

CLouDERA

eBook

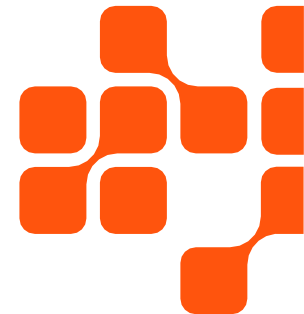
データガバナンスの 導入を成功させる 方法

最近、データの扱いに
限界が来たと感じませんか。

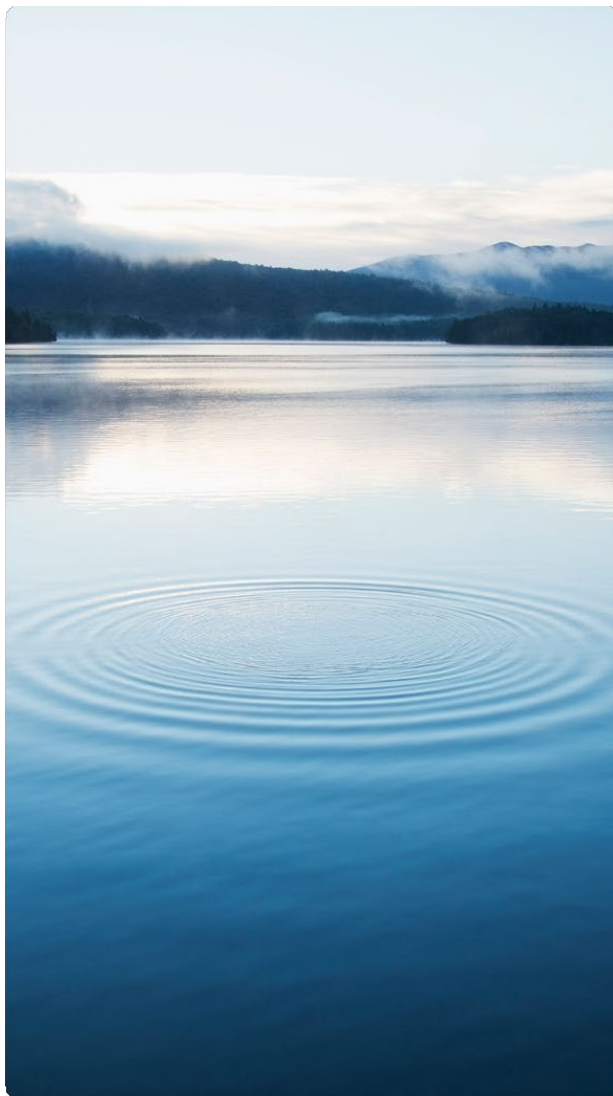


目次

データガバナンスソリューションの導入を成功させる方法	3	データガバナンスフレームワークの構築	10
ガバナンスの全体像を見据えた計画作り	4	データとその利用に関するポリシーを策定	10
的を絞って小さくスタート	4	データの命名と保存に関する基準と手順を確立	11
データガバナンス導入のリーダーを選定	4	データの利用と共有に関するルールを定義	12
このすべてが必要な理由	5	前述のルールと手順を実施するための役割と責任を割り当て	12
データガバナンスマネージャーを部門横断型のチームでサポート	5	導入	13
成果を測定する準備	6	あらゆる経験から学ぶ	13
適切な支援ツールとソリューションを選択	7	データガバナンスを成功させましょう！	14
データを深く掘り下げる	9		
データの動きを把握	9		
データの使われ方をデータリネージから把握	9		
最初のデータクレンジングを実施	9		



データガバナンスソリューションの導入を成功させる方法



最近、データの扱いに限界が来たと感じませんか。
必要なときに見当たらず、ごく簡単