

データ・サイエンスはチームスポーツ。  
チーム・プレイヤーとしてのスキルはお持ちですか

IBM

1

---

はじめに

2

---

データ・サイ  
エンティスト

3

---

データ・エン  
ジニア

4

---

開発者

5

---

ビジネス・アナ  
リスト

6

---

データ・サイ  
エンス・チーム:  
新たな変革の  
担い手

7

---

参考資料

---

## はじめに

ビッグ・データ革命にとって大きな課題は、今日の広大なデジタル・エコノミーによって生まれる情報を、すべて理解することです。利用可能な範囲のデータを残らず収集できれば上出来ですが、その大量の情報から価値を引き出すにはどうすればいいのでしょうか。

手元のデータが増えるほど、レポートも戦略上の指針も質が向上します。もちろん、そのデータを素早くインテリジェントに分析し、有益なインサイトによって実践につなげることができればの話です。そうでなければ、データが増えるほど問題も増えかねません。データの散乱、ストレージの惨状、セキュリティー・リスクといった事態に陥り、ビジネス・チームは機能せず、IT スタッフは過剰な負担を抱えることになるでしょう。



**では、本当にデータから得たいものは何か、どのデータにその答えがあるのかを見極めるにはどうすればいいのでしょうか。**

生の指標や記録を実際にビジネスの役に立つ(また、ビジネスを後押しする)情報に変える方法は、あるのでしょうか。ビジネス上の成果をあげるために必要なスキルを、お持ちでしょうか。

組織として成功するために、少数の優秀な個人に頼るのではなく、さまざまなスキル・セットと職務を集めたデータ・サイエンス・チームを築きます。

現実には、複数の人がチームに属してデータから成果を生み出しています。分析は、その基盤となるデータを収集、構築、分析しているチームの能力に依存します。

**データ・サイエンス・チームは、どのように機能するのでしょうか。**データ・エンジニアやフロントエンドの開発者などは、プログラミングにより、データ・サイエンスに使いやすさをもたらします。数学の専門家あるいはデータ・サイエンティストの立場では、統計アルゴリズムを用いてデータにパターンを見いだします。分析に関わ

るプロセス全体では、優先されるビジネス目標にあらゆるものが合致していなければならない、それを導くのはビジネス・アナリストの視点です。こうしたメンバーが一体となって、データ・サイエンス・チームは今日の課題や問題を克服し、明日の新しいチャンスと可能性を作り出します。

では、データ・サイエンス・チームのメンバーにふさわしいのはどんな資質や能力で、お互いはどう補い合うのでしょうか。**[このまま続けてお読みください。](#)**

## データ・サイエンティスト

データ分析を通じて問題を解決し、疑問に答えることが、データ主導の今の世界ではにわかに常識となりつつあります。現在のビジネスでは、現実的な実験がどこでも可能になってきたため、データ・サイエンティストはこうした企業の原動力を駆り立て、方向を決めて加速する役割へと進化しつつあります。

優れたデータ・サイエンティストは、ビジネス上の問題にただ対処するのではなく、組織にとって最も大きい意味のある問題、その一点に集中します。

データ・サイエンティストは、「ユニコーン」と称されることもあります。数学、統計学、各分野の専門知識、コミュニケーション能力などのスキル・セットと、さまざまな職務を担える、希少な資質をあわせ持っているからです。基本的に、データ・サイエンティストの職務は隠れたパターンを見つけ出すことにあります。それを達成するために駆使するのが、機械学習、モデリング、統計や視覚化といった（もちろん、それだけではありません）高度な分析手法です。

多くの場合、データ・サイエンティストはモデルを構築して結果を予測し、基本パターンを探り出します。そのときの戦略が、今後の成果を改善するうえで実践的に利用できる知見を導き出すことです。



データ・サイエンティストは、データの成果を確立する過程のどの段階にも関与するため、データによって問題解決を総合的な視野で見渡す必要があります。とは言え、あらゆる分野の専門家になることはできません。だからこそ、チームが重要になってきます。

データ・サイエンティストは、新しい予測モデル、ビジネス・ルール、統合ロジックを配備して「ネクスト・ベスト・アクション」に応用することで、たえず実験を続けています。

熟練のデータ・サイエンティストは、複数の異質なソースから集めたデータを探究・調査します。入手したデータはすべて丹念に研究しますが、その目的は過去のデータに新しい情報を結び付け、関係性や傾向を見いだすことです。それが、競争上の決定的な優位につながったり、ビジネス上の差し迫った問題を解決したりします。ただ単にデータを集め、レポートにするだけで

はありません。さまざまな角度からデータを確かめて、それが意味することを判断し、結果の応用方法を推奨するのがデータ・サイエンティストの仕事です。クエリーが正しいことを確認しなければならず、健全なモデルと信頼できるデータによって結論を補強する必要もあります。データ・サイエンティストは、経営陣や上層部への提言を求められることも少なくないからです。



