

NTT DATA

# データ駆動型ビジネス実現の 課題と解決法

NTTデータ先端技術株式会社

# データ駆動型ビジネス実現の課題と解決法

課題に対しては、下記の対応策をご紹介します。

## 課題

1

AIとデータ活用を推進する  
人材の不足

2

部門や役割間での  
分析環境のサイロ化

3

データ準備における  
リードタイム発生

4

セキュリティ・ガバナンスの煩雑さ

## 対応策

1

技術的ハードルの緩和・育成環境の整備



2

分析環境の共通化



3

データ準備系ツールの導入



4

アクセス制御や変更履歴の一元管理化





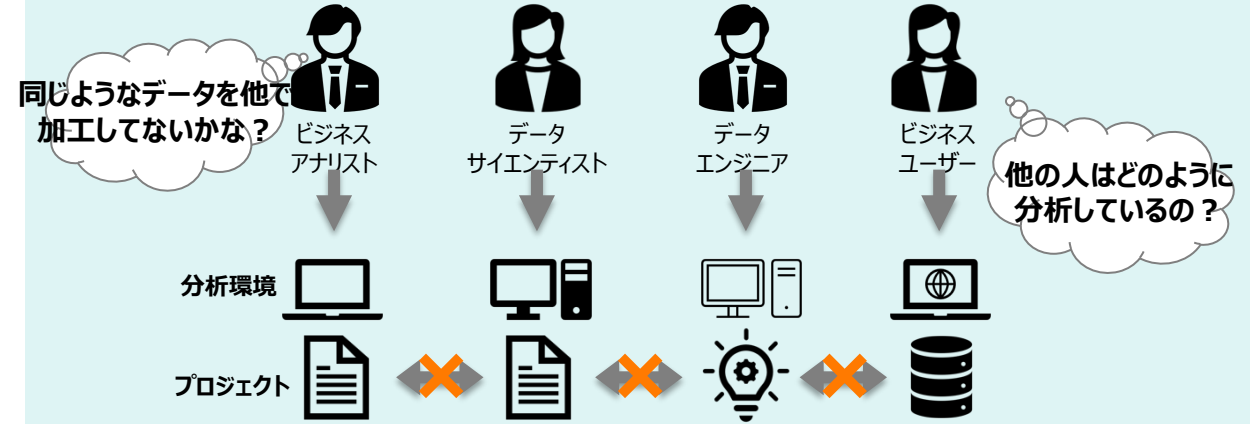
# データ駆動型ビジネス実現の課題

データ活用のさらなる推進を妨げる代表的な課題は以下の通りです。

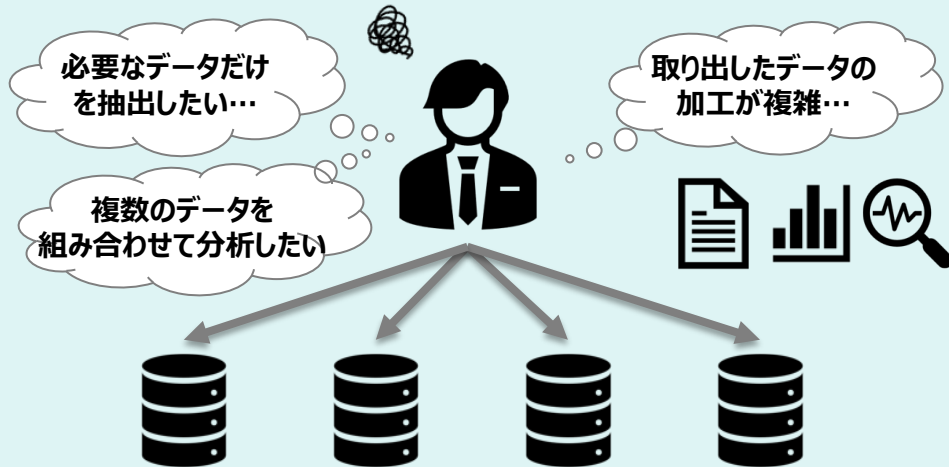
## ① AIとデータ活用を推進する人材の不足



## ② 部門や役割間での分析環境のサイロ化



## ③ データ準備におけるリードタイムの発生



## ④ セキュリティ・ガバナンスの煩雑さ

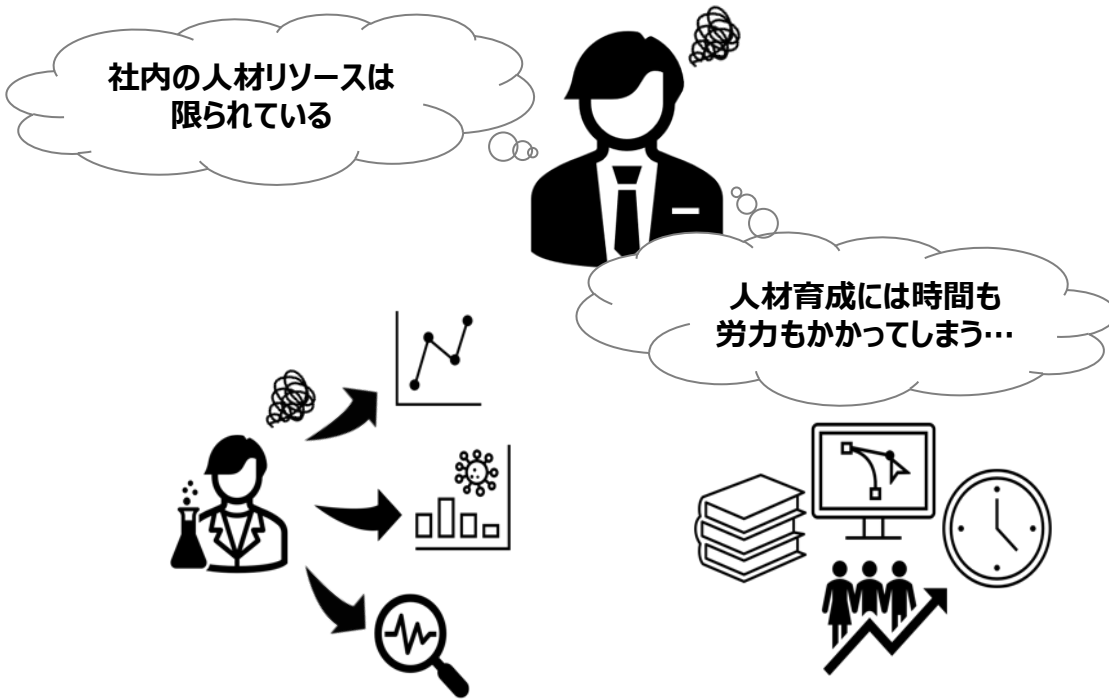


# 課題① AIとデータ活用を推進する人材の不足への解決法

AI・データ活用の民主化を進めるためには、技術的な専門知識が無いユーザでも作業できるように技術的なハードルを下げる仕組み作りや効率的に学習できる仕組み作りが必要です。

## 課題

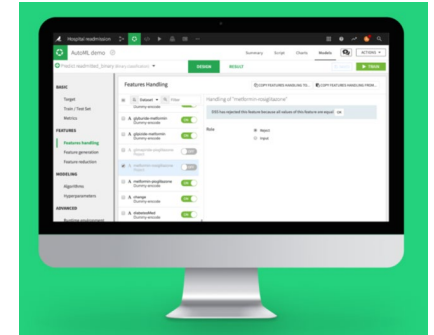
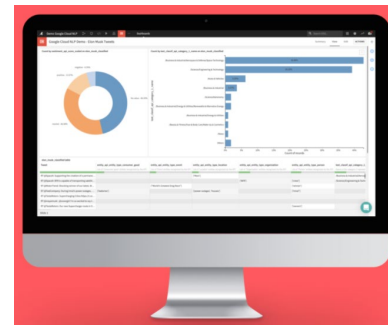
蓄積したデータから必要な形に加工し、課題や改善点を導き出して、課題の解決に役立てることができる人材がいない。



