

アナリティクスを すべてのユーザー に解き放つ

SAS®とオープンソースのアナリティクスの
利点の融合



アナリティクスをすべてのユーザーに解放つ

企業や組織のアナリティクス環境をあらゆるタイプのプログラミング言語、あらゆるレベルのユーザーに開放するには、どうすればよいのでしょうか？ また、取り組みが組織内のどこで開始されるかに関係なく、分析モデルとその結果を踏まえた行動の一貫性を確保するには、どうすればよいのでしょうか？

今日のアナリティクス・テクノロジーをもってすれば、オープンソース・アナリティクスと商用のアナリティクスをめぐる議論の焦点は、もはや二者択一ではありません。今では、SASとオープンソース、それぞれのアナリティクス・テクノロジーがもたらす利点を融合することが可能です。

アナリティクス・ライフサイクル全体を考えると、アルゴリズムだけでなく、データの準備、業務への実装、パフォーマンス、スケーラビリティ、ガバナンスについても熟慮することが重要です。このサイクル内では、オープンソース・アナリティクスと商用のアナリティクスの両方が様々な役割を果たすことができます。

例えば、機械学習アルゴリズムをSASまたはPythonで開発し、SAS Event Stream Processingを用いてリアルタイムのデータ・ストリームに適用することができる一方で、それらのアナリティクス資産をJavaやC向けのAPI群、RESTful Webサービス、Apache Kafka、HDFSなどを通じてオープンシステムと連携させることも可能です。

このe-bookでは、様々な業種の組織がオープンソース・アナリティクスとSASソリューションを併用し、両方から最大限の価値を引き出している事例をいくつかご紹介しながら、アナリティクス・ライフサイクル全体においてSASソリューションが果たす役割を説明します。

目次



オープンソース・アナリティクスについて
オープンな姿勢を維持することの重要性

3



モデリングの工業化が
参加無料型ネットゲームの利益拡大を促進

6



オープンソース・アナリティクスに関する5つの疑問

9



アナリティクスの新世界で高まるオープンソースの機運

12



ビジネスのためのオープン化

15



すべてのユーザーのためのアナリティクス、
すべての機能を一元的に管理

18



オープンソース・
アナリティクスについて
オープンな姿勢を
維持することの重要性



どのアナリティクス・ソフトウェアが最良かをめぐる議論をやめ、代わりにすべてを活用することを決断したとしたら、どうなるのでしょうか？

今日のデータ・サイエンティストたちは、多種多様なバックグラウンドの持ち主の集まりであり、幅広いスキルを現場にもたらしめます。彼らが多様なアナリティクス・ツールに加え、モデルの統制や業務への実装を一貫した方法で行うシステムを利用できれば、複雑な課題を解決するための選択肢が増えます。

これを実現して成果を上げている組織の例として、本稿では Cleveland Clinic と Cox Automotive を取り上げます。この取り組みの結果、両社のデータ・サイエンス施策は好調に進んでおり、組織全体の業績も向上しています。

「当院には複数の言語やテクノロジーについてトレーニングを受けた従業員が何人もいます。それぞれが最も快適に感じる言語を使いながらも、共通のアプローチで仕事を進められるようにしたいと考えています」と話すのは、Cleveland Clinic のアナリティクス担当エグゼクティブ・ディレクター、クリス・ドノバン (Chris Donovan) 氏です。

同氏によると、Cleveland Clinic の目標は医療システム全体でアナリティクスの成熟度を高めることであり、今現在は、アナリティクスのスキルと機能を1つのチームに集中させるのではなく、組織全体を対象とした広範囲な仕組みの構築を進めています。「それを可能にするプラットフォームを整備することが、アナリティクスの成功には極めて重要です」(同氏)。

アナリティクスに対するこうした注力は、Cleveland Clinic だけでなく、医療業界においても、緩和ケアから予防ケアへと重点を移すための変革に役立ってきました。ドノバン氏は次のように説明します。「個々の患者が ER (救急外来) や医院を訪れたときにケアするだけの医療から脱却し、患者全体を視野に入れ、そもそも病気になるのを防ぐにはどうすればよいかを考える医療へとシフトすることは、どうすれば可能になるのでしょうか？」

「プログラミング・コードを問わない」環境の実現

Cox Automotive 社の意思決定科学担当副社長、ショーン・ハッシュマン (Shawn Hushman) 氏によると、Cox Automotive では、SAS® Viya® がアナリティクス部門全体を結びつける接着剤として機能しています。「ソースシステムをめぐる政治的な議論が不要になるため、モデリング自体、モデルのバージョン管理、モデルの業務への実装に集中することができます。このオープンなプラッ

トフォームでは、誰もが好みのオープンソース・コードを活用できるようになり、また、新しいコードベース (SAS Viya の機能) が全員に開放されます。

ハッシュマン氏が特に称賛しているのは、異種混在のコード、プロセス、情報を統合することで、首尾一貫した情報提供を行う1つのハブを実現できる点です。

同氏のチームは、Python、R、SAS でコードを書く開発者を世界各地に抱えています。「複数のインターフェイスでプログラミングを行う開発者がおり、彼らはモデル開発のコラボレーションにも様々な方法を利用しています。また、結果の発行や表示に用いる方法の好みや、モデルを業務へ実装する方法も様々です。SAS Viya なら、そのすべてを取りまとめることができます。」

このソリューションのおかげで Cox Automotive は「コードを問わない」環境を実現できている、とハッシュマン氏は言います。好みのコードベースについて論争するのではなく、誰もが課題に対する最良の解決策と一緒に議論できるのです。「相手のコーディング方法の好みを気にすることはありません。モデリングにどのパッケージを使うかで争うのではなく、組織に真の変革をもたらすフレームワークに目を向けましょう、ということです。」



