

(解説)

全社データ分析基盤 DataLab[®] の構築と活用

南 和男*¹・藤平雅信*²・逢坂武次*²

Building and Utilizing Company-wide Data Analytics Platform, DataLab[®]

Kazuo MINAMI・Masanobu FUJIHIRA・Taketsugu OSAKA

要旨

ものづくりにおけるデータ活用はますます重要性を増している。当社では、IoTやビッグデータの利用を早期に開始し、製品開発、サービス向上に貢献してきたが、データの全社的活用を効率的に進める統一プラットフォームが不足していた。本課題に対応するため、データ分析基盤DataLab[®]を構築した。この基盤は、大量データの効率的な収集・分析を可能にし、先進的分析ツールを活用して新しい洞察を提供する。DataLab[®]の目的は、データの一元管理、柔軟なインフラストラクチャ、高度な分析能力、厳格なセキュリティ、ユーザーフレンドリーな操作性を実現、提供することである。本稿では、材料開発と設備診断の事例を通じて、DataLab[®]の取り組みを紹介する。

Abstract

The utilization of data is becoming increasingly important in the art of manufacturing. Kobe Steel is an early adopter of IoT and big data, which has contributed to product development and service improvement, but lacked a unified platform for efficiently utilizing data across the company. To meet this challenge, a data analysis platform, DataLab[®], has been established. This platform enables efficient collection and analysis of large amounts of data and leverages advanced analytical tools to provide new insights. The purpose of DataLab[®] is to provide centralized data management, flexible infrastructure, advanced analytical capabilities, strict security, and user-friendly operability. This paper introduces the DataLab[®] initiative through material development and equipment diagnosis examples.

検索用キーワード

データ収集, IoT, MI, マテリアル・インフォマティクス, データ分析, 異常予兆, 状態監視

ま え が き = 情報通信技術やAI技術などのテクノロジーの進歩は、現代のビジネス環境における競争力を高めるうえで不可欠な要素である。とくに操業データや製品データの活用は、新たな製品開発や製品の改善、市場戦略の立案、顧客体験の向上での多岐にわたる分野で、企業の成長に不可欠な役割を果たしている^{1), 2)}。

当社においても2014年頃からIoTやビッグデータの重要性を認識し、それぞれの事業部門で製品開発や製造現場の改善、お客様サービスの向上などへの活用を開始した。しかし、当時はデータの蓄積や分析を全社で活用できるプラットフォームがなかったため、これらの活動をそれぞれ独立して行い、個別にデータやノウハウが蓄積された分析システムを開発していた。

その後、2016年から始まった当社の中期計画では、IT企画部と技術開発本部が共同で、操業データや製品データの分析支援や、データサイエンティストの育成を開始した。また、デジタル技術の進化により、コンピュータやネットワークの処理能力が大幅に向上し、クラウドサービスが発展したことによって、以前は困難だった大量データの収集、伝送、処理が格段に容易になった。

このような背景を踏まえ、当社はデータとノウハウを統合的に管理し、全社規模での組織的なデータ活用を促進するための基盤として、データ蓄積から前処理、分析

までの機能を提供するデータ分析基盤 (DataLab[®]^{注1)}の構築に着手した。

DataLab[®]は、従来活用が難しかった、研究開発データや操業設備の時系列データなど、あまり構造化されていない大量のデータを扱える環境を全社に提供することを目指した。そのため、膨大なデータを迅速に収集・蓄積し、必要に応じて柔軟にリソースを追加したり、容易にアクセスしたりできる環境を準備した。また、先進のデータ分析ツールやAIアルゴリズムを組み込み、データから効率的に洞察を抽出して、新たな価値創出を加速することを目指した。

本稿では、1章でDataLab[®]の概要、2章で最初にテーマとして取り組んだ材料開発と設備診断の事例の概要について述べ、DataLab[®]の取り組みを紹介する。

1. DataLab[®]の概要

データ分析基盤の構築をするうえで重要なことは、データの統合と管理、柔軟に計算資源を変更できるインフラストラクチャ、高度な分析ツールや技術、データのセキュリティ保護、操作の容易性である。

