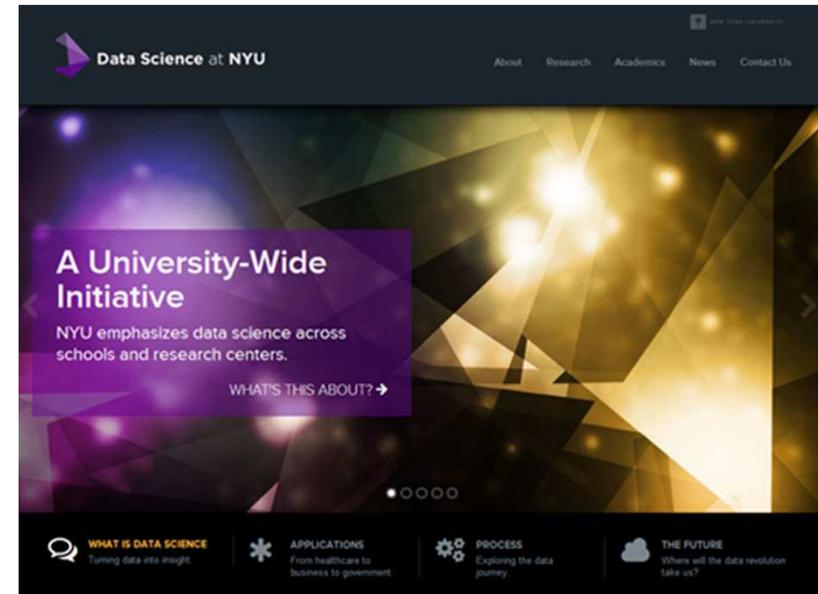
大学等におけるデータサイエンティスト育成プログラム

企業名	プログラム名	目的・概要	対象	発表月
立教大学	ビッグ・データをテーマにした授業	学生が自ら想定顧客を決定し、ビッグ・データを活用した新しいビジネスモデルやユースケースの提案を最終目標。EMCジャパンが講義を支援	立教大学経営学部国際経営学科3年次生	2012年5月
会津大学	ベンチャー基本コース	ビッグデータの活用事例や効果、データ活用手順、システム構成、分析の手法など、ビッグデータ関連の最新情報を7回にわたってレクチャー。福島県を起点とした産業の活性化や起業支援などに貢献	会津大学200名の学生	2012年10月
東京大学、大阪大学等8大学、6学会、8団体	文科省大学間連携共同教育推進事業における取組「データに基づく課題解決型人材育成に資する統計教育質保証」	具体的な分析の手順・方法とそれらの数理的基礎、そしてそれらを適用する際の注意事項等を講述。講義に出席し必要な課題を提出し合格した受講生へは単位認定を実施	実質科学においてデータ解析を研究の道具として実際に使う学生（+研究者）	2013年8月
ヴァージニア大学	ビッグデータ・インスティテュート (Big Data Institute)	建築や医学、芸術といった幅広い分野の知識を学ぶことで、従来コンピュータには関心を持っていなかった建築や医学、エネルギーや芸術等、様々なセンスと能力を持った人材を養成	ヴァージニア大学 研究所の学生	
コロンビア大学	定量分析学 (Quantitative Studies)	ビッグデータに関する新しい修士コース	コロンビア大学のデータサイエンスおよび工学にかかわる学生	
スタンフォード大学		大規模なデータセットから情報を抽出・分析するためのデータマイニングや機械学習の習得		
ニューヨーク大学	データサイエンス	データサイエンスにおける大学レネイの研究施設を開設	18の大学など	

海外大学におけるデータサイエンス関連ページ



コロンビア大学
Institute for Data Sciences and Engineering
<http://idse.columbia.edu/>



ニューヨーク大学
Data Science at NYU
<http://datascience.nyu.edu/>

ワシントン大学 <http://www.pce.uw.edu/certificates/data-science.html>
バークレー大学 <http://requestinfo.datascience.berkeley.edu/index.html>

企業におけるデータサイエンティスト育成プログラム

EMCジャパン：データ・サイエンティストトレーニングコース 「Data Science and Big Data Analytics」

コース概要

1日目	2日目	3日目	4日目	5日目
ビッグデータ分析入門+ データ分析のライフサイクル	データ分析の基本 ～[R]を使って	先進的分析の理論と手法	先進的分析の技術とツール	分析結果の可視化と プレゼンテーション +課題チャレンジ
<ul style="list-style-type: none"> ● Big Dataの概要 ● 分析実務の現状 ● データサイエンティストとは ● 業界別のBig Data分析 ● データ分析のライフサイクル 	<ul style="list-style-type: none"> ● R言語の基礎知識 ● データの調査と分析 ● モデル構築と評価理論 	<ul style="list-style-type: none"> ● K平均法クラスタリング ● アソシエーション・ルール ● 線形回帰 ● ロジスティック回帰 ● 単純ベイズ分類機 (Naive Bayesian Classifier) ● 決定木 ● 時系列分析 ● テキスト分析 	<ul style="list-style-type: none"> ● 非構造化データの分析 (MapReduceとHadoop) ● Hadoopエコシステム ● In-database 分析-SQLの要点 ● In-database分析で活用する先進SQLとMADlib 	<ul style="list-style-type: none"> ● 分析プロジェクトの実施と運用 ● 最終成果物の作り方 ● ビジュアル化のテクニック <p>+課題チャレンジデータ分析ライフサイクルの適用業務(ケーススタディ)</p>