ショーン・ハッシュマン
(Shawn Hushman) 氏

Cox Automotive 社、
意思決定科学担当副社長

「SAS Viya ではデータ・サイエンティストのコミュニティの力を結集させることができますが、これは他のソリューションでは不可能なことです。」

そうした考え方により、Cox Automotiveでは、AutotraderやKelley Blue Bookといったオフライン事業をオンラインの世界へ移行する際にデータを最大限に活用することができています。「SAS Viyaではデータ・サイエンティストのコミュニティの力を結集させることができますが、これは他のソリューションでは不可能なことです。我々はモデル管理をハブの中心だと考えていますが、それは、特定の何かに依存することなく、すべてを確実に結びつけることができるからです」(同氏)。

「我々の責任は、結果を効率的に提供することと、そのプロセスをシームレスにすることです。我々が求めているのは、アルゴリズムの多様性に由来する良い面をすべて活用した上で、提供内容や提供方法をめぐる整合性・連携性を確保することです」(同氏)。

アナリティクスを経営幹部にも開放

Cleveland Clinicでは、より多くの人々がアナリティクスを利用できるようにすることも優先課題となっています。ドノバン氏をはじめとするCleveland Clinicの経営陣は、コードの記述や高度なアナリティクスの実行に精通しているデータ・サイエンティストやプログラマーだけでなく、経営幹部やマネージャーもドラッグ&ドロップ操作と簡潔なインターフェイスを用いてデータを容易に利用できるようにしたいと考えています。

「経営幹部は予測モデルの作成方法を知らなくてもよいかもしれませんが、より優れた意思決定のためにデータを活用できる必要はあります。全員がデータ・サイエンティストである必要はありませんが、自身の職務に求められるレベルでデータを活用できるスキルは、誰にとっても必要です」(ドノバン氏)。

ドノバン氏によると、Cleveland Clinicでは今、「分析担当者とは何か？」を再定義することと、あらゆるレベルのユーザーに共通の入口を提供することに取り組んでいます。「以前は、至るところにデータがあり、複数のツールが使われていましたが、今はその状況の転換に取り組んでいます。データを人々のところに持っていくのではなく、人々をデータのあるところに連れていくのです。世界最高水準のプラットフォームを確立すれば人々を引き込むことができ、それによって一貫性が促進され、プラクティスの共有コミュニティが形成され、組織全体にわたって人々の連携が実現し、最終的に全組織レベルの標準アプローチを見つけ出すことができる、我々はそう確信しています」(同氏)。

アナリティクスによるビジネス戦略の開拓

Cox Automotiveのハッシュマン氏も、組織全体でアナリティクスを活用することの重要性を強調しています。「アナリティクスとは組織の心臓および魂であり、あらゆる業務の基盤要素であるという見方を私は好みます。弊社ではアナリティクスが、すべての製品の改善、新製品の開発促進、すべての製品スイートにおける収益拡大を実現しています。アナリティクスが弊社のビジネスを推進しているのではなく、アナリティクスこそが弊社のビジネスなのです。」

あらゆるユーザーをサポートできるようにアナリティクスの活用範囲を拡大することは、テクノロジー戦術のレベルを超えた取り組みであり、ビジネス戦略の一部です。「アナリティクスはコア戦略をサポートする能力の1つにすぎないではありません。それ自体がコア戦略の1つとなるべき取り組みです。組織全体でアナリティクス成熟度を高め、その領域で世界最高水準を実現する必要があるものであり、それができなければ他社に先を越されてしまうでしょう。」



クリス・ドノバン

(Chris Donovan) 氏

Cleveland Clinic、アナリティクス担当
エグゼクティブ・ディレクター

「アナリティクスはコア戦略をサポートする能力の1つにすぎないではありません。それ自体がコア戦略の1つとなるべき取り組みです」



モデリングの工業化が 参加無料型ネットゲーム の利益拡大を促進

Wargaming 社、1日に数テラバイトの
顧客データを解釈できる規模に
アナリティクス環境をスケールアップ

連日、数百万の人々が「World of Tanks」に参加し、仮想の戦場で戦車同士の対戦を楽しんでいます。プレイヤーの多くは全く料金を支払っていません。それでも、このゲームの開発元であるWargaming社は、ゲーム業界で1年に数十億ドルの収益を上げています。

このゲーム業界の巨人は、どのようにして参加無料型のゲーム事業を採算化し、全く料金を支払わないプレイヤーも含め、あらゆる顧客の“プライスレス”な価値を把握できるようになったのでしょうか？ その実現を支えている要因の1つはアナリティクスの活用です。

Wargaming社の目標は、あらゆるレベルのプレイヤーに優れたエクスペリエンスを提供することです。MacでもWindowsでも、コンソール版でもモバイル版でも、誰もが無料で参加することができ、もっと深く楽しみたいプレイヤーはゲーム内でアイテムを購入することもできます。

Wargaming社のゲームの中でも最多のプレイ回数を誇る「World of Tanks」には全世界で1億1,000万人のオンライン・ユーザーが登録しており、同社はすべてのユーザーのすべてのプレイについて、発射されたすべての砲弾、戦車がとったすべての動きに関するデータを収集しています。Wargaming社のビジネス・インテリジェンス・データサービス部門責任者であるアレクサンダー・リャーボフ (Alexander Ryabov) 氏は、次のように説明します。「毎日どの瞬間でも、約400万人が弊社のゲームをプレイしています。プレイヤーは複数のバトルをプレイし、各バトルでは複数のイベントが発生するため、そこで生成されるデータの総量は1日で数テラバイトに及びます。」

Wargaming社では、プレイヤーがゲームにログインした瞬間からログアウトするまでの間、データを収集し続けます。また、ゲーム内で交わされるチャットのログのほか、ソーシャルメディアやゲーム愛好者コミュニティで自社のゲームについて語られている内容も、収集・分析しています。こうしたデータを活用することで、顧客の維持、他のゲームのクロスセル、無料プレイヤーから課金ユーザーへのコンバージョン、「プレイヤー・ジャーニー」のモニタリング、ゲーム内で不便な点の軽減などを推進するためのモデルを実行できます。

総計すると、Wargaming社は毎月30テラバイト以上のデータを処理しています。データの98%はOracle Big Data Appliance上のHadoopに保管されており、Hadoop環境の管理にはClouderaを使用しています。データがHadoopに入った後は、SASのソリューションと連携するデータマートをETL開発者が作成し、この統合環境でモデルの作成と本番環境への展開が行われます。

