

Fact Sheet



主な機能

SAS Data Loader for Hadoop では、直感的な操作画面を通して Hadoop データにアクセスし、管理することができます。幅広い業務部門ユーザーがデータの準備作業をセルフサービス方式で行うことができ、トレーニングも最小限ですみます。技術スキルを持つユーザーは、Hadoop 上で SAS コードを記述・実行し、パフォーマンスやガバナンスの改善を図ることができます。

ビジネスメリット

大量のデータを保管するために Hadoop へと切り替える企業は増えていますが、一般的な Hadoop 環境では特別なコーディング・スキルが要求される場面が多いため、データ管理作業が難しい課題となっています。SAS Data Loader for Hadoop はこうしたスキルギャップを解消し、技術的な能力に関係なく、幅広いユーザーがデータに容易にアクセスできるようにします。

対象ユーザー

このソリューションはビジネスユーザー向けに設計されており、コードを記述しなくてもビッグデータにアクセスして活用できます。また、SAS 開発者やデータ・サイエンティストは、処理速度や生産性の向上を図るためにこのソリューションを活用できます。

SAS® Data Loader for Hadoop

セルフサービス方式のビッグデータ統合環境により、データ管理における IT 部門の負担を軽減

今日では、ほとんどの企業・組織がビッグデータの重要性を認識しています。ビッグデータを活用するためのアナリティクスやその他の先端テクノロジーは実用レベルに達しており、Hadoop などのシステムへの蓄積・保管も進んでいます。

しかし、データを保管することと、それを適切に管理することは、全くの別問題です。Hadoop データにアクセスするにはコーディングが必要ですが、コードの作成や保守には専門的な技術スキルが必要なため、Hadoop データ管理に関するスキル不足に悩む企業・組織が少なくないのが実情です。必要なデータを必要なときに利用できないのであれば、データを収集して“一等地”に保管しても意味がありません。しかし、業務部門ユーザーに残された選択肢はわずかであり、IT 部門に頼るか、自分でコーディングの勉強をするか、あるいは、スキルギャップを解消してくれるソリューションを探しかありません。

SAS Data Loader for Hadoop なら、直感的な操作画面で Hadoop データのプロファイリング、管理、クレンジング、移動を行えます。つまり、コーディングの方法を知らなくてもデータを操作することが可能です。また IT 部門側でも、基本的なデータ管理の負担が減るため、処理速度の向上やセキュリティの強化といったテクノロジー自体の利点を引き出す業務に集中できるようになります。

主な利点

最小限のスキルでデータ管理が可能

高度なトレーニングを実施したり、専門家を高給で雇ったり必要はありません。SAS Data Loader for Hadoop を導入すると、高度な IT 知識のないユーザーでも、データ統合、データ品質管理、データ準備の作業を行えるようになります。

ビッグデータのパワーを最大限に活用

SAS Data Loader for Hadoop によってスキル面の壁がなくなれば、蓄積したデータを業務に活用できる可能性は無限に広がります。データのプロファイリング、クレンジング、ジョイン、変換を (IT 部門や専門家の手を借りずに) 実行し、高品質なデータを生成した上で、高度なアナリティクスのために活用できます。

拡張性とパフォーマンスが向上

SAS Data Loader for Hadoop は、業務部門ユーザーの分析や意思決定を支援するのみならず、データ・サイエンティストや SAS 開発者による処理速度、効率性、俊敏性の改善も支援します。このソリューションでは、コード・アクセラレーターを活用して Hadoop のパワフルな処理をさらに高速化できます。また、データ移動が最小限に抑えられるため、データのセキュリティも強化することができます。

無償体験版

SAS Data Loader for Hadoop 体験版は、ダウンロード後 45 日間、Cloudera または Hortonworks の本稼動クラスターで利用いただけます。この体験版は、後日ライセンスをご購入いただいた場合は、ソフトウェアをインストールし直さなくても、そのまま正式版としてお使いいただけます。

製品概要

SAS Data Loader for Hadoopは、Hadoop上で全てのデータ管理操作を提供するソリューションであり、データ統合とデータ品質管理に関する機能を包括的にカバーしています。

使いやすさと高度な技術機能を融合させたSAS Data Loader for Hadoopは、業務部門とIT部門の双方にメリットをもたらします。

直感的なユーザー・インターフェイス

SAS Data Loader for Hadoopは、ビジネスユーザー向けに設計されています。直感的なウィザード方式の操作画面で、Hadoopデータのアクセスや管理を容易に行えるため、作業のたびにIT部門の支援を受けたり、Hadoopに詳しい専門家を雇う必要はありません。

Hadoopデータ管理のために構築されたツール

SAS Data Loader for Hadoopは、Hadoopでビッグデータを管理するために新たに構築されています。既存のIT部門向けツールの焼き直しではありません。

ビッグデータの品質管理

Hadoop環境内でデータを適切に管理することができます。SAS Data Loader for Hadoopには、データのプロファイリングを行い、全体的な品質を把握するための機能が用意されています。その上で、軽量型のSAS実行エンジンであるSAS Embedded Processを用いて、標準化、パース、同一性の判定といったデータ品質管理機能をHadoopの内部で実行することができます。

データ品質管理の新たに追加された機能には、ケーシング(大文字/小文字の変更)、ジェンダー分析、パターン分析、フィールド抽出などがあります。これにより、データに含まれるアルファベットの大小文字を変更する、データ値にもとづいて性別を推測することで顧客セグメンテーションの精度を高める、許容されるデータのパターンをフィールド値にもとづいて推測する、といった処理が可能になります。フィールド抽出機能は、フィールド内の非構造化テキストや自由入力テキストから、名前、組織名、住所、メールアドレス、電話番号などの有用なトークンを拾い出します。

プロファイリング機能は、Hadoopクラスター上で並列実行されるためパフォーマンスが向上し、時系列に沿ってデータの変化を追跡できるトレンドグラフも生成します。識別分析機能は、列に含まれるデータのタイプを判断します。例えば、列のデータ内に「NC」、「North Carolina」、「N Carolina」が含まれる場合には、「州」として分類されるため、データ探索に役立ちます。



図1:セルフサービス機能を利用して、必要なデータをHadoopへコピーすることができます。



図2: SAS Data Loaderでは主要な操作指示(ディレクティブ)が最初から組み込まれているため、トレーニングが最小限で済みます。