

ビッグデータから真のビジネス価値の獲得へ  
スピードとガバナンスを両立させるプラットフォームの最新像

「データマネジメント」「アナリティクス」「ビジネス活用」を包括的に支援

データをビジネス価値に変えるためには、経験や勘を排し、データ分析から得られた洞察に裏付けられた意思決定を“最適なタイミング”で実施していくことが大切だ。すなわち「データマネジメント (データ準備)」→「アナリティクス (洞察)」→「ビジネス活用 (意思決定)」のステップを着実に実行する必要がある。その方法論としてSASが提唱しているのが、分析精度・成果向上/維持のためのアナリティクス・ライフサイクルだ。Hadoopと緊密に統合されたビジネス分析プラットフォームが、その取り組みを包括的に支援する。

## タイム・ツー・マーケットのニーズに応える データ分析のスピードが重要

SASが提唱するアナリティクス・ライフサイクルには、データから洞察を得るための準備や探索、モデル作成といった作業から、分析結果を実務の現場での活用につなげていくための業務適用、行動、評価まで、広範なプロセスが定義されている。このアナリティクス・ライフサイクルを繰り返し実行することで分析精度およびそれに裏付けられた意思決定の精度を高めていくことができるのだ(図1)。

ただし、エンタープライズ領域を対象としたアナリティクス・ライフサイクルならではの要件として、必ず満たさなければならないことがある。まずはタイム・ツー・マーケットのニーズに応えるスピードだ。SAS Institute Japanの山本圭一氏は、「業務側が求めるタイミングで分析結果を出し、アクションにつなげていかないと、せっかくのビジネス機会を損失してしまいます」と警鐘を鳴らす。

また、アナリティクス・ライフサイクルを通じて扱うデータそのものも従来の構造化データのみならず、テキストや画像などの非構造化データ、IoTデータなどへと多様化・大容量化している。そして当然のことながら、これらのビッグデータについてもセキュリティ対策や職務分掌に基づいた権限管理、アクセス監査、メタデータ管理、プロセス・モデル管理、バッチ処理など厳重な“統制”を効かせることが求められる。

大前提となるこれらの要件を踏まえ、ビッグデータを対象としたデータマネジメント、アナリティクス、ガバナンスの包括的なソリューションを提供することが、SASのビジネス分析プラットフォームの基本コンセプトである。