ゲームプレイ体験と顧客向けオファーをアナリティクスで改善

Wargaming社ではデータ・サイエンティスト・チームがモデルを開発し、各モデルのスコアリング結果は、ゲーム内のイベント処理コンポーネントや会社のCRMシステムに送信されます。そして、そこから開発チームにフィードバックが戻され、開発チームはそれを活かしてさらなるモデリングを行います。

例えば、開発チームは最近、ゲーム内の特定の場所でプレイヤーが次々に「死ぬ」ことをデータから認識しました。「そこで、マップのバランスを調整するために、その場所に丘陵を作りました。弊社のデータ・サイエンティストはヒートマップを作成しており、特定の期間内に放たれたすべての砲撃をゲームマップ上で確認できるようになっています」(同氏)。

また開発チームは、プレイヤーがゲームの特定の要素を見逃していないかどうかを調べ、必要に応じて通知メッセージを送ることで、プレイヤーの次回のエクスペリエンスを向上させる目的にも、アナリティクスを活用しています。例えば、何らかの武器を入手できる場所や、前回のゲームで見落としていた場所をプレイヤーに教えることができます。

「こうしたメッセージは、プレイヤーが次にゲームをプレイしたときのエクスペリエンスを向上させるために役立ちます。これはほんの一例であり、適切なモデルを作成して本番環境に展開することで、同様のことを数多く実現することができます」(同氏)。



アレクサンダー・リャーボフ
(Alexander Ryabov) 氏

Wargaming社、ビジネス・インテリジェンス・データサービス部門責任者

「そして、総合的な満足度を高め、さらに優れたプレイヤー・エクスペリエンスを実現する取り組みには、SASが役立っています」

Wargaming 社ではカスタマー・エクスペリエンスのさらなる向上を目指し、ソーシャルメディアや顧客との会話から収集したフィードバックに対してテキスト・アナリティクスを適用しています。「ソーシャルメディア・データに特定のフィルターをかけることで、プレイ体験全般に関するセンチメント分析を実行できるようになります。センチメント分析は、顧客サポートに関して利用することも、チャネル別に“オールスター・プレイヤー”を識別するために利用することもできます」(同氏)。

アナリティクスのキャパシティを MMO (massively multiplayer online) ゲームに適した規模に拡張

Wargaming 社は3年前にビジネス・インテリジェンス・プログラムを構築しましたが、それを境にオープンソース・テクノロジーに引き寄せられていきました。「詳細なデータ分析とデータマイニングが必要だと分かった後は、初期の“高度なアナリティクスによるモデリング”を開始し、R、Spark、Python、その他のあらゆるオープンソース・ソリューションを試していきました」(同氏)。

しかし、やがてチームは、そうした初期の取り組みを、数千種類のモデルや日々増え続けるデータ量に合わせてスケールアップすることは、極めて難しい取り組みであることに気付きました。同社におけるオープンソース・アナリティクス活用の初期段階について、リャーボフ氏は次のように振り返ります。「弊社にとっての最大の課題はスケーラビリティでした。データ・サイエンティストはモデルのコンセプトを思いつくと、データ抽出やデータ加工を行います。都度同じ作業を繰り返すのではなく、結果を得るまでの工程を自動化する必要が出てきます。当時はすべてを手作業で行っており、それが開発者の大きな負担となっていたのです。」

同氏によると、チームが開発した初期のモデル群は業務への実装までに3~6ヶ月ほどかかっていました。「すべてのゲーム、すべての地域、すべての時間枠に対応するためには、数百、さらには数千のモデルを実行する必要があることに気付いた後は、そのレベルのスケーラビリティを実現できるソリューションを探し始めました。」

詳細なリサーチを経て、リャーボフ氏とチームはニーズを満たすソリューションを見つけました。「弊社の用途には SAS Factory Miner と SAS Model Manager がパーフェクトでした。なぜなら、

1つのモデルを時間枠別、地域別、製品別に何度も再利用できるからです。各モデルは実質的に同じモデルですが、ビジネス・コンテキストに合わせて本番環境に展開し、一種の工業化された方法でモデルの実行、維持管理、改良を繰り返すことができます。私たちのリサーチでは、唯一の有効な選択肢が SAS でした」(同氏)。

リャーボフ氏によると、現在では、データの準備とモデリング方法論の確立が済んでいれば、同一モデルから数千種類の派生モデルを作成する作業は担当者が1人で行えます。「これほど多くのモデルを手作業で作成・維持管理するには、10人から20人が必要になり、当然、人為的なミスも発生するでしょう。SAS を用いて自動化された本番環境はミスをすることがありません。」

全体的に見ると、Wargaming 社のデータ・サイエンティストは、より多くのモデルをより短時間で作成・展開できるようになっており、それがひいては、収益の増大、リソース利用率の向上、機会創出費用の削減につながると期待されます。今後、市場が成長し、Wargaming 社が他のプラットフォームへの多角化展開を続けていく中でも、既にモデリングの自動化と工業化に成功している同社は、より多くのモデルの実行、より多くの顧客の維持、より多くの顧客の獲得、より複雑なアナリティクスの適用といった取り組みのすべてを、同一のアナリティクス・プラットフォーム内で遂行していくことができるでしょう。

最も重要な点は、プレイヤーにもメリットがあることです。「弊社のデータ・サイエンティスト・チームは才能あふれる人材の集団であり、どうすればプレイヤーの希望通りのものを適切なタイミングで提供できるかについて、実に革新的なアイデアを持っています。そして、総合的な満足度を高め、さらに優れたプレイヤー・エクスペリエンスを実現する取り組みには、SAS が役立っています」(同氏)。

また、ゲームのプレイ体験を改善する取り組みには、お金をかけてゲームをもっと楽しみたいと考える長期プレイヤーを増やす効果も期待できます。「当社の創業者であるビクトル・キースリュイ (Victor Kislyi) が言うように、『弊社の目標は究極的にはプレイヤーにハッピーになってもらうこと』です。プレイヤーがハッピーなら、他のことはすべて何とかなるものです」(同氏)。