## データ・サイエンティストの主なスキル

データ・サイエンティストを特徴付けるのは、鋭いビジネス感覚と、営業および IT 双方のリーダーに結果を伝える能力です。その伝え方が、ビジネス上の課題に対する組織の臨み方を大きく左右します。データ・サイエンティストが、IT 部門と経営幹部との連絡役になることも少なくありません。したがって、データ・サイエンティストはどちらの言葉でも話ができて、各階層のデータを理解できる必要があります。単にデータの専門家というだけでは済まされません。つまり、ビジネスを確実に理解するとともに、反論があったときには断固として結論を支持しなければならないということです。

データ・サイエンティストは、探究心と好奇心が旺盛です。精査し、疑問を持ち、what-if 分析を実行しては、従来の仮定やプロセスを疑います。データ・サイエンティストの技術スキルとして一般的に求められるのは、複数のプログラミング言語、Apache Hadoop や Spark のようなビッグ・データ管理および分析ツールに関する知識、データとインサイトを視覚化するツールの経験です。

## データ・エンジニア

データに隠れたパターンの発見を担うのはデータ・サイエンティストですが、では、そのデータはどこで入手するのでしょうか。ソリューションを数千ユーザーの規模に拡大しなければならないときや、機密情報を扱うときには、どうするのでしょうか。データ・サイエンティストがそこまで担うためには、業務を運用可能にする適切なサポートが必要です。

**ここで登場するのが、データ・エンジニアです。**おおまかに言うと、データ・エンジニアはデータ・サイエンティストが最終的に分析を構築するときに使用するデータを収集、編成、クリーニングする役割を果たします。データ・サイエンティストがデータから価値を引き出すとすれば、データ・エンジニアはソースから終着点までデータがスムーズに流れ、処理されるよう努めます。

他のデータ関係者(データ・サイエンティストも含む)が利用し、データ処理時に依存するシステムおよびプロセスをセットアップするのも、データ・エンジニアの仕事です。データ・エンジニアは、データ・フローの遅れを最小限にとどめ、俊敏なアナリティクスを実現するために、その微妙な扱い方を理解しなければなりません。データ・サイエンスのプロジェクトを本番に移行する際には、フロントエンドの開発者とも連携します。

多くの組織では、複数のソースから集まる大量の一時データについて、データ取得と統合パイプラインを設計、構築、測定するなど、データの統合もデータ・エンジニアに任せられます。たとえば、データベースの抽出、アプリケーション・サーバーのログ、スキャンした画像、音声記録、Twitter のストリーム、Web サイト、健康状態

データ・エンジニアは、データ・アナリストまたはデータ・サイエンティストが評価や実験に必要なデータを容易に取り出せるように土台を築く役割も担います。

のセンサー・データなどが該当します。フィルタリングした情報の膨大な「プール」との間で、継続的なパイプラインが確立すれば、該当するデータ・セットをデータ・サイエンティストが引き出して分析します。



## データ・エンジニアの主なスキル

データ・エンジニアは、データベース・ソフトウェアの内部動作を理解している中核的なエンジニアです。データベース・システムのコンパイルおよびインストール、複雑なクエリーの記述と複数のマシンへのスケールアップ、バックアップの管理、災害復旧システムの展開を行います。データベースや大規模データ処理システムなどを開発、構築、テスト、保守するのもデータ・エンジニアです。

優れたデータ・エンジニアは学習を怠らず、ビジネスを推進できる新しいテクノロジーがどれか常に考えています。そのため、深いプログラミングのバックグラウンドを身に付け、MapReduce、Hive、Pigといった Hadoop ベースのテクノロジーにも精通するようになります。データ・エンジニアはたいてい、SQL ベースの技術や NoSQL テクノロジー、データウェアハウスの手法、抽出、変換、ロード (ETL) などのソリューションについても経験が豊富です。

## 開発者

開発者は、データ・サイエンス・チームの中で常に専属のポジションにあるとは限りませんが、不可欠な役割を果たすことは間違いありません。プロジェクトが成功すれば、データ・サイエンスの業務が形を結び、内部では利害関係者、外部ではお客様の需要に応えることとなります。データ・サイエンスをビジネス・プロセスとして成り立たせることこそ、企業がデータを通じて競争上の優位に立つことにほかなりません。

多くの場合、開発者の出番はデータ・サイエンスのワークフローの終盤になります。その役割は、モデルが成立するところにアプリケーション（データ・サイエンス・チームで他のメンバーが引き出したインサイトとデータを活用するアプリケーション）を構築することです。そのためには、かなりのプログラミング時間が必要であり、しかも仕事は困難を極めるものかもしれません。統合されたテクノロジーがなければ、開発者の仕事は行き詰まり、データ・サイエンス・チームの分析を組み込むことはできなくなります。



開発者がデータ・サイエンスに  
使いやすさ (Consumability)  
を持ち込んだおかげで、内外  
の関係者が日々の業務でデー  
タとアナリティクスを活用でき  
るようになりました。