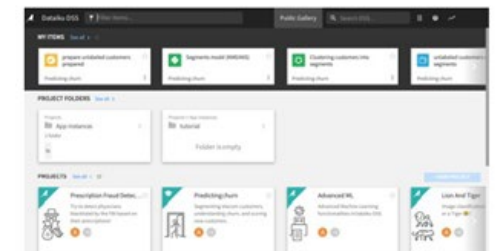
## 対応策

AI・データ活用の技術的ハードルを下げる仕組み作りや効率的に学習できる仕組み作りを行う。

### NoCode、GUI、AutoML



### Eラーニング



## 課題② 部門や役割間での分析環境のサイロ化への解決法

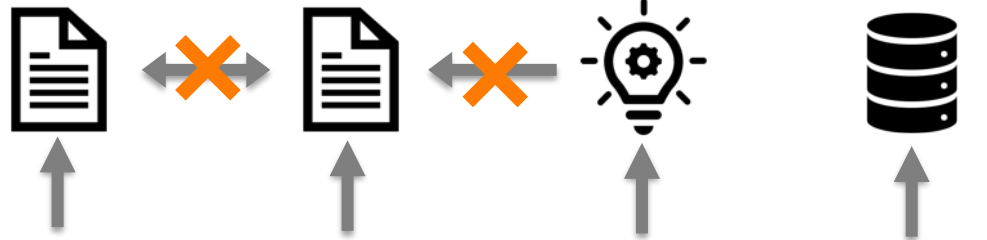
部門・役割間でのシームレスな連携を実現し、コード・データの再利用やノウハウ共有を可能とすることで、AI・データ活用を促進させる必要があります。

### 課題

部門・役割間での分析環境のサイロ化により、データ・コード・ノウハウの連携が困難となる。

### プロジェクト

部品Aの  
需要予測      部品Bの  
需要予測      需要予測  
モデルのノウハウ      部品Aの  
売上データ



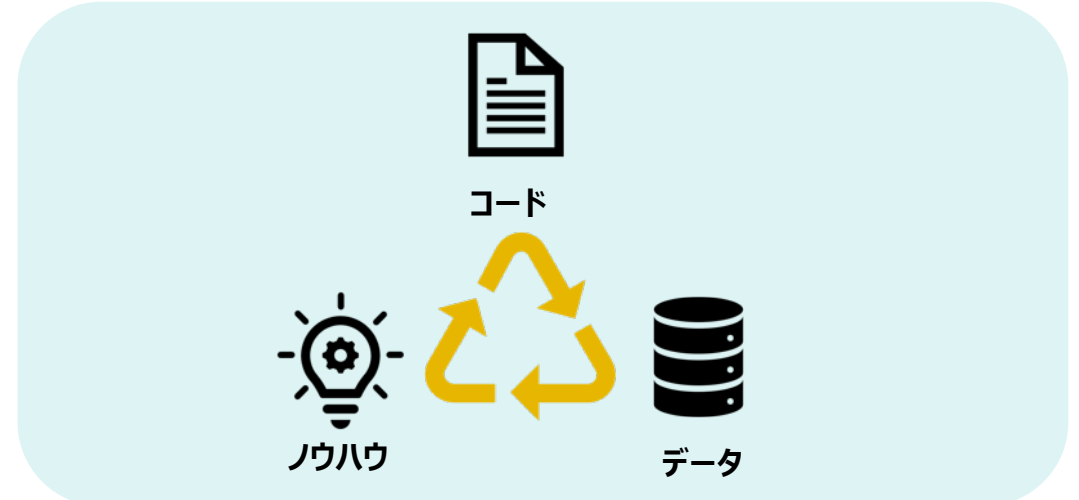
### 分析環境



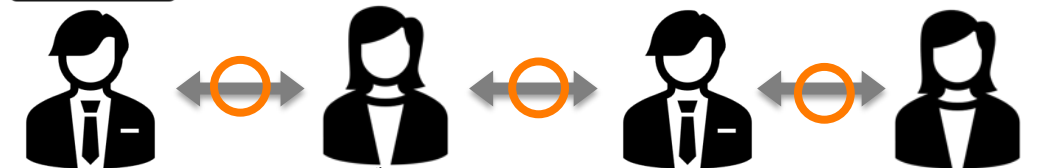
ビジネスアナリスト      データサイエンティスト      データエンジニア      ビジネスユーザー

### 対応策

部門や役割に関係なく、作業者が共通で使用できる環境の構築を行う。



### 分析環境



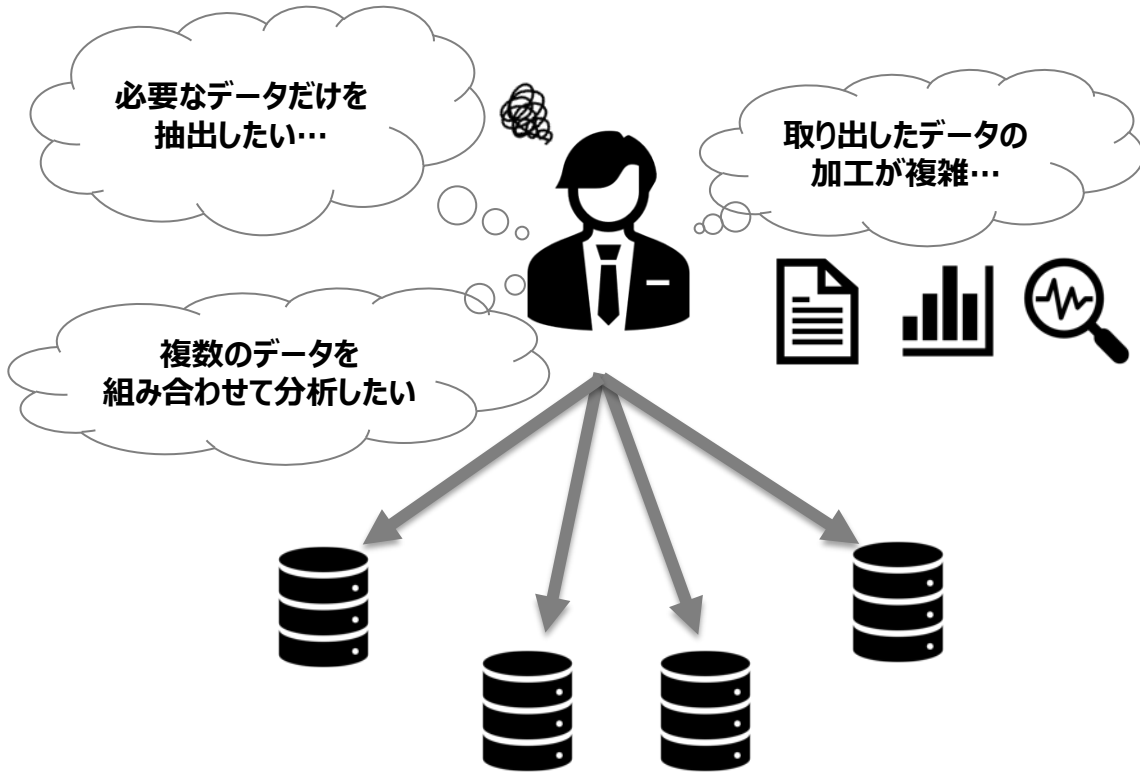
ビジネスアナリスト      データサイエンティスト      データエンジニア      ビジネスユーザー