モデリングを工業化するメリット

モデリングの自動化と工業化は、Wargaming 社に多くのメリットをもたらしています。

- コーディング作業の大部分がポイント&クリック操作中心のワークフローに移行したことで、モデル作成の作業効率が向上
- モデルの開発と展開にかかる時間が60%短縮
- モデルの展開や自動化のためにデータウェアハウスを管理する必要性が80%軽減



オープンソース・ アナリティクスに関する 5つの疑問

アナリティクス施策に
オープン性を積極的に取り込んでいる
Wyndham Destination Network社

380万人の会員に200万件の物件情報（個別のSKU情報）を提供するバケーション向けレンタル物件予約サイトにおいて価格設定を最適化するためには、正確で信頼性が高くノンストップで稼動し続けるアナリティクス環境が必要です。

Wyndham Destination Network 社では、価格弾力性やリゾート運営、顧客維持といった複雑な課題について革新的な解決策を見出すために、アナリティクス・チームがオープンソース・アナリティクスと全社規模のアナリティクスを組み合わせて活用しています。

そして、Wyndham Destination Network 社のアナリティクス担当上級副社長であるジェレミー・ターブッシュ (Jeremy TerBush) 氏から、同氏のチームにおけるアナリティクス（オープンソース・テクノロジーから全社規模のソリューションまで）の利用状況についてお話を伺いました。

オープンソースとオープン・ソフトウェアをそれぞれどのように定義していますか？

ジェレミー・ターブッシュ氏：私にとってオープンソースとは、システムをその場でダウンロードし、一切のライセンス取得手続きなしで自分のマシンにプログラムを直接インストールできることを意味します。オープンソースは無償であり、使いやすく、そのソフトウェアについて議論したり改善に取り組んだりする数多くのコミュニティが存在します。

一方、オープン・ソフトウェアとは、他のソフトウェアとの接続が容易なソフトウェアを意味します。弊社には予測を生成するシステムが複数あり、それらのシステムは予測結果を別のルーチンに送信する必要があります。オープン・ソフトウェアを利用すると、開発や実行の観点からシームレスかつ容易にそれを実行できます。

Wyndham 社ではオープン性にどのように取り組んでいるのでしょうか？

ターブッシュ氏：全員が同じツールを使うように義務付けてはいません。新しい課題に突き当たり、自社で構築し実装も済んでいるシステムでは対応できないときに、分析担当者がオープンツールに頼った例は何度も目にしています。弊社には、小規模なプロジェクトの開発や課題解決をオープンソース・ソフトウェアで成功させた例が多数あります。今では、そうした成果を現場業務に組み込む必要がある場合には、より大規模な環境にそれを移行して、SASを活用して実装するのが典型的なやり方です。

課題を割り当てられたグループは希望のツールを使用することが許されているため、弊社の環境は一種のオープン・エコシステムと言えます。オープンソースは一回限りの課題のプロトタイプ段階で使われています。オープンソース・テクノロジーを活用した成功例が数多く見られるのは、この領域です。

オープンソース・テクノロジーから全社規模のソフトウェアに切り替えるのは、どのタイミングでしょうか？

ターブッシュ氏：我々がセットアップしているプロセスのうち日々実行する必要があるのは、典型的には、レンタル物件の需要予測や価格弾力性のような価格設定に関するプロセスです。それを行う



ジェレミー・ターブッシュ
(Jeremy TerBush) 氏

Wyndham Destination Network 社、
アナリティクス担当上級副社長

「ですから、最初のプロトタイプではオープンソース・アナリティクスを用いて予測を実行します。その第一段階の結果に満足できたら、それをSASで実行するのです。弊社の場合は、SASを利用する方が、競争力のある商品を容易に作成できます」

には、最近の履歴や取引行動のデータをすべて取りまとめ、それを予測ルーチンに入力することで、最終的に当該物件の日次価格をレコメンドする必要があります。当初のモデルには、オープンソースのアナリティクス・ソフトウェアを用いてプロトタイプ化した最適化や予測が含まれています。その結果に満足できたら、そのアルゴリズムをSASに移行した上で、日々の実行スケジュールに組み込みます。

価格設定の予測のために日次ベースで実行する必要があるジョブの場合、オープンソース・ツールによる実装が、日次ベースの価格最適化に求められる正確性を実現できるかどうかについて、私には確証がありません。ですから、最初のプロトタイプではオープンソース・アナリティクスを用いて予測を実行します。その第一段階の結果に満足できたら、それをSASで実行するのです。弊社の場合は、SASを利用する方が、競争力のある商品を容易に作成できます。

両方を活用する利点は何でしょうか？

ターブッシュ氏: SASとオープンソース・ソフトウェアの両方を活用する主な利点の1つは、各自が最も快適に感じるプラットフォームを使えることです。人々が発想を完全に転換するためには、それ相応の学習曲線を辿る必要があるため、誰もが自分にとって快適なプラットフォームを利用できる環境は、チーム全体の生産性が大きく向上することを意味します。私は、解決策の質も高まると考えています。チームを挙げてベストプラクティスを収集し、すべてのプラットフォームのコミュニティから知見を引き出し、より大きなアイデア・プールを形成できるため、チーム内部の柔軟性が高まり、多種多様な方法で課題を解決できるようになるのです。

オープン性を維持するためには、他にどのようなテクノロジーが重要でしょうか？

ターブッシュ氏: 我々の組織内でも、他の組織の間でも、パブリックAPIを開発し、それを介してデータを共有することは、以前と比べ遥かに容易になりつつあります。例えば、我々が最近行った作業では、Google MapsのAPIを利用する必要があったので、すぐにPythonの使用を決めました。

企業や組織が活用できる外部データはますます増えています。ソリューションを社内で開発する場合は、モデルを複数のプラットフォームに展開した上で、モデルの実行結果は、それを利用する必要があるプラットフォームにAPI経由で送信するようにしたいと思います。これこそが間違いなく、あるべき方向性です。目標は、どのプラットフォームでも利用可能な結果を生成できるようになることであり、それを実現する手段がAPIなのです。