とくにデータの統合と管理は、データの正確性、完全

脚注1) DataLab[®]は当社の商標である。

*¹ IT企画部 *² 技術開発本部 デジタルイノベーション技術センター

性、信頼性を確保し、分析のための有効な情報とするために不可欠である。業務システムや設備など様々なデータソースから収集したデータは形式の不一致や不完全なデータが存在するため、データクレンジング（形式や不完全データの修正）、データ変換（データを共通のフォーマットや標準形式に変換）、データ保存（データを後続のビジネスロジックや分析モデルに適合させて保存）を通してデータの質を高める。

つぎに、柔軟な拡張性をもったインフラストラクチャを確立することで、データ量の増加や計算処理量の増加に柔軟に対応できるようになる。クラウドサービスの活用は、このような拡張性を実現するうえで非常に効果的である。

さらに、高度な分析ツールと技術を導入することで、統計分析、機械学習、ディープラーニングなど、目的に応じた分析手法を適用し、データからの洞察を効率的に抽出できるようになる。また、データのセキュリティ保護は、データアクセスの厳格な制御や暗号化技術の導入で対処している。

最後に、より多くの人が容易に分析を行えるよう、汎用的なデータ可視化ツールを組み込むことで、意思決定者も情報の結果を迅速かつ直感的に理解し、行動を起こせるようになる。

DataLab[®]は、データ分析に必要な機能を具備した基盤であり、大きくはデータを蓄積するデータベース層と、データ分析やソリューション開発を行うソリューション層に分かれている（図1）。以降はそれぞれの層について概要を述べる。

1.1 データベース層

データ管理における主要な課題として、元データが個人端末や設備内部のコンピュータ、共有ファイルサーバなど様々な場所に分散して保存されている点がある。これに加え、データ間の関連付けルールが不在であったため、分析に適した形式へのデータ変換や、同じテーマで別の人が作成したデータの関連付け作業に大きな労力が必要であった。また、設備内部の制御用コンピュータに

保存されたデータは、そのコンピュータのデータ容量の制限があり、数箇月で消失することもあることから、必要なデータが存在しない、利用できないとの問題も生じている。

効率的な分析を行うためには、収集した生データを分析テーマに合わせて適切な形式に変換し、目的に応じたデータベースに保存することも必要である。

これらの課題に対応するため、DataLab[®]のデータベース層では、ExcelやCSVの構造化データ、XMLのような半構造化データ、画像や音声ファイルなどの非構造化データを含む、様々な形式の生データを同一のデータ集約領域に一元的に保存、管理できるようにした。

また、データ収集からデータベース層への保存までのプロセスを標準化することでデータ収集・保存を容易にするとともに、データ関連付け機能を具備させるなど、データ分析のための前処理工程の手間を大幅に削減できるようにした。その際、未加工の状態生データを保管し、将来的な分析ニーズの追加にも対応できるようにしている。これらのデータはクラウド上に保存されるため、無制限にデータを保存し続けることが可能になる。

以上の処理を経て、データベース層に整形・蓄積されたデータは、つぎのソリューション層で作成するモデルやツールから参照して利用される。

1.2 ソリューション層

ソリューション層の目的は、蓄積されたデータを基に、具体的なビジネスソリューションを開発することである。ソリューション層では当社が独自に開発した分析ツールの利用や市販の分析ソフトウェア、BI（Business Intelligence）ツールなどの可視化ツールを利用して、アプリケーションの開発や分析を行う。

作成された独自開発分析ツールには、当社のデータサイエンティストがそれぞれの分析テーマに合わせて機械学習・チューニングしたモデルを組み込んでおり、データベースに蓄積されたデータを高度に分析することが可能になる。