### 課題③ データ準備におけるリードタイムへの解決法

データの準備作業に全体の80%を要すると言われており、この部分をできるだけ効率化し、データ準備にかかるリードタイムを短縮することが必要です。

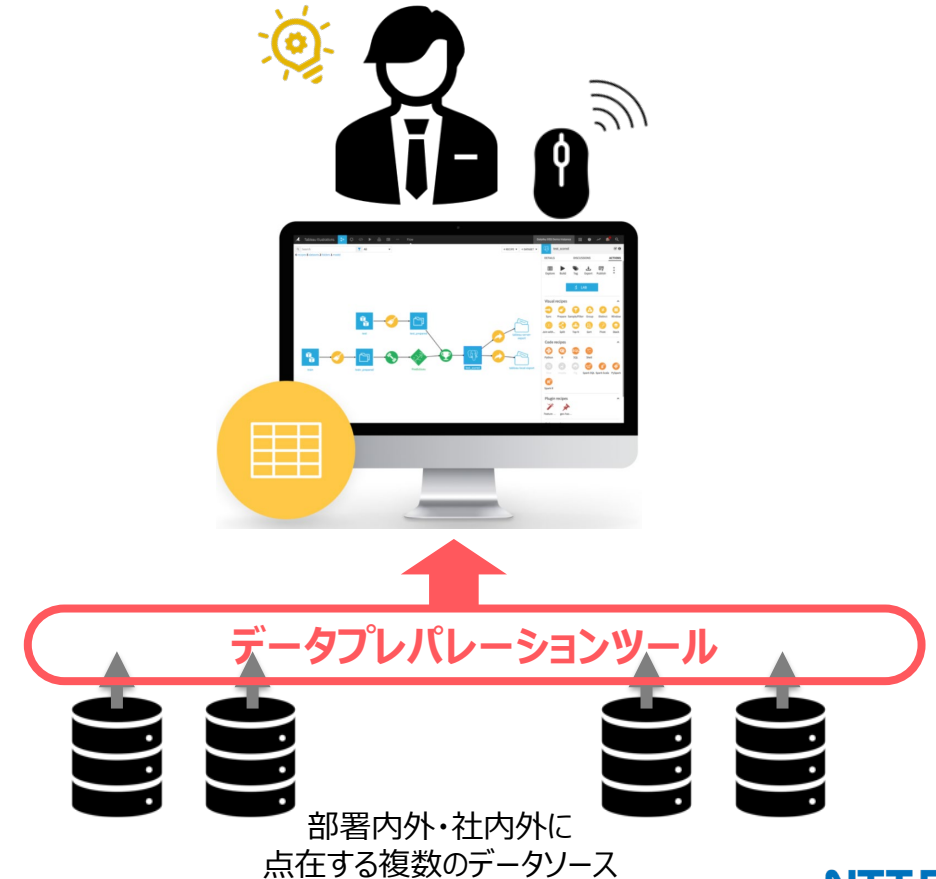
#### 課題

データの抽出や加工に時間を要し、本来やりたい分析業務に十分な時間を充てられていない。



#### 対応策

様々なデータソースからデータを抽出し、簡単に加工・保管できる環境の構築を行う。

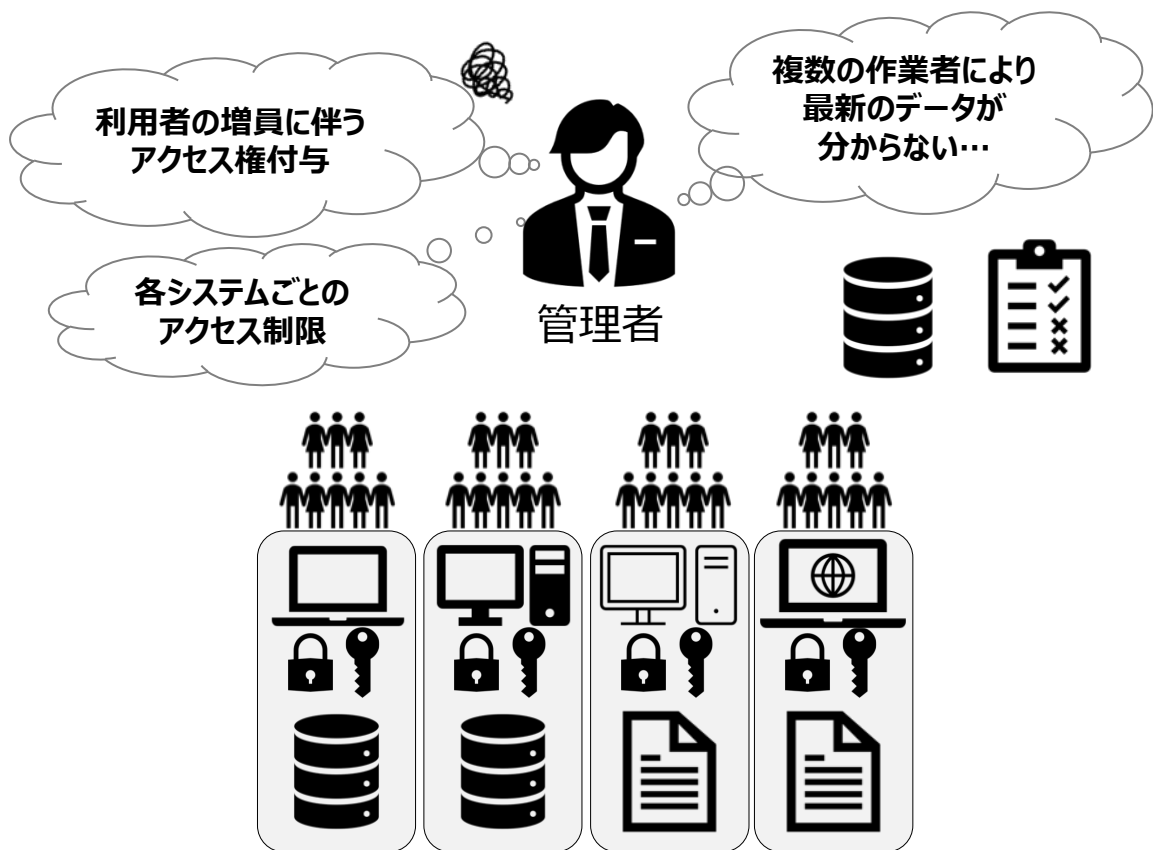


## 課題④ セキュリティ・ガバナンスの解決法

各データへのアクセス権を一元的に管理できる仕組み作りが必要です。また、監査対策として作業者の作業履歴やデータの変更履歴を追える仕組み作りが必要です。

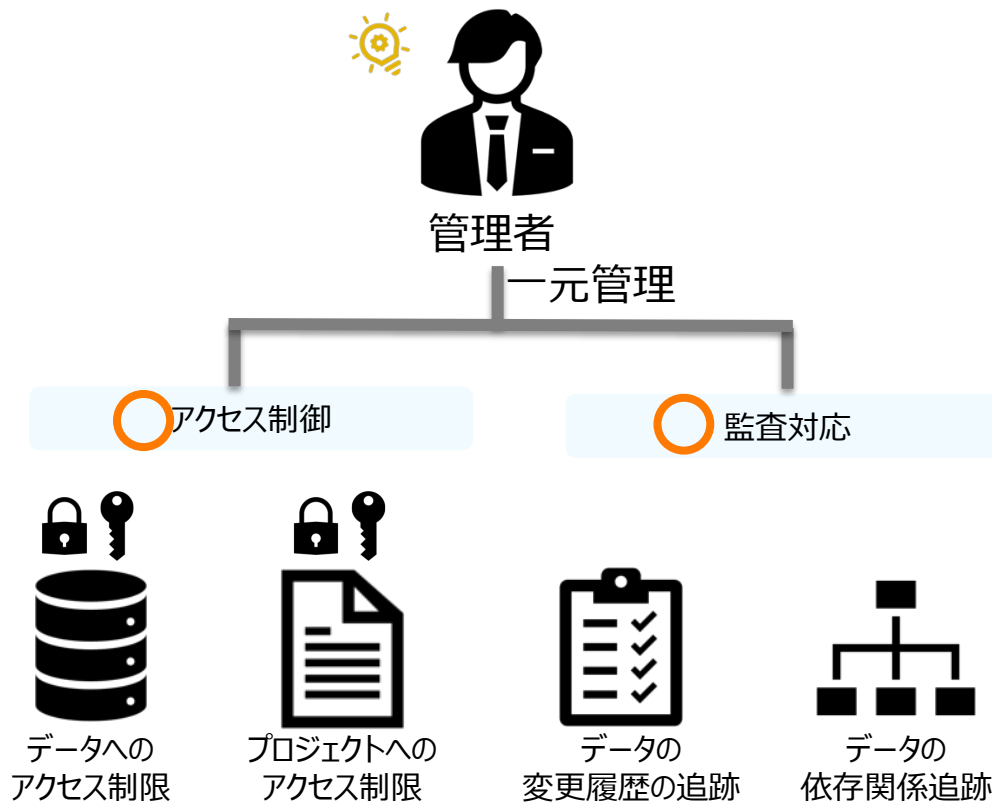
### 課題

利用者の増加に伴いアクセス制御が煩雑となり、また、誰がどんな作業をしているのか管理が難しい。



### 対応策

アクセス制御が一元管理できる環境の構築やデータの変更履歴が確認できる仕組み作りを行う。



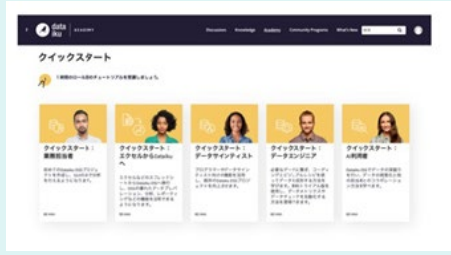
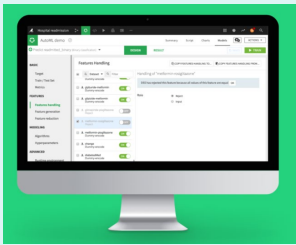
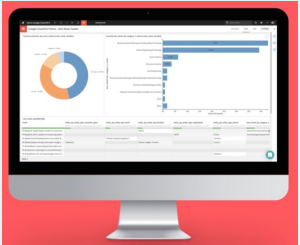