アナリティクスの 新世界で高まる オープンソースの機運

Quantum Insights 社による調査レポートの
抜粋

企業戦略に関する意思決定において業務部門／事業部門の果たす役割が強まり、また、コア開発プロジェクトの進行速度が従来よりも格段に高まっている中、アナリティクスにオープンソースを取り込むことが意味を持ち始めています。大規模なグローバル製造企業のイノベーション担当CTOが語る言葉の意味をよく考えてみましょう。

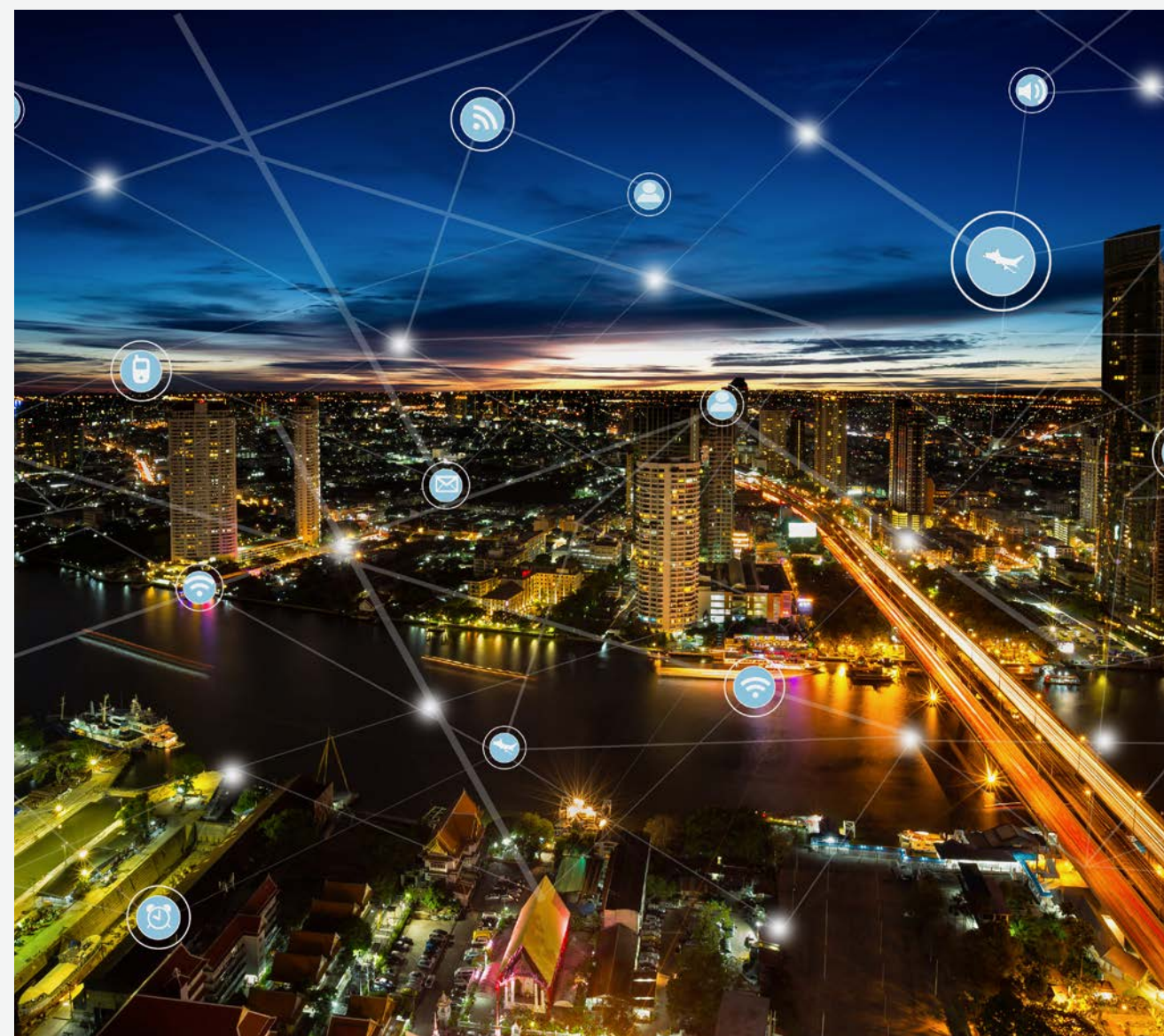
「オープンソースを活用する理由としては、もちろんコストの問題があります。これは特にデータ・アナリティクスの場合に顕著です。この領域では担当者がビジネスケースに確信が持てないケースも多く、彼らは多大な時間と資金を投じる前に様々なアイデアを試す必要があると感じます。投資を正当化できるかどうかは必ずしも明確ではないケースでは、オープンソースの導入障壁がかなり低くなります。つまり、リスクの低さとスピードの向上を両立できる点が魅力となるのです。」

同様に、ある上級IT管理チームが「組織内外での使用を見据えた統合型のサービス管理・提供のための強固なプラットフォームを18ヶ月で実現せよ。この工期には、開発から本稼働開始までの全工程が含まれるものとする」という任務を課された事例も見られています。

こうした状況は、オープンソース・テクノロジーへの大規模な移行を促す直接的な要因となっています。極めて短期間で結果を提供する必要性は、「綿密に準備し、一切の失敗の可能性を排除する」のではなく、「低コストのフェイル・ファスト（失敗を早期に認識して何度でもやり直すこと）を通じて前進する」というアプローチを採用しやすくなることを意味します。

新たなリスク方程式が、新たな業務遂行手法とともに発展しつつあります。[ITを活用した]業務遂行における権限移譲は今や当たり前になっており、このアプローチでは、個々の部署やチームが各々の目標に適合する特定のプロジェクトの実現に取り組む一方で、全社レベルのIT部門は、そうした取り組みのすべてが1つのエコシステムに収まっている状態を確保することに責任を負います。