また、可視化ツールは、前述のデータベースにアクセ

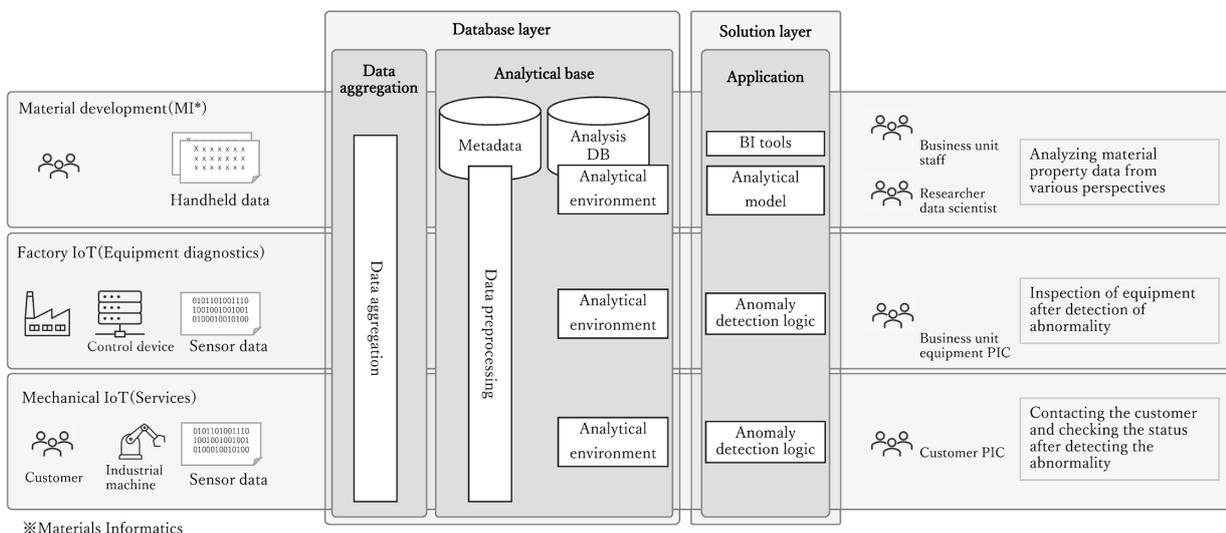


図1 DataLab[®]の概要図
Fig.1 Concept of DataLab[®]

スして必要なデータを取得し、可視化することでユーザーの意思決定を補助する役割を担う。

さらに、外部のデータ分析サービスとセキュリティを確保した状態でデータ連携が可能であるため、他社が準備したデータ分析モデルを容易に利用できる。そのため、開発のマンパワーや、AIに関する専門的な知識がなくても簡単にデータ分析や予測が可能となる。

2. 活用事例

データ分析では、一つのシステムで全ての分析ニーズに対応することは難しいため、材料開発や設備診断などの大きなテーマごとに最適な構成を設計し実装することが必要になる。例えば、材料開発で扱うデータは非定型かつ少量データであるいっぽう、設備診断では定型的で大量のデータを扱うため、データの整理の手順や最適なデータ構造も異なる。

今回、当社で幅広く活用が見込める材料開発と設備診断をユースケースとして設定し、DataLab[®]の構築に取り組んだ。本章では、材料開発と設備診断の事例（本号「材料開発および利用におけるMI技術」p.57、本号「加古川製鉄所第2分塊工場における設備状態監視・異常予兆検知システムの構築」p.17、参照）に基づき、データ分析基盤としての概要を説明する。

2.1 材料開発

材料開発では、一般的に、新しい材料の開発や既存材料の改良に関連する多様な実験データが生成される。これには、物理的特性、化学的反応、耐久性試験などのデータが含まれ、それぞれが異なる形式で記録される。また、担当者ごとにデータを管理していることが多く、整理の仕方も担当者、実験時期、内容によって統一されていないことが多い。そのため同じ項目名称でも担当者によってその項目にひもづくデータが異なることや、逆に同じ意味のデータであっても項目名称も異なることが多く、データ分析可能な形式への整理負荷が大きかった。

この整理を簡単にできるように、収集したデータを自動的に分析し関連データを分類したうえで、解析用デー

タベースに体系化して保管する仕組みを組み込んだ。

材料開発では、特定のルールに従って記述されたcsvファイルなどの実験データをデータ集約領域に保存する。そのデータを分類分けし、新たにテーブルを作成するとともに、キー項目をもとにテーブル間の関連を抽出し、逆スタースキーマとして構成する。そして、リレーショナルデータベースにデータを保管し、必要なデータテーブルを自由に組み合わせることで分析データを取得できるようにした（図2）。

データ分析時は、使用したいデータを目的別にテーブル形式でまとめ、市販の表計算ソフトウェアやBIツールなどの可視化ツール、当社のデータサイエンティストが機械学習・チューニングすることにより作成した様々なモデルを組み込んだ独自開発ツールで分析する。本ツールは、分析者が目的に応じたモデルを選択し、変数や条件を入力して計算を実行することで、材料特性の予測や目的材料の設計探索を行うことができる。また、高度な計算をクラウド上で実行することができるため、ローカル端末のスペックにも依存しない高速な計算が可能となる（図3）。