# 解決法のまとめ

データ活用における代表的な課題の対応策をまとめると以下の通りです。

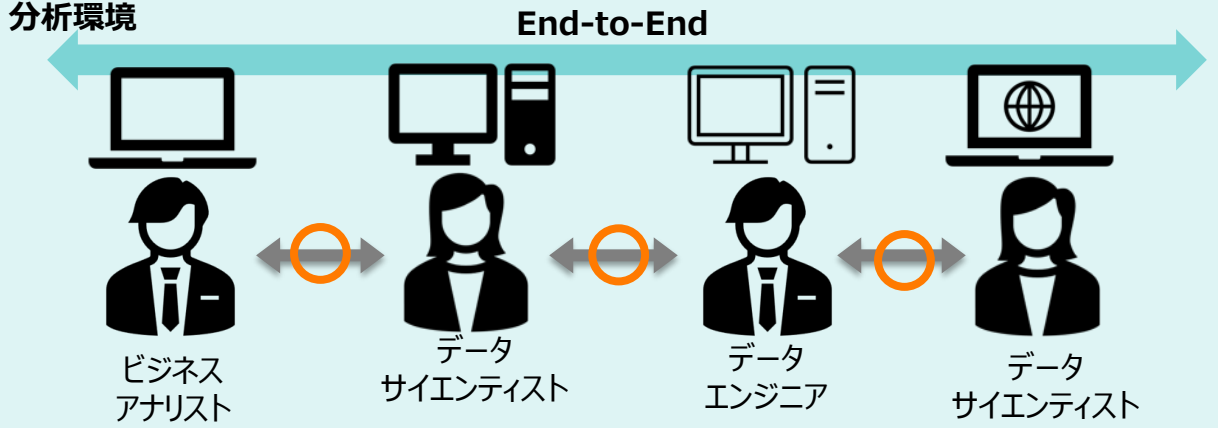
## ① 技術的ハードルの緩和・育成環境の整備

NoCode、GUI、AutoML

Eラーニング



## ② 分析環境の共通化



## ③ データ準備系ツールの導入

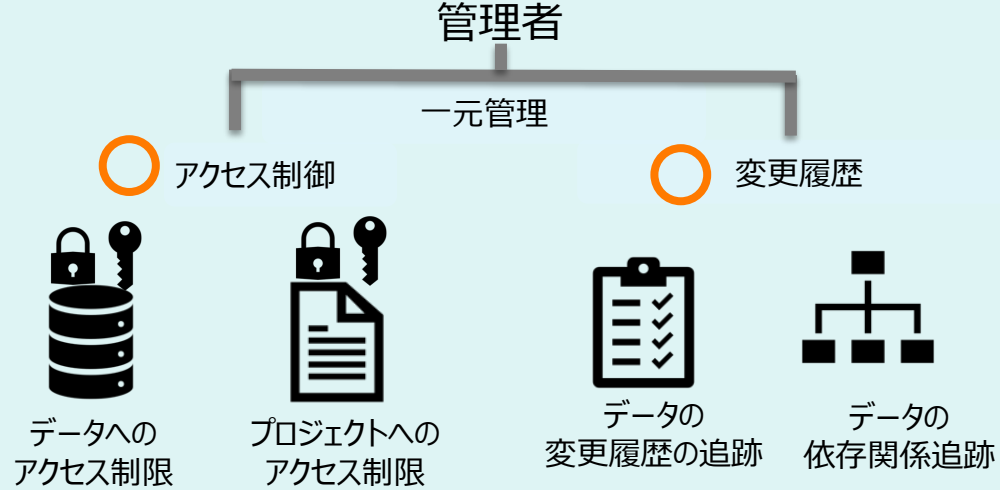


データプレパレーションツール

部署内外・社内外に点在する複数のデータソース



## ④ アクセス制御や変更履歴の一元管理化





# 対応策の具体的検討



# 専用ツール（クラウドサービス）を組み合わせたアプローチ

専用ツール（クラウドサービス）を組み合わせてスモールスタートさせる方法があります。

## 課題

1

AIとデータ活用を推進する  
人材の不足

2

部門や役割間での  
分析環境のサイロ化

3

データ準備における  
リードタイム発生

4

セキュリティ・ガバナンスの煩雑さ

## アプローチ

1



NoCode/AutoMLツール等

2



コラボレーションツール

3



データプレパレーション

4



管理ツール



専用ツール（クラウドサービス）の組み合わせは、スモールスタートしやすい。



しかし、運用段階で、ITエンジニア等の人件費も合算し計算すると、逆にコストが悪くなることも...