このアプローチが今、多くの企業で「社内文化の革命」を引き起こしており、内部の境界をまたぐ連携だけでなく、多くの場合は小規模な専門性の高いパートナー企業との連携、さらには外部開発者のコミュニティとの連携を通じて、以前よりも格段に高いレベルの協働態勢が実現しています。



以上から導き出される結論とは？

オープンソースの利用拡大が進むにつれ、企業のIT環境は、より複雑になっていきます。その結果、場合によっては制御が困難になるほど大規模化および多様化する恐れのあるソフトウェア・スタックを適切に管理するために、極めて強力なプロセスを整備する必要性が生じます。

CIOは、大規模なIT環境の管理を「当然の職務」として受け入れ、それを粛々と実現する必要があります。ただし、オープンソースのコスト低減効果を通じて実現される明白な節約額の大部分は、多様性の増大に適切に対応できる強固かつセキュアな統合型のソフトウェア・スタックを構築する上で生じる「追加のコンサルティング・コスト」や「内部IT支出の上昇」によって相殺される可能性があります。

したがって、コスト計算は、オープンソース・ソフトウェアの利用を想定してプロジェクトを依頼するビジネスユーザーには単純に見えるかもしれませんが、CIOの立場からは全く異なる様相が見えてきます。

アナリティクスの新世界に向けた推奨事項

調査結果を踏まえると、論理的な答えは「モジュール方式を採用しており、定常的な更新を通じて常に業界のベストプラクティスが反映されるようになっている、特定ソフトウェアに依存しない業界標準のアナリティクス・プラットフォームを整備すること」だと思われます。

この方針に従えば、大部分の機能を導入後すぐに利用できる標準的な市販パッケージでも、保守リスクの抑制、投資コストの低減、運用面での高度な柔軟性を実現できます。

また、大規模な企業のビジネスユーザーも、より広大なIT環境の完全性をリスクに晒すことなく、オープンソースのアナリティクス・ソフトウェアを自由に利用できるようになります。

このような専門特化したアナリティクス管理システムは、どのような特徴を備えることになるのでしょうか？

- まず、今日的な「全社規模のIT環境」のすべてに共通して見られる汎用目的の統合レイヤーとは別物です。

- 代わりに、アナリティクス・ワークロードに専門特化した運用環境でなければなりません。
- アナリティクス資産と開発ツールを管理する目的のために設計および最適化されます。
- これにより、幅広いオープンソース製品を対象として、既存のアプリケーション資産群も含める形で、セキュアなプロフェッショナル・サービス管理を適用することが可能になります。
- この環境には、適切に管理されたオープンソース・スタックの開発、合理性に裏打ちされた頑健な評価手法、統一の戦略ロードマップ、採用／開発／利用に関するガバナンス手法が含まれることとなります。

今日ではビジネス管理のあらゆる側面がアナリティクスに依存するようになっているため、アナリティクス・ツールの業者選定、調達、実装、管理、継続的進化について近い将来になされる意思決定は、極めて重要な意味を持つこととなります。

これからの「全社規模のIT管理」に求められているのは、業務部門／事業部門が要求するスピードと局所的自律性を提供しながら、それと同時に、企業戦略に沿った意思決定を実行したり、全社規模の首尾一貫したデータフローのポテンシャルを最大化したりするための統合性も維持することです。

この記事は、SASの委託に基づき作成されたQuantum Insights社の独立調査レポート「Evolution or Disruption: The New World of Analytics? An Informed View From Inside Enterprise Business」(進化か破壊か：アナリティクスの新世界とは？ エンタープライズ・ビジネスの内部から得られた情報に基づく見解)からの抜粋です。

今日ではビジネス管理のあらゆる側面がアナリティクスに依存するようになっているため、アナリティクス・ツールの業者選定、調達、実装、管理、継続的進化について近い将来になされる意思決定は、極めて重要な意味を持つこととなります。



ビジネスのための オープン化

Navistar社、データ・サイエンティストの業務
環境におけるオープンソース・ソフトウェアと
エンタープライズ・ソフトウェアの共存につい
て議論

データ・サイエンスやデータ・サイエンティストをめぐる論争や議論は活況を呈していますが、多くの企業は行動を起こすことを選択し、多数のデータ・サイエンティストの採用・雇用に踏み切っていることから、有望な候補者の供給が需要に追いつかない状況が生まれています。

その結果、「データ・サイエンティストの獲得と維持を効果的に行う」という概念は、ITやビジネスのコミュニティにおける話題として一般的になっています。これほどの売り手市場において、統計学者とビジネス分析担当者のスキルを併せ持つ引手あまたの人材に対し、企業は自社の魅力をどのように伝えればよいのでしょうか？

Navistar International Corporationのデータ・サイエンス担当ディレクターであるジャイアシ・ダパー (Gyasi Dapaa) 氏によると、データ・サイエンティストを獲得および維持するためのカギは、サポートと創造性です。手元の分析課題に最適と感じられるツール群だけでなく、チームが協働で絵を描くことのできる「クリエイティブなパレット」も提供することには、強力なモチベーション向上効果があります。

「技術的バックグラウンドやスキルレベルが様々に異なる20人のスタッフの力を結集させるためには、全員の心を魅了する何かを備えた環境を提供しなければなりません。入社してくるスタッフの中には、高度なプログラムをゼロから書ける者もいれば、既存のプログラムやモデルを再利用して個々のプロジェクト向けに調整することが得意な者もいます。また、プログラムを全く書けない者もいますから、彼らのために、使いやすいGUIを備えたツールを用意する必要があります」(ダパー氏)。

Navistar社は、商用トラック、バス、軍用車両、エンジンなどの世界的なメーカーです。同社のデータ・サイエンティストは、保守作業の優先順位を設定したり、突然の故障を予防したりするために、顧客のトラックから収集されるセンサーデータ(エンジンの稼働状況や故障コードなどが含まれます)を分析します。また、マーケティング・モデルや価格弾力性モデルの開発や、設備機器の修理時期の予測なども行います。

適切な目的のための適切なソフトウェア

ダパー氏によると、彼のチームはNavistar社におけるすべてのアナリティクス・プロジェクトをサポートするCoE (Center of Excellence) として、会社のアナリティクス・エンジンを動かし