2.2 設備診断

設備診断の分野では、工場の設備稼働率の向上や操業の安全性を確保するために、様々なセンサからのデータがリアルタイムで収集される。これらのデータには、設備の動作状態、温度、圧力、振動レベルなどの情報が含まれ、設備の健全性や性能のモニタリングに利用されるとともに、予防保全や障害予測にも役立てることができる。

そのセンサー数は非常に多く、取得周期も極めて短いため、非常に大量のデータを扱う必要がある。いっぽう、取得するデータは定型化されたデータであるため、データ収集の仕組みを構築すれば他設備へも容易に展開することができる。また、設備診断については、多くの開発ツールが市販されており、市販ツールと独自開発ソフトウェアを組み合わせることで、効率的に分析システムを開発できる。

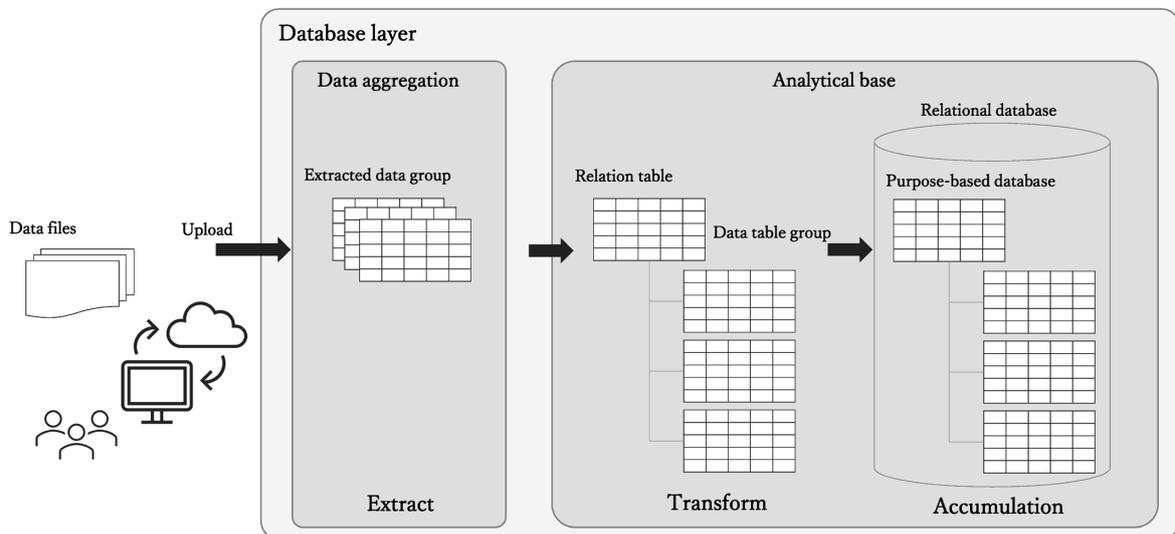


図2 データベース層におけるデータ処理概略図
Fig.2 Data processing image on the database

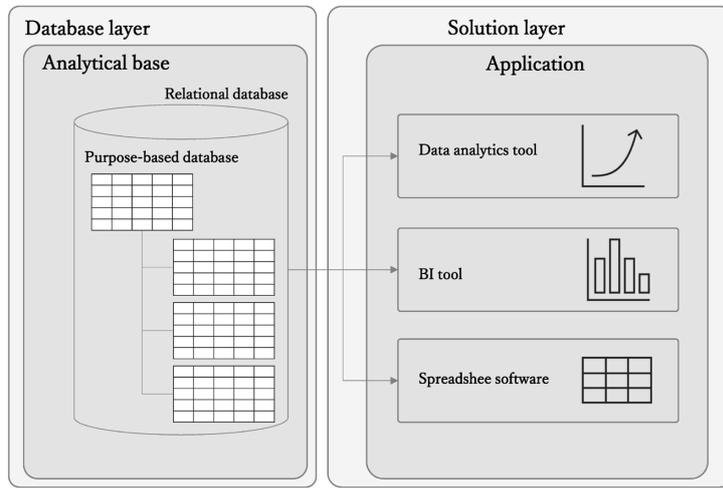


図3 材料開発 DataLab[®]の構成図
Fig.3 System architecture of DataLab[®]