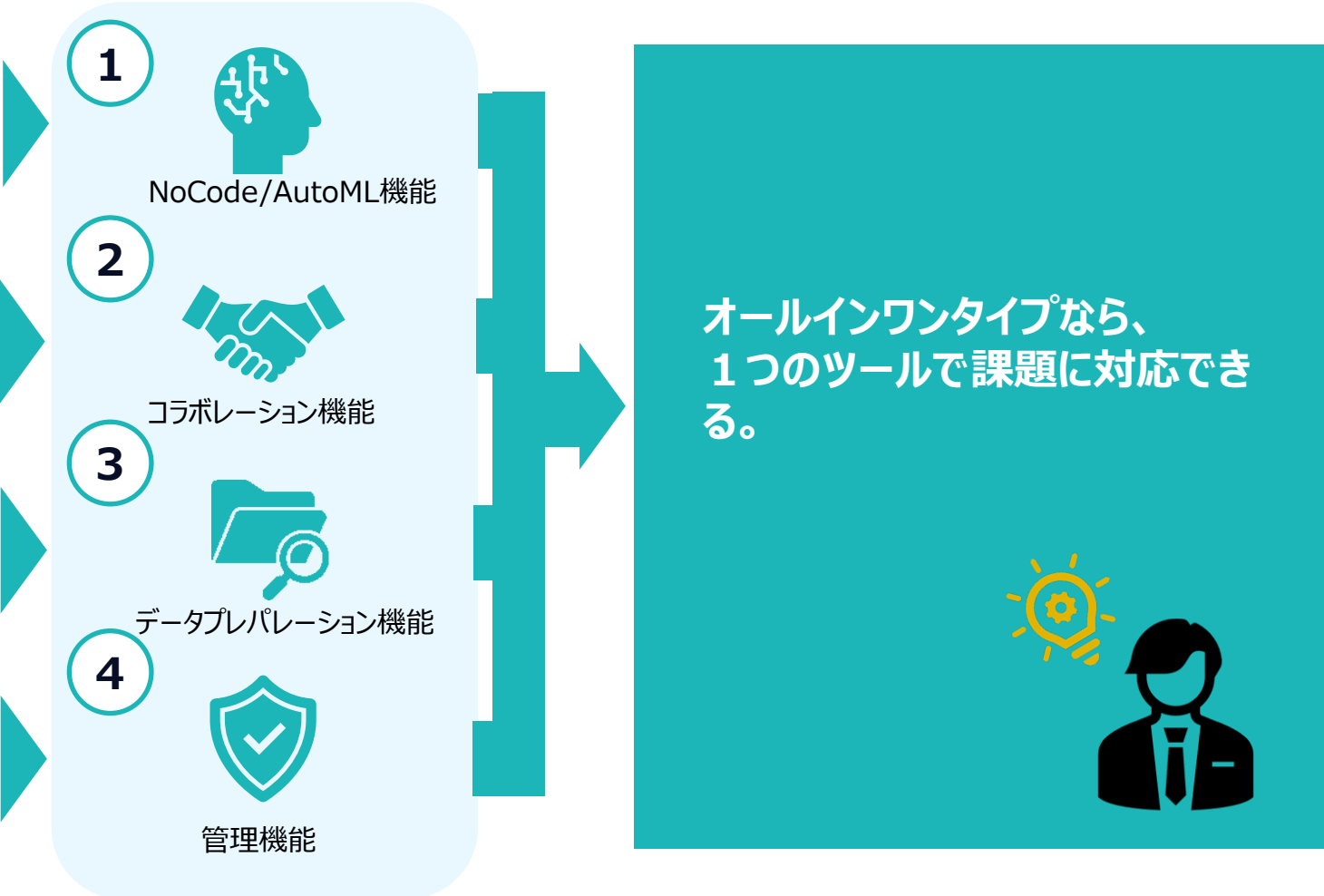
# オールインワンタイプによるアプローチ

AutoML機能やデータプレパレーション機能などが1つになったオールインタイプも誕生し、近年、注目され始めています。

## 課題

- 1 AIとデータ活用を推進する人材の不足
- 2 部門や役割間での分析環境のサイロ化
- 3 データ準備におけるリードタイム発生
- 4 セキュリティ・ガバナンスの煩雑さ

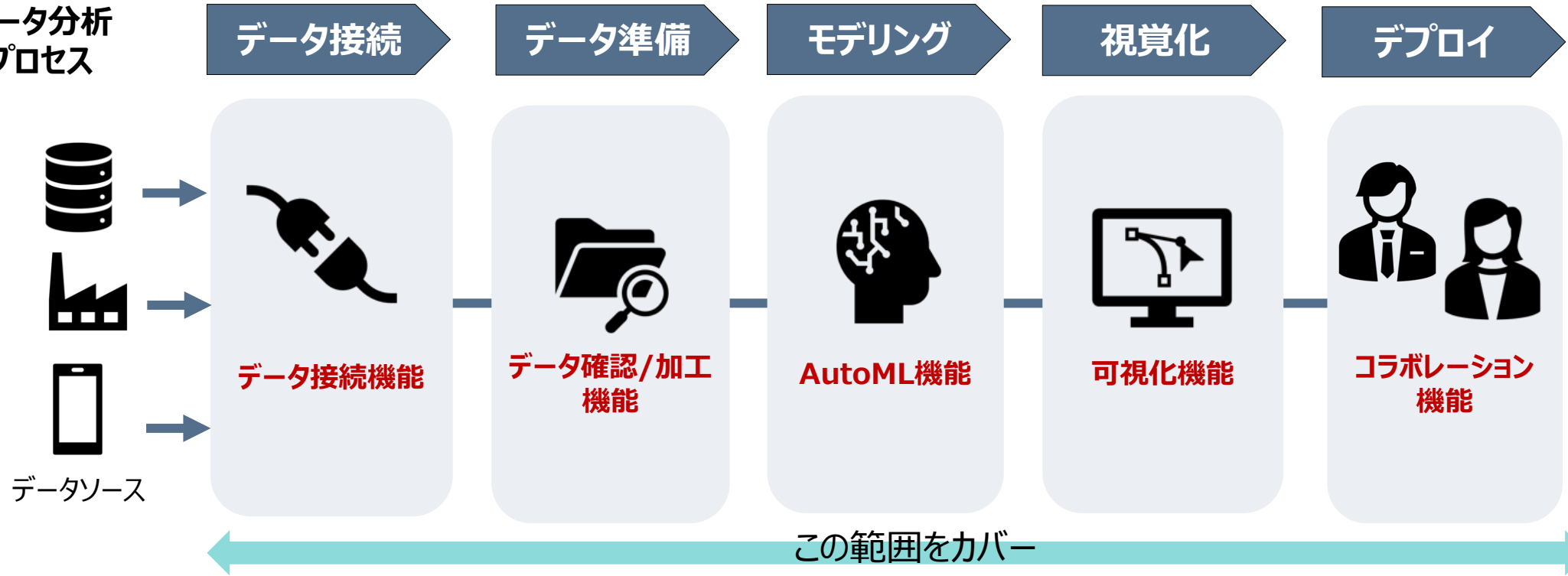
## アプローチ



# オールインワンタイプの詳細

オールインワンタイプは、下記のデータ分析プロセスをカバーいたします。

データ分析  
プロセス



オールインワンタイプは、データサイエンス&機械学習プラットフォームとして注目されています。

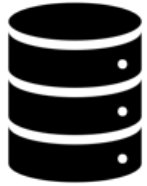
# Dataikuとは



Dataikuは、オールインワンタイプのデータサイエンス&機械学習プラットフォームの1つです。

データ容量/構造化・非構造化データ/オンプレ・クラウド環境を問わず、ひとつのプラットフォーム上での作業が可能です。

## End-to-End



### データ分類化 データソースへの接続

### データ探索 データ加工

### 機械学習

### インサイト共有

### オーケストレーション 本番展開・モニタリング

- 40種類以上のデータソース標準接続
- オンプレ・クラウド環境を問わず接続可

- データ統計分析の視覚化
- データ変換機能のプリセットを90以上内蔵

- AutoML
- ノートブックML
- カスタムモデルの作成

- ビジネスダッシュボード
- プロジェクト結果を視覚化

- DataOps
- MLOps

コラボレーション

セキュリティ・一元化されたガバナンス

NoCodeで実装可能・GUIだけで分析可能

業界ソリューションカタログ・教育コンテンツ



# Dataikuの詳細

# Dataiku社について

## データサイエンス&機械学習(DSML)プラットフォームのリーダー

- 2013年創業のユニコーン企業
- Gartner Magic Quadrant 2021で  
データサイエンス&機械学習(DSML)プラットフォームのリーダーに2年連続選出
- 従業員 1,000名以上 (2021年10月時点)
- 2022年にDataiku Japan株式会社を設立

### 製造や金融、官公庁まで 主要顧客500社以上



### オフィスは8カ国に展開

- 東京
- ニューヨーク
- パリ
- ロンドン
- アムステルダム
- フランクフルト
- シドニー
- シンガポール
- デンバー
- ワシントンDC
- ロサンゼルス