基本から応用的な分析方法の知識や、分析に関する技術とツール、分析結果の表現方法など、データ・サイエンティストとしての実務内容を網羅

対象

BI / 分析 / ビッグデータの専門チームを統括するマネージャー、現役ビジネス・アナリスト、データ・サイエンスおよびビッグデータの分野に進出したい大学新卒者および大学院生など

出所： <http://japan.emc.com/microsites/japan/data-science-team/pdf/data-scientist-training-2h13.pdf>

データサイエンティスト育成を推進する団体

一般社団法人 データサイエンティスト協会



一般社団法人
データサイエンティスト協会

▲ サイトマップ ■ お問い合わせ

HOME データサイエンティスト協会とは ニュース 活動内容 資料ダウンロード 入会のご案内

データサイエンスに精通した
プロフェッショナルの育成を目指して

2013年7月16日スタート

データサイエンティスト協会について

当協会は、新しい職種であるデータサイエンティストに必要なスキル・知識を定義し、育成のカリキュラム作成、評価制度の構築など、高度IT人材の育成と業界の健全な発展への貢献、啓蒙活動を行っています。

入会のご案内
MEMBERSHIP

<http://www.datascientist.or.jp/>

データサイエンティストに必要なスキル・知識を定義し、育成のカリキュラム作成、評価制度の構築など、高度IT人材の育成と業界の健全な発展への貢献、啓蒙活動を実施

データビジネス創造・ラボ

慶應義塾大学SFC研究所、ブレインパッド、アクセントチュアが、データサイエンティスト育成カリキュラムの構築、ならびに先進的なビッグデータの集計・解析手法などに関する共同研究を実施する「データビジネス創造・ラボ」を創設

IT、アナリティクス、デザインの融合方法の研究

各種データの収集に向けたあるべき情報システムの研究

データ処理方法などのアーキテクチャ構築の研究

データ分析手法の研究

データを組織戦略に活用できる人材の研究

人材育成

データサイエンティストの育成

教育カリキュラムの立案

ケーススタディ執筆

非営利・営利組織の事例調査、およびケースの作成

<http://www.brainpad.co.jp/news/2013/09/19.html>

(学生向け) データサイエンス・アドベンチャー杯

・実務で統計・データ分析を行う社会人ならびに統計学や情報学一般に関心を持つ学生の方々を対象に、分析アイデアおよび分析スキルの優劣を競うコンテスト



使用必須のデータとしてJSTが長年収集し整理・体系化してきた科学技術分野の論文情報を中心とするデータ (JST科学技術データ) が、使用必須の分析ツールとして「SAS(R) Analytics Pro」や「SAS(R) Enterprise Miner」などのSASソフトウェアを貸与

審査委員長 長尾 真 氏 (JST科学技術情報特別主監, 京都大学名誉教授)
審査委員 (50音順)
大向 一輝 氏 (国立情報学研究所 コンテンツ科学研究系 准教授)
小野寺 夏生 氏 (一般社団法人情報科学技術協会 (INFOSTA) 会長)
鈴木 良介 氏 (株式会社野村総合研究所 主任コンサルタント)
林 雅之 氏 (NTTコミュニケーションズ株式会社 主査, 国際大学GLOCOM 客員研究員)
南川 敦宣 氏 (KDDI株式会社 新規ビジネス推進本部 課長補佐)
森川 富昭 氏 (慶應義塾大学 大学院政策・メディア研究科/環境情報学部 准教授)

本コンテストの実施を通じて、**未来のデータ・サイエンティストが分析アイデアや分析スキルを切磋琢磨できる機会**を提供

主 催 : SAS Institute Japan株式会社, 独立行政法人科学技術振興機構
後 援 (50音順) : KDDI株式会社, 株式会社ジー・サーチ, 日本統計学会, 株式会社マイナビ
課 題 : JST科学技術データを使用して※, 柔軟な発想でテーマを自由に設定し, 統計・データ分析を行う
※JST科学技術データを使用していれば他のオープンデータ等を組み合わせた分析も歓迎
応募作品 : 分析テーマ, 分析手法および分析結果について, 分析概要 (400字程度) とプレゼンテーション資料 (PowerPoint形式) を提出
日 程 :
<エントリー開始> 2013年10月8日 (火) より公式Webサイトにてエントリー開始
<エントリー締切> 2013年12月6日 (金)
<応募作品締切> 2014年1月31日 (金)
<予選 (非公開) > 2014年2月17日 (月) 予定
<本選 (公開) > 2014年3月8日 (土)

出所 : データサイエンス・アドベンチャー杯 ホームページ

まとめ

<ビジネス>

- ・ビッグデータにおけるビジネスエコシステム

<データサイエンティスト>

- ・ビッグデータ解析における人材・スキル不足
- ・データ解析を学ぶ卒業生は日本は減少傾向
- ・各企業においても多くはデータサイエンティストは不在
(事例は増えてきているが、専門部署はこれから)
- ・政府(文科省)や企業なども支援プログラム
(海外の大学のカリキュラムと比べると見劣り)

最後に

自律的に動き、自己学習する特性を持つ
⇒スマートマシーン+クラウド

「Movers（移動する）」

「Sages（賢者）」

「Doers（行動する）」

出所：ガートナー「2014年の戦略的テクノロジー・トレンドのトップ10」2013.10

**2030年には、
世界中の全雇用の50%、
20億人分の仕事が機械化でなくなる**

調査会社 ダビンチインスティテュート

マイクログリッド、自動運転、3Dプリンター、ロボットの4つのテクノロジーの進化

**2045年には、
コンピューターの性能は現在の約
264万倍の性能にコンピューターの
能力が全人類の知能を上回る可能性**

ムーアの法則「半導体チップの集積度は、約18カ月で2倍になる」

ご清聴ありがとうございました