続けるために「SASのパワーを徹底的に活用」していますが、それと同時に、チームが幅広いオープンソース・ソフトウェアも利用できるようにしています。その理由を同氏はこう説明します。「オープンソースは無償ですから、コスト面を気にせずに利用できます。また実際問題、データ・サイエンティストに幅広いツールを提供することは重要であり、これは画家が幅広い色や道具を使うのと同じことです。ツールの選択に関して一定レベルの自由度と柔軟性を認めることは、チームのモチベーションを高め、スタッフの意欲を維持するための強力な手段です。」

ただし、ダパー氏はオープンソースのみのアプローチには実践面での制約があると考えています。「すべてのオープンソース・ソリューションが、SASと同じように大規模なデータセットを処理できるとは思いません。私自身は、データのクエリー、統合、クレンジング、プロファイリング、さらには、GLM (一般化線形モデル) 手法やクラスタリング・アルゴリズムを用いた分析を実行するための



ジャイアシ・ダパー
(Gyasi Dapaa) 氏
Navistar社、データ・サイエンス担当
ディレクター

「今では、SASとPythonのスキルを持つデータ・サイエンティストのチームがプロジェクトの様々な部分を分担し、各自の好みのインターフェイスで作業することが可能になっており、互いに作業結果を共有するのも容易です」

ワンストップ・ショップとして、SASを活用できるようにしたいと考えています。SASがなければ、そのすべてを実行するために2~3種類のツールが必要になるでしょう。」

また同氏は、Navistar社ほど大規模なオペレーションの基盤として、各種プロセスに品質管理機能がほとんど組み込まれていないオープンソース・プログラムを利用することには、スピードや正確性に関する懸念があると指摘した上で、「弊社が行う分析は、SASを用いることで高速性と信頼性が向上するのです」と付け加えました。

SAS Viyaの革新性について質問すると、ダパー氏は次のように語りました。

「SAS Viyaでは、スキルセットの異なる分析担当者たちが、よりの確かつ迅速にコラボレーションできるため、SASの魅力が更に強化されます。例えば今では、SASとPythonのスキルを持つデー

タ・サイエンティストのチームがプロジェクトの様々な部分を分担し、各自の好みのインターフェイスで作業することが可能になっており、互いに作業結果を共有するのも容易です。SAS Viyaでは、SASを使用する分析担当者が、Pythonで作業された部分を簡単にSASに取り込めます。こうした利便性は、ほとんどのデータ・サイエンティストが高く評価するでしょう。」

ダパー氏によると、Navistar社の場合、SASから得られた結果を現場業務に組み込むことで信頼性の高い洞察やビジネス価値を引き出せるため、SASへの投資効果は指数関数的に増大しています。ただし、最終的に重要なのは「方法」ではなく「理由」であると同氏は指摘します。「ほとんどのデータ・サイエンティストにとってモチベーションを高める要因は、数学や統計のエlegantさではなく、解決策がビジネス戦略の推進にどう役立つかという点です。自分の仕事がビジネスにどのような違いをもたらすのかを理解できるように彼らを支援する必要があります。」





すべてのユーザーの
ためのアナリティクス、
すべての機能を一元的
に管理

ベスト・オブ・ブリード（最良の組み合わせ）型
のアナリティクス環境を単一のアナリティクス・
プラットフォームで統制

\$45.00
\$52.00
\$23.00
\$20.00
\$30.00
\$10.00
\$7.00
\$9.00

ここでは、ある成長中の国法銀行（訳注：米国連邦政府の監督下にある商業銀行のこと）に勤務する聡明でアナリティクスに精通した従業員をご紹介します。いずれも所属部署でアナリティクスを駆使し、ポジティブな成果を上げています。



ジェームズ

ジェームズは事業開発グループのデータサイエンティストです。同行の上得意客向けにカスタマイズしたオファーを作成するためにデータを分析しています。



スーザン

スーザンはデジタル・マーケティング担当者です。住宅ローンの見込客のカスタマージャーニーを追跡し、要所要所の意思決定に影響を及ぼすことが仕事です。



クリスティーナ

クリスティーナはリスク分析担当者です。同行の融資ポートフォリオに関するリスクモデルを構築しています。



マービン

マービンはリテール・バンキングの業績がトップの支店に所属する市民データサイエンティストです。地元の顧客に関するデータを自発的に分析しています。

同行では、彼らのように進取の気性に富んだ従業員たちが、アナリティクスやデータ管理のために十数種類の異なるパッケージを利用しています。商用のデータ統合ツールから、SAS、Python、Rまで、オープンソースのアプリケーションやクラウドベースのアプリケーションもあれば、ホスティング方式の全社規模のアプリケーションもあります。

この銀行のIT部門が「これらのプロジェクトのすべてにおいて、信頼性の高いデータ、正確なモデル、厳格な分析プロセスが使用されており、それによって結果の有用性が保証されている状態」を確保するためには、何をどうすればよいのでしょうか？ また、カスタマー・ジャーニーの追跡やその他のアナリティクス活用機会の発掘に利用できるように、こうした異種混在のコードベースやビジネスシナリオのすべてを取りまとめるのは、誰の責任なのでしょうか？

アナリティクスのための単一のエコシステム

この銀行が必要としているのは、様々なデータソースを組み合わせることで分析用データを作成する機能から、トラッキングやモニタリングのためのすべてのアナリティクス資産を管理する機能、モデルへの適切なアクセスをすべての部門に提供する機能までを完備した、一元管理型のアナリティクス環境です。しかし、単一のプラットフォームだけを用いて、すべてのアナリティクス活動を取りまとめ、あらゆるレベルのユーザーに容易なアクセスを提供し、さらにはオープンソース・プロジェクトの統制も行うことなど、本当に可能なのでしょうか？