### 財務状況

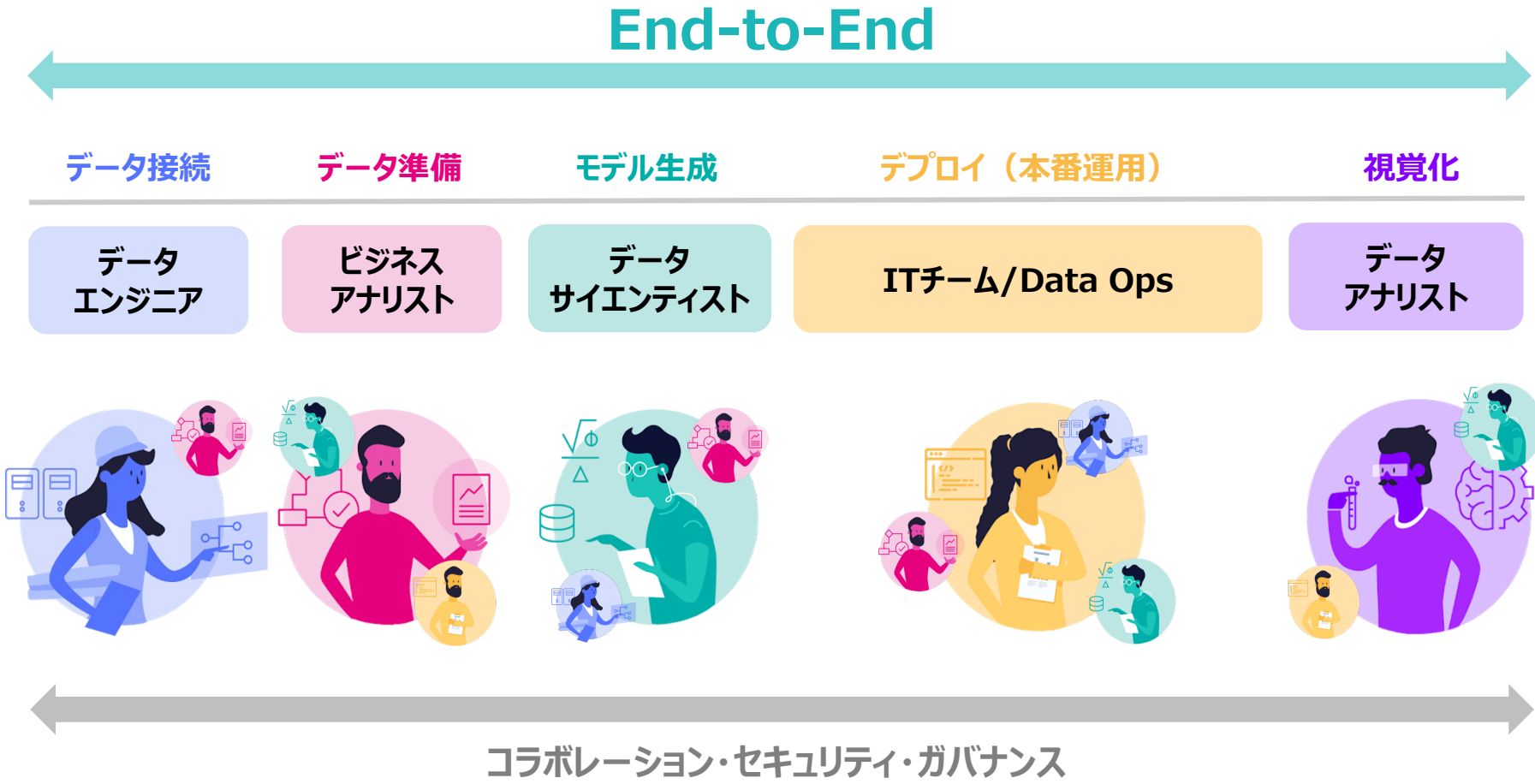
- 4億ドルを新たに調達し、  
Dataikuの評価額は46億ドル  
(2021年8月時点\*)

\*Dataiku Series E: Unleashing  
Everyday AI



# オールインワンAI・機械学習プラットフォーム「Dataiku」

AIエンジニアリング基盤の1つDataikuは、オールインワンAI・機械学習プラットフォームとして、「データ準備～モデル生成～デプロイ」を**エンドtoエンド**で**シームレスにデータを統合し、あらゆる役割、スキルレベルの人の連携を実現**いたします。



NoCodeで実装

GUIのみで分析

AutoML

業界ソリューション  
カタログ

教育コンテンツ





# オールインワンAI・機械学習プラットフォーム「Dataiku」



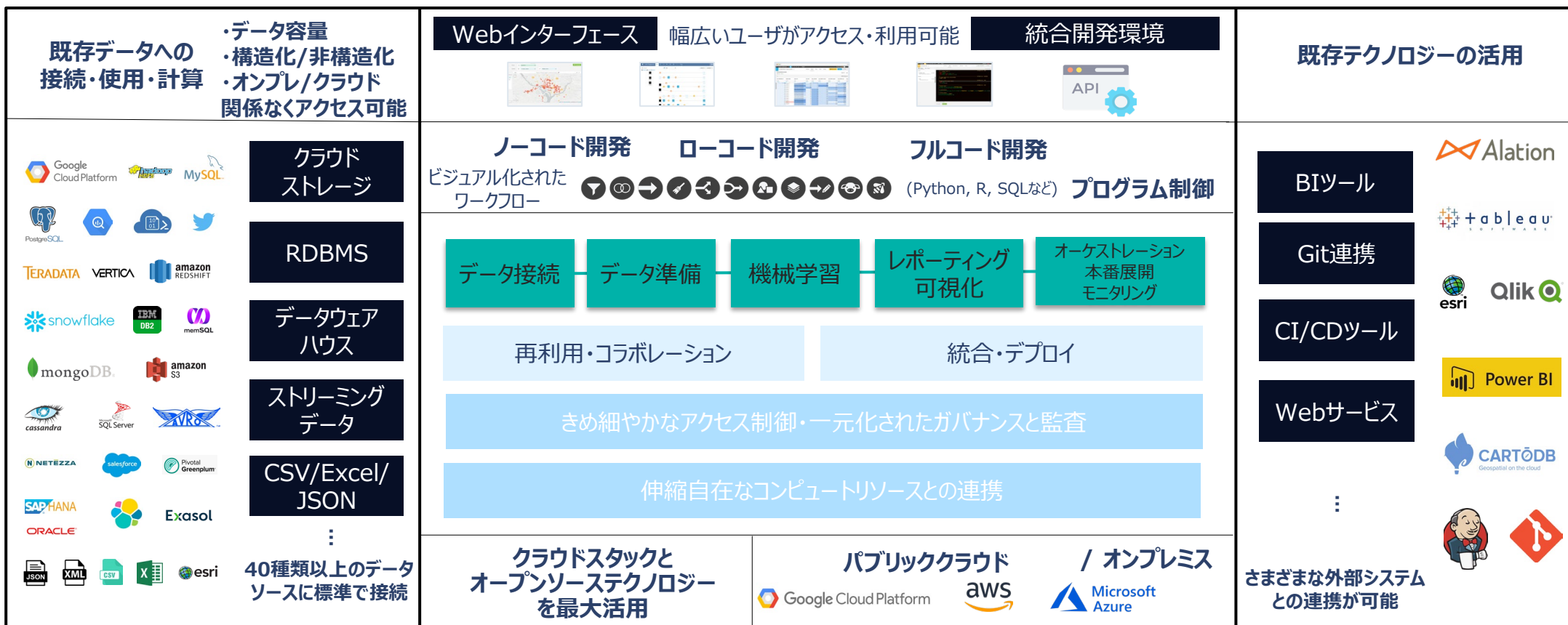
- ・データ容量
- ・構造化/非構造化
- ・オンプレ/クラウド
- 関係なくアクセス可能

# AI/機械学習プラットフォーム

-ノーコード・ローコードでAIのビジネス活用を加速する-

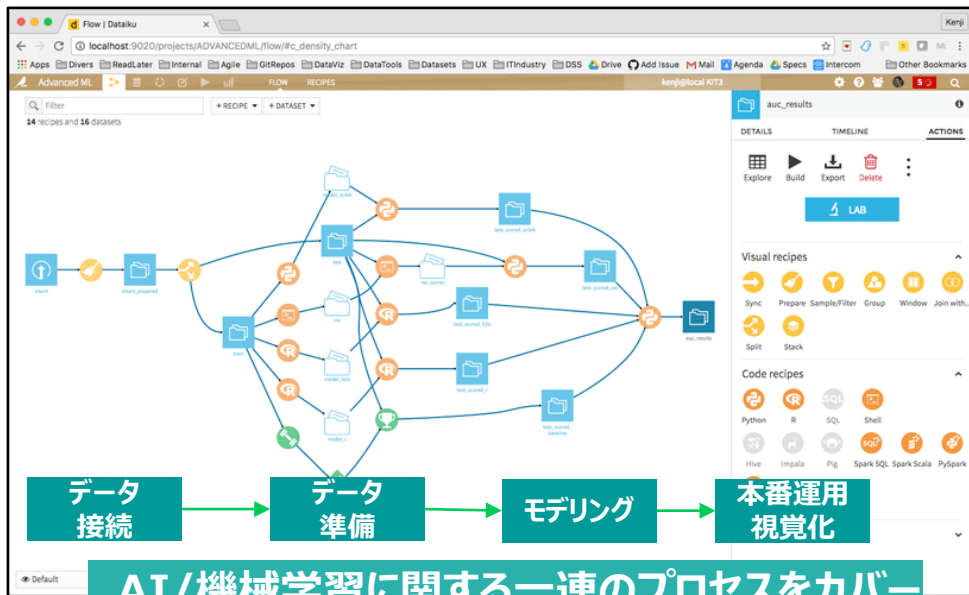


Dataiku(データイク)は、**だれもが使いやすい**エンドツーエンドのプラットフォームです。AIプロジェクトに関わる様々なユーザのコラボレーションを実現し、**生産性を向上**します。**業界・業種別のユースケース**に対応したテンプレートを使うことで、**AIを業務で活用するまでの時間を大幅に短縮**でき、より迅速に価値を提供できます。