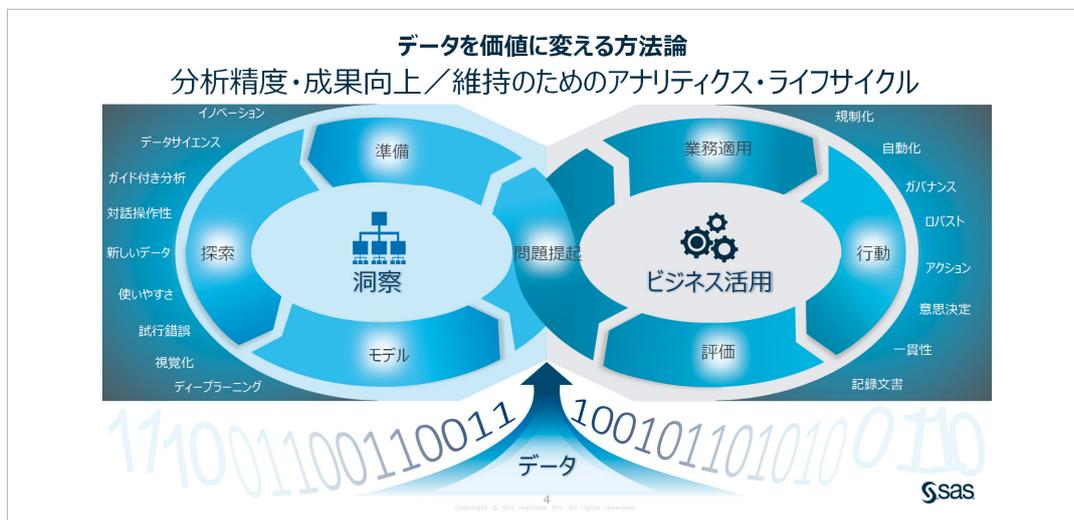


図1：SAS Instituteが提唱するアナリティクス・ライフサイクルの概要

## アナリティクス・ライフサイクルをHadoop上で実施可能

既存のデータ活用環境を補完するビッグデータ分析基盤として、SASが“本命の選択肢”と位置付けているのがHadoopだ。ビッグデータを格納・利用するための高速な分散ファイルシステムをリーズナブルなコストで構築することが可能。分散格納されたビッグデータに対して並列分散処理を実施するフレームワークがあり、その演算能力をリニアにスケールアウトすることも容易。冗長化によるデータ保護といったオープンソースソフトウェアならではのテクノロジーの先進性に注目したものである。

「SASのビジネス分析プラットフォームとHadoopは統合されています」と山本氏は強調する。具体的には、必要なデータにアクセス・変換・統合して、分析に必要なタイプや品質に加工する「データ準備」から、現状把握ならびに原因・関連性・予測などの洞察を獲得する「視覚化／分析」、プロトタイプ・モデリングやプロダクション・モデリング、機械学習／ディープラーニングに対応した「予測モデルの開発」、そのモデルをアクション（意思決定）に活かす「業務への組み込み」まで、アナリティクス・ライフサイクル全体を一貫してHadoop上で実施することが可能だ（図2）。

なお、SASとHadoopの連携には大きく2つの方法がある。1つは「from Hadoop処理」と呼ぶもので、SASからHadoopに対してSQLクエリーを実行し、得られた結果をSASサーバー上に抽出して分析処理を行う。要するにHadoopを高速なデータベースの一種として扱うわけだ。「これまで社内に蓄積されてきたSASの豊富な分析資産をそのまま活かすことができます」と山本氏は、そのメリットを語る。