もうひとつの要件として、開発者は十全なプログラミング・スキルをいつでも使える必要があります。これには、Web サービス構築の専門技能、フロントエンドの開発スキル、ユーザー・インターフェースの機能と特徴に関する深い知識が必要です。また、アプリケーション・プログラミング・インターフェース (API) にも精通し、それを利用してさまざまなデータ成果物とソースをアプリケーションに組み込めなければなりません。開発者のスキルが適切に機能すれば、そこから作られるアプリケーションは強固であり、その表面下にデータ成果物が利用されているとは意識もされないほどです。

モバイル、Web、エンタープライズ向けに、市場を大きく変えるほど革新的なアプリケーションを開発するために開発者に必要なのは、そのビジョンを現実に変えるデータ、そしてツールです。開発者はデータ・サイエンティスト、データ・エンジニア、データ・アナリストと協力して、ビジネス目標とアナリティクス・バックエンドとの間で連携を図らねばなりません。

## ビジネス・アナリスト

ビジネス・アナリストは、シチズン・アナリストと呼ばれることもあるように、ビジネスレベルの専門知識と指針を、データ・サイエンス・チームに提供します。その職務は、分野の専門知識を応用して、データに基づく意思決定を下すことです。

ビジネス・アナリストは、企業のニーズを理解しますが、詳細な分析を展開できるような技術的なバックグラウンドは持ち合わせていません。テクノロジーが進歩し、ユーザー・プロファイルの

理解が深まることで、ビジネス・アナリストは「分析の消費者」となり、データの洞察に基づいて意思決定ができるようになります。柔軟で使いやすいテクノロジーがあれば、ビジネス・アナリストはコーディングを必要とせず、またデータ・サイエンティストにも相談することなく、ビジネスレベルの分析を展開できます。そうすれば、ビジネス・アナリストは独立性を保てるため、データ・サイエンス・チームによる結果を日々の分析で利用する機会も増えます。



## データ・サイエンス・チーム: 新たな変革の担い手



熟練のデータ・サイエンティスト、データ・エンジニア、開発者、ビジネス・アナリストは、現代のビジネスを変革する人物です。ビッグ・データ経済の中核とも言えます。それは、単に彼らが新しいシステムを設計するからではありません。データの新しいソースと、そのデータの新しい使い方を探ろうとするからです。もちろん、IT部門も引き続きシステムを構築しなければなりません。各部署が協力して問題を解決し、イノベーションを加速するうえで重要なのは、データ・サイエンスのプロフェッショナルです。

最適なデータ成果物は、エンド・ユーザーには気付かれさえしないものです。これを実現するには、インターフェースや、取得されたデータ、アナリティクス・バックエンドによって有効になる自動化に注目する必要があります。今では、大量のビジネス・データを収集して分析するテクノロジーが利用できるようになり、それを会社の利益のために活用することができます。データ・サイエンスという、ここ一番の大勝負に、チーム・プレイヤーとして参加する準備はできていますか。

### 将来のデータ・サイエンス・プロフェッショナルを育成する

しっかりしたチーム・プレイヤーとして、ぬかりなく備えるには、知識を蓄え、スキルを磨いて自分自身の向上を図る必要があります。業務の水準と速度を改善しようとするデータ・サイエンスのプロフェッショナルを、IBM はサポートします。

成長の著しい分野はどこでも、新しいスキルと専門技能を持つプロフェッショナルを必要としています。IBM は、データ・サイエンス・コミュニ

ティーを一貫して積極的にサポートしてきましたが、今後もオンラインで、また直接的な指導の場でこのサポートを続けていく予定です。イベント、ミーティング、トレーニング・コース、コンテンツ、オープンソース・コミュニティーへの貢献などを通じて、IBM は現職あるいは将来のデータ・プロフェッショナルをサポートし、大容量・高速という今後のデータ需要への対応をお手伝いします。



## 参考資料

データ・サイエンスのスキルを強化する準備はできましたか。今すぐ始めるなら、参考資料はこちらをご覧ください。

1. データ・サイエンスのスキルを磨くなら、[Big Data University](#) へ
2. 必要なツールを使い始めたければ、[IBM Data Science Experience](#) へ



---

© Copyright IBM Corporation 2016

日本アイ・ビー・エム株式会社  
〒103-8510  
東京都中央区日本橋箱崎町 19-21

Produced in Japan  
June 2016

IBM、IBM ロゴおよび [ibm.com](#) は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、[ibm.com/legal/copytrade.shtml](#) をご覧ください。

本資料の情報は最初の発行日の時点で最新であり、予告なしに変更される場合があります。すべてのサービスが IBM の操業国すべてにおいて提供されているとは限りません。

本資料の掲載情報は現存するままの状態を提供され、第三者の権利の不侵害の保証、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任なしで提供されています。IBM 製品は、IBM 所定の契約書の条項に基づき保証されます。



Please Recycle

---