# オールインワンAI・機械学習プラットフォーム「Dataiku」

## ■主な特長



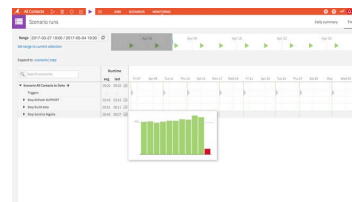
AI/機械学習に関する一連のプロセスをカバー  
インターフェースが統一されており、使いやすい

arg	Initl Mins	Initl Calls	Initl Charge	CustServ Calls	Churn?
11.010000	10.000000	3	2.700000	1	False
11.450000	Split column on False	False			False
7.320000	Replace False by ...	False			False
8.800000	Extract text like False...	False			False
8.010000	Remove rows containing text like False...	False			False
9.180000	Flag rows containing text like False...	False			False
9.070000	Remove rows containing False...	False			False
9.030000	Remove rows equal to False...	False			False
9.720000	Keep only rows equal to False...	False			False
14.000000	Filter on False...	False			True
9.400000	Filter on False...	False			False
8.020000	Show complete...	False			False
8.000000	Copy m...	False			False
8.000000	123000000	False			False
9.140000	13.000000	6	4.000000	4	False

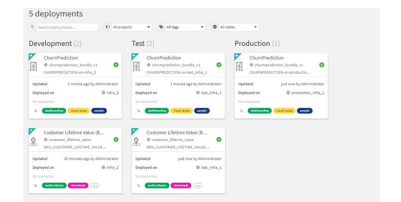
データ準備

The screenshot shows the 'Create prediction model on Churn' interface. It features three main sections: 'AutoML' (Let Dataiku create your models), 'Interpretable Models for Business Analysis' (Predict and generate more explainable models), and 'High Performance Models' (Predict and generate more accurate models). There are also sections for 'Expert' (Use your own models) and 'Deep Learning' (Use the power of deep learning).

機械学習



DataOps



ダッシュボード

### コラボレーション機能

- ・**ワークフロー・モデリングの共同作業**  
(データサイエンティスト、運用エンジニアのシームレス連携)
- ・ワークフローへのドキュメント追加
- ・統合されたGitリポジトリを介して変更をドキュメント化 (変更管理)

### データ準備

- ・**NoCode、LowCode、FullCodeでデータソースの加工が可能**
- ビジュアルレシピ：ビジュアルインターフェースで使用可能な、事前パッケージングされた複数のオペレーションを通して、入力データセットをインタラクティブに迅速に変換
- コードレシピ：コーディング技能のあるユーザーが、ビジュアルレシピ機能を超えて、サポートされる任意の言語（`SQL、Python、R など）を使用して、変換をコントロール

### 機械学習

- ・**最新の機械学習ライブラリ** (scikit-learn、MLlib、XGBoost、TensorFlow、Keras など) を搭載した、**設定可能な AutoML を活用**
- ・**Python**や**R**で**コードを直接カスタマイズ**し、高度なカスタム機械学習も利用可能
- ・上級レベルのデータサイエンティストは、よく使用するコードノートブックやJupyter、PyCharm、RStudio、VSCodeなどの統合開発環境 (IDE) を使用できる。

### セキュリティガバナンス

**すべてのデータの変更履歴や変更過程、依存関係の追跡、データ・プロジェクトに対するきめ細かなアクセス制御、監査**など企業が求めるセキュリティとガバナンスを確立することができます。これにより、**組織全体でのデータ活用を安全に加速**していくことができます。

### その他

- ・**40種類以上のデータに対するコネクタ**
- ・KubernetesをベースとしたAPI化によるモデルのデプロイ
- ・分析エンジン (R、Python、DB、Hadoop/Sparkなど) へのプッシュにより、ローカル、inDB、分散環境での分析実行
- ・**構築したモデルの根拠などを自動でレポート**

# 業界ソリューションカタログ

業界ソリューションカタログは過去の実績を元に作成したユースケースパッケージ化プロジェクトです。

ユーザーは業界ソリューションカタログを使用することで開発時間を大幅に短縮し、Dataikuの価値をより早く提供します。

## 業界ソリューションカタログ (ユースケースのパッケージ化)



Spin up your DSS App



Visualize & leverage

### 特定の業界と業種に重点

☑複雑な問題の解決策のインスピレーションも得ることが可能

### ダッシュボードを含むビジュアルツールをDSSアプリに組み込み

### データを当て込み、ワンクリックスタート

☑すぐに使用可能なオブジェクトとワークフローを提供

☑カスタマイズ可能

☑拡張可能



小売  
CPG

**分布空間フットプリント**  
流通ネットワークで影響力のある意思決定をサポート  
地理空間分析

**マーケットバスケット分析**  
売上を最大化するために購入間の関係を分析  
分析

**RFMで強化された顧客生涯価値**  
顧客の価値を評価して、販売戦略を改善  
機械学習

**需要予測**  
顧客の需要を予測し、供給と価格戦略に役立てる  
機械学習

**テリトリー  
balancingデザイン**  
需要を理解し、機械学習で調整された予測を作成  
機械学習



金融サービス  
保険

**アクチュアリー向け  
GLMツールキット**  
他のMLモデルとの比較およびGLMの開発を簡便化  
機械学習

**ESG分析のための  
ドキュメントインテリジェンス**  
ドキュメントをスキャンし、ESG分析のための洞察提供  
NLP; OCR;感情分析

**ニュースベースの  
株式アラートシステム**  
株価の監視、ニュースを活用した異常予測  
機械学習

**AIで強化された  
AMLと詐欺**  
ルールベース・MLアラート、グラフ調査による対策強化  
グラフ分析;機械学習

**不動産価格  
設定エンジン**  
ルールベース・MLアラート、グラフ調査によるセットアップ  
グラフ分析;機械学習



ヘルスケア  
医薬品

**ドラッグリポジショニング  
タイミング特定**  
オープンデータを活用し、ドラッグリポジショニングの機会を迅速に特定  
グラフ分析

**オムニチャネル  
マーケティング**  
営業活動の推進要因を理解し、影響力を向上  
機械学習

**患者コホートの選択**  
臨床検査の影響を改善するため、患者選択戦略を改善  
機械学習



製造  
エネルギー

**監視と監視の改善**  
複数のデータソースを統合し、石油精製のcockpit構築  
機械学習

**予知保全**  
保守活動組織を最適化し、コストを削減  
機械学習

**外れ値の検出**  
マシンを監視、異常な動作を予測し再調整  
機械学習

- エネルギー消費予測
- ESG投資管理
- 製品の推奨事項
- 製品感情分析装置
- 顧客セグメンテーション
- 特許モニタリング
- クレジットリスク
- 販売地域の最適化
- CO2排出量の最適化
- 保険価格設定エンジン
- P&L予測

■ 需要予測 事例

- ・ベトナムのオンラインショップ (約4時間で導入し、自社データを活用)
- ・日本の某製造業

More...