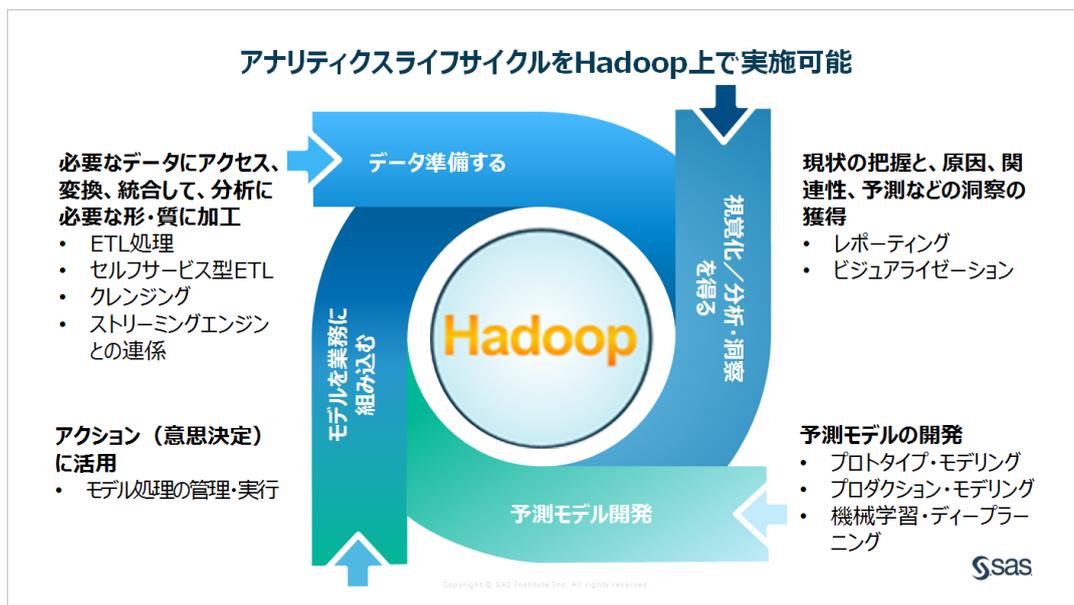


図2：Hadoopとの完全統合によって、あらゆるデータの分析：活用を加速する

ただし、SASとHadoop間のデータ移動は、従来のSQLインタフェースによる処理となるため、抽出容量が大きい場合はその時間的な遅延が、分析業務の大きなボトルネックとなる可能性がある。この課題を解決すべく提供しているのが、2つめの「in Hadoop処理」という連携方法だ。ビッグデータが格納されたHadoopに並列分散フレームワークにSASモジュール (SAS Embedded Process) を配置。データ加工やデータ品質改善処理を含むSAS処理を直接実行し、すべてのプロセスをHadoop上で完結させるのである。

さらにSASは、このin Hadoopのアーキテクチャを補完・強化すべく、Hadoopと統合可能なインメモリー分析サーバー「SAS LASR Analytics Server/SAS High-Performance Analytics Server」を提供している。「ビッグデータをメモリー上に並列高速ロードし、これまでBIツールで行っていたダイス (分析軸の切り替え) やドリルダウン (詳細データへの掘り下げ) など、インタラクティブな操作によるデータ視覚化やアドホック分析、レポートニングから各種機械学習までカバーします」と山本氏は語る。

## ビジネスユーザーの“スキルギャップ”を解消する セルフサービスのHadoop専用データ準備ツールを提供

さらにSASが重要テーマとするのが、データ準備すなわちデータマネジメントに関する作業の生産性向上と時間短縮である。実際、多くの企業がデータ分析にいたる前のデータ収集、コードの整備や標準化に手を焼いており、一般的には分析業務における80%の時間がデータ準備に費やされているとも言われている。

このような非効率性はHadoopに格納されたビッグデータを利用する場合には、より深刻になる。前述の通り、Hadoopデータを従来のSQLインタフェースで利用するにはスピード的な問題が発生する。本格的に利用するためには、Hadoopの並列分散フレームワークを利用することになるが、新たなプログラム言語を習得する必要となる。そこで、多くの場合IT部門などの専門家にデータ準備を依頼している。

SAS Institute Japan  
ソリューション統括本部 プラットフォームソリューション統括部  
マネージャー  
山本 圭一 氏



しかしIT部門は業務を知らないため、ビジネスユーザーが本当に欲しいデータが得られるまで、何回もやり直しが発生して、ビジネスユーザー、IT部門の双方ともに労力、時間がかかる結果となってしまう。この担当者間の“スキルギャップ”を是正できないか？

この課題を解決するため、SASが新たに提供を開始したのが「SAS Data Loader for Hadoop」である。ビジネスユーザーのために用意されたセルフサービスによるHadoop専用のデータ準備ツールだ。「SAS Data Loader for Hadoopを利用すれば、GUIのドラッグ&ドロップ操作でHadoopに格納されたビッグデータに直感的にアクセスし、管理することができます。幅広い業務部門のビジネスユーザーが複雑なプログラムを書くことなくデータの準備作業を簡潔に行うことができ、トレーニングも最小限ですみます」と山本氏は語る。

ビジネスユーザーやデータサイエンティストは、IT部門の手を借りることなくデータのプロファイリング、クレンジング、ジョイン、変換などの操作を自ら実行し、高度なアナリティクスのために活用できる。これによりIT部門もまた煩雑なデータマネジメントの負担から解放され、高度で信頼性の高いデータ活用基盤の構築・維持活動に、より重点的なリソースを割くことが可能となる。

ビッグデータを価値に変えるアナリティクス・ライフサイクルが、こうしてSASのビジネス分析プラットフォームによって確立されていくのである。



●お問い合わせ先



SAS Institute Japan 株式会社  
[www.sas.com/jp](http://www.sas.com/jp)

マーケティング本部宛て

03-6434-3700  
[JPNSASInfo@sas.com](mailto:JPNSASInfo@sas.com)