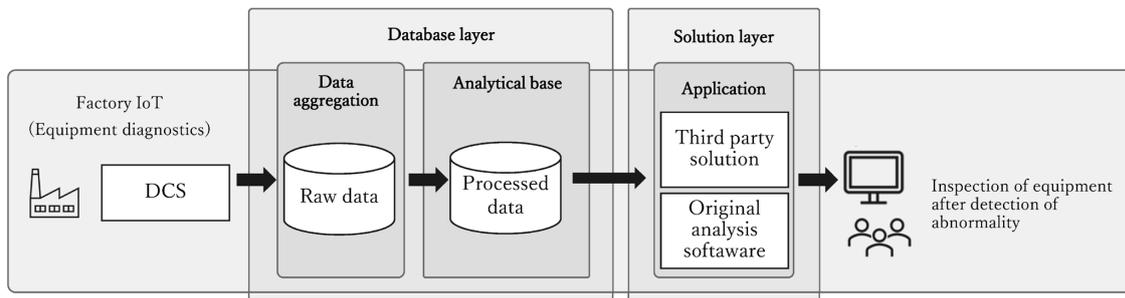


図4 設備診断におけるデータ処理概略図
Fig.4 Data processing image for Equipment Diagnosis

設備のデータ収集から分析まで構築した構成例を図4に記す。

データの収集には、DataLab[®]へのデータ伝送ソフトウェアを搭載したゲートウェイ装置を用いる。本装置はDataLab[®]への伝送と保存先があらかじめ組み込まれており、データ発生元の制御装置（DCSやPLC）と接続すればDataLab[®]へデータ伝送できる。伝送されたデータは所定の形式でデータ集約領域に生データとして保存される。そのデータは後続の分析モデルや分析ソリューションで分析しやすいように処理され整形データとして保存される。

設備診断の場合、各設備に接続された様々なセンサなどから多様なデータが収集されるため、データ構造としては膨大なデータ種類（列項目）が生成されるが、いっぽう、実際の異常検知モデルで有効なデータは検知したい異常ごとに限定されたものとなる。今回、設備診断のための解析用データベースとしては、時系列データを所定時間ごとに分割したうえで列指向データベースを構築して格納し、分析しやすさとパフォーマンスの両立を図った。

分析ソリューションとしては、市販ツールを活用し、あらかじめ用意されたアルゴリズムを活用した異常検知モデルや、正常時の運転データをもとにした異常予兆の検知機能を実装したが、社内のデータサイエンティストがPythonなどの統計解析向けプログラムを駆使して独自分析ソフトウェアを構築し実装してもよい。また、データの可視化ツールも具備しており、それを用いた設備

監視も可能である。

3. 今後の展望

材料開発と設備診断に適した環境を整備したが、まだ一部の事業部門で利用を開始したばかりである。今後は他事業部門へ展開していくとともに、検査や材料開発で要望の多い画像データへの対応など新たな機能を追加していく。

また、現場の装置や個人のパソコンなど、将来価値を生み出す可能性のあるデータが様々な場所に多く存在するが、データ集約により事業部門間でのデータ連携につながる可能性もあるため、データ集約を進めていく。

さらに、手元にデータはたまっており、簡易にデータ分析をしたいというニーズも存在するため、多くの人がデータ分析を簡単にできるように、DataLab[®]を用いたデータ可視化支援やノウハウの蓄積・共有も推進していきたい。

むすび=本稿では、組織的データ活用を促進するためのDataLab[®]のコンセプトを活用事例も交えて紹介した。ものづくり力の向上にデータ活用は必要不可欠だと考えている。そのため、DataLab[®]を用いて様々な工場や設備のデータ、実験のデータの蓄積と分析を推進していくことで、当社のものでづくり力の向上に貢献していく。

参考文献

- 1) 森田 登氏ほか. 三菱電機技報. 2022, Vol.96, No.5, p.40-43.
- 2) 田谷文彦ほか. JFE技報. 2022, No.50, p.44-49.