SASのアナリティクス・プロダクト・マーケティング・マネージャーであるフィオナ・マクニール (Fiona McNeill) は、次のように説明します。「アナリティクスについても、それぞれの部門内で既にうまく機能している“影のIT部門”に任せればよいと考えるかも知れません。しかし、アナリティクス・ガバナンスのためのプラットフォームを整備すれば、気の利いた分析を部門別に行うだけの状態から脱却し、真のビジネス価値を全社規模で実現できる態勢に移行できます。」

マクニールが推奨するのは、一元管理型で包括性、柔軟性、弾力性に優れた単一プラットフォームです。「必要なのは、IT要件の変更を強いるようなプラットフォームではなく、組織のIT要件に順応できるプラットフォームです。」

アナリティクス・ガバナンス・プラットフォームの仕組み

アナリティクス・ガバナンス・プラットフォームとは、アナリティクス・エコシステム内のすべての資産を一元的に管理・活用することを可能にするソフトウェアのことです。ユーザーには見えないところで機能し、データ・サイエンティストによるモデルの管理や、IT部門によるアナリティクス・コンプ

ライアンスの監査を支援します。また、モデルのバージョン管理、監査証跡、モデルのリネージ、ソースデータに関する情報を提供します。

では、こうしたアナリティクス・ガバナンス・プラットフォームは、どのように機能するのでしょうか？ まず、データ・サイエンティストはモデルをインベントリに登録し、ビジネス分析担当者はモデル内の各オプションの適用条件を記述したビジネスルールを適用します。すべての事項が、実業務環境への展開のために文書化およびテストされます。その結果、IT部門は、開発されるすべてのモデルについて、それぞれに関連するビジネス・コンテキストを把握できるようになります。このシステムでは、インベントリ、パフォーマンス、展開プロセスのモニタリングとその結果を踏まえた継続的なモデル改善を通じて、結果の正確性が維持されます。

分析担当者の観点からすると、アナリティクス・ガバナンス・プラットフォームは、各種のアナリティクス手法を一元管理できるインベントリと、あらゆる展開アクティビティ／アクションに取り込める標準化されたコードベースを提供してくれます。コードの可搬性が高いため、いったん記述したコードはどこでも実行可能であり、どのような規模の課題の解決にも対応できます。また、モデルが



フィオナ・マクニール
(Fiona McNeill)

SAS、アナリティクス・プロダクト・マーケティング・マネージャー

「アナリティクス・ガバナンスのためのプラットフォームを整備すれば、気の利いた分析を部門別に行うだけの状態から脱却し、真のビジネス価値を全社規模で実現できる態勢に移行できます」

チャンピオンか、チャレンジャーか、あるいは自動生成されたものかの判別も容易に行えます。

「データ・サイエンティストがモデルを別の言語で再コーディングしたときに、元のモデルと同じ結果が得られず、その理由も不明というケースは珍しくありません」と話すのは、SASの上級副社長および最高技術責任者（CTO）であるオリバー・シャーベンバーガー（Oliver Schabenberger）です。「アナリティクス・ガバナンスが確立されていれば、異なる言語からでも同じデータに対して同じアナリティクスを実行することが可能となり、アルゴリズムの実装方法、各種のオプション、データモデルに関する特異性や相違点について心配する必要がなくなります。バックエンドにおける一貫性と標準化が、フロントエンドにおける相違点に関する懸念を取り除くのです。」

アナリティクス・ガバナンス・プラットフォームにより、すべてのデータユーザーが常に最良のアルゴリズムを活用できる環境が実現します。どの言語でプログラムを書くかや、それをどこから実行するかに関係なく、常に正確な結果が得られます。

「このガバナンスが確立されていれば、RパッケージまたはPythonライブラリで書かれたアルゴリズムのどちらが優れているかといった議論は、もはや無意味です。ユーザーがどのような方法で実行をトリガーするかに関係なく、どの言語でも、同じデータモデルに基づき同じ分析コードが実行されます。また、どの言語でも、マルチスレッド処理、分散コンピューティング、共通のデータアクセス、セキュリティ・モデルのメリットを享受できます」（シャーベンバーガー）。

シャーベンバーガーはアナリティクス・ガバナンスをWebガバナンスになぞらえ、1つのWebページを複数の異なるノートPC、タブレット、スマートフォンで呼び出しても常に同じコンテンツが表示されるようなもの、と説明します。それを可能にしているのがWebガバナンス標準です。

アナリティクス・ガバナンスも同様に機能します。ユーザーが組織全体の様々なシステムからデータを取り込んでモデルを作成できる全社規模のアナリティクス環境において、アナリティクス・ガバナンス・プラットフォームは一貫性、正確性、パフォーマンスを確保する役割を果たします。

オープンソース・アナリティクスと全社規模のアナリティクスを単一のプラットフォームで一元管理

この記事の冒頭で触れた銀行に話を戻しましょう。この銀行はアナリティクス・ガバナンスを確立することで、十分なレベルのオープンソース対応を実現できます。具体的には、ジェームズ、スーザン、クリスティーナ、マービンの全員が自分好みのツールを使えるようになるほか、同行は様々なパッケージのすべてを対象として、一貫したエンタープライズ・クラスの結果を保証でき

るようになります。また、IT部門はアナリティクスの実装とデータを統制する機能を手にし、誰もが同じデータモデルに基づいて作業する状態を確保できるようになります。

様々なユーザーが様々な方法でデータにアクセスする環境では、アナリティクス・ガバナンスこそが優れた意思決定のための基盤となります。「コンセプトからイノベーション、実装、ROIに至るまで、幅広い領域でアナリティクスの真価を発揮させたいのであれば、アナリティクス・ガバナンス・プラットフォームの整備は不可欠です」（マクニール）。



SASとオープンソースの連携をお考えなら、[SAS® Viya®](#)への移行をぜひご検討ください。

SASをフォローしてください：



SAS Institute Japan 株式会社 www.sas.com/jp

jpnasinfo@sas.com

本社 〒106-6111 東京都港区六本木6-10-1 六本木ヒルズ森タワー 11F
大阪支店 〒530-0004 大阪市北区堂島浜1-4-16 アクア堂島西館 12F

Tel: 03 6434 3000 Fax: 03 6434 3001
Tel: 06 6345 5700 Fax: 06 6345 5655