# Dataiku Cloud (SaaS版) 概要

データサイエンス／機械学習のフルマネージド・AIプラットフォームサービスとして、お客様のAI活用、アナリティクスによる洞察を加速させます。

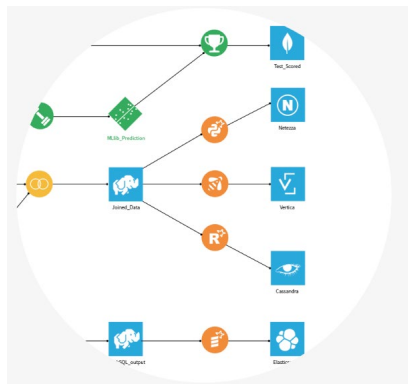
## 最新のクラウドデータスタックに対応

Snowflake、Amazon Redshift、Google BigQueryなどのビルトインのデータコネクタや事前に構築された統合のための機能を利用し、最新のクラウドプラットフォームのアジリィーを活用することができます。



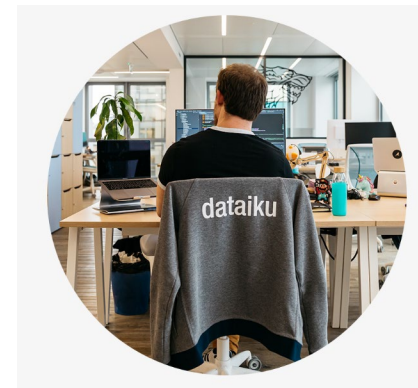
## 必要ものがすべて1つの場所に

お客様のデータを接続し、使いやすい視覚化されたデータ準備の機能や自動機械学習 (AutoML)、ビジネスのための統合レポートや視覚化ダッシュボードなどが利用できます。



## 管理とアップグレードはDataikuが対応

サーバーや管理、アップグレードなどについて、Dataiku社がすべてに対応するため、ビジネスのためのアナリティクスやAIにフォーカスすることができます。





# 教育コンテンツ

# Dataiku教育コンテンツ

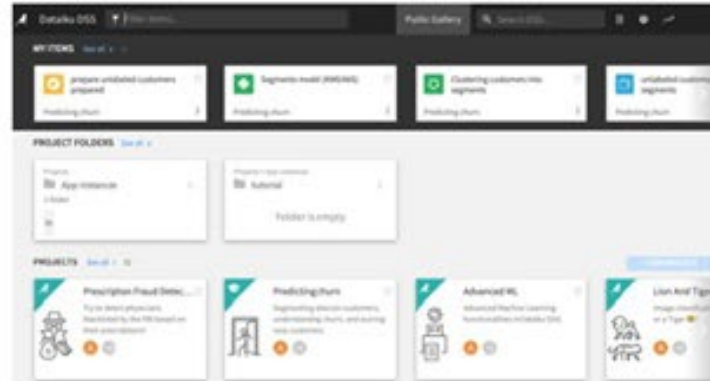
Dataikuでは、「AI人員を教育・育成する」ための3つの教育コンテンツを用意しています。

## Dataiku Academy



- ロール別のチュートリアル
- 各自のスキルにあった学習
- 学習目標を設定可能

## Dataiku Gallery



- サンプルのプロジェクト、ダッシュボード、アプリなどが多数準備
- Dataikuの操作
- AI・機械学習の流れを理解・学習

## 過去のウェビナー



- 過去に実施したウェビナーを動画配信
- インストラクターによるデモなどが多数

# Dataiku Academy

2022年4月現在



## 共通 (ビジネスユーザー)

日本語

### クイックスタート： 業務担当者

Dataiku DSSプロジェクトを作成し、GUIのみの分析法を学ぶ

データ準備・データ加工・AI構築

日本語

### クイックスタート： AI利用者

データの深掘りを行い、データの視覚化と他の担当者とのコラボレーション方法を学ぶ

AI構築

日本語

### クイックスタート： エクセルからDataikuへ

エクセルから移行し、優れた機能の活用を学ぶ

データ準備・データ加工

英語

### Dataiku DSS - Project Walkthrough

DSSプロジェクトを通じてAIのライフサイクルを巡る

データ準備・データ加工・AI構築



## データ エンジニア

日本語

### クイックスタート： データエンジニア

データ接続・コーディングとビジュアルレシピを使ってデータを成形する方法を学ぶ

データ準備

英語

### Core Concepts

ビジュアルレシピをカバーする短いビデオレッスン、プロジェクトを構築するハンズオン、クイズ問題を学ぶ

データ準備・データ加工

英語

### Advanced Data Preparation

高度なPrepare(準備)レシピとフローの使用法、イベントデータの生成法を学ぶ

データ準備・データ加工

英語

### Advanced Code

コードを記述、探索、実行、共有、およびデバッグする方法やプロジェクト変数の概念を学ぶ

データ準備・データ加工・AI構築



## データ サイエンティスト

日本語

### クイックスタート： データサイエンティスト

データサイエンティスト向けの機能を活用し、既存のDataiku DSSプロジェクトを向上させる

データ加工

日本語

### ML プラクティショナー

統計分析の概要について学ぶとともに、探索的データ分析の実行方法について学ぶ

データ準備・AI構築

英語

### Advanced Analytics

時系列、テキスト分析の為にツールやプラグインの使用法、概念を学ぶ

データ準備・データ加工

英語

### Advanced Code

ビジュアルMLインターフェースでカスタムMLモデルを作成して使用する方法を学ぶ

AI構築

英語

### Visual Machine Learning

機械学習の基本的な概念、モデルの比較、トレーニングについて学ぶ

AI構築



## 運用エンジニア

日本語

### ML プラクティショナー

モデルの作成、使用方法、新しいデータに予測モデルをデプロイする方法の概要を学ぶ

AI構築

英語

### Advanced Analytics

モデルの構築、プラグインの使用法について学ぶ(主に分析フローの形式化)

AI構築

英語

### Visual Machine Learning

機械学習の基本的な概念、モデルの運用方法について学ぶ

AI構築

英語

### Operationalization

実環境にモデルをデプロイし運用する方法を学ぶ

モデルデプロイ

英語

### MLOps

AIライフサイクルにおけるMLOpsの主要なコンセプトやフローを自動化する方法を学ぶ

パイプライン構築



# Dataiku Version 12.5- 日本語UIの提供を開始

<https://www.dataiku.com/ja/プレスリリース/dataiku-starts-the-japanese-localization-of-its-everyday-ai-platform/>

## Dataiku、Everyday AIのためのプラットフォームの日本語化を開始

企業のデータとAIの活用を阻む壁をまたひとつ取り除き、日本の企業にデータとAIの活用を楽しむカルチャーと仕組みの導入を推進

Dataiku Japan株式会社 2024年2月1日 10時00分



Everyday AI プラットフォームの

日本語化を開始



The screenshot displays the Dataiku interface in Japanese. The main window shows a script named 'compute\_customer\_prepared' with 2 steps. The 'Script Output' section displays a table with 10,000 rows. Below the table, the 'Processor Library' is visible, listing various data processing actions such as 'Filter', 'Rename columns', and 'Fuzzy join with other dataset (memory-based)'. The 'Fuzzy join' processor is highlighted, showing its description and usage examples in Japanese.

顧客ID	生年月日_parsed	年齢	性別	IP_continent	IP_geopoint
string Text	date Date	bigint Integer	string Gender	string Text	string GeoPoint
ID000003	1982-08-23T00:00:00.000Z	41	F	North America	POINT(-94.57
ID000007	1981-09-13T00:00:00.000Z	42	F	North America	POINT(-97.82
ID000011	1987-05-22T00:00:00.000Z	36	F	Africa	POINT(24 -29)
ID000015	1990-01-25T00:00:00.000Z	34	F	North America	POINT(-84.68
ID000019	1987-12-22T00:00:00.000Z	36	M	Europe	POINT(-3.002
ID000023	1956-07-24T00:00:00.000Z	67	F	Asia	POINT(113.72
ID000027	1960-09-25T00:00:00.000Z	63	F	Asia	POINT(28.951
ID000031	1980-05-31T00:00:00.000Z	43	M	North America	POINT(-82.51

**NTT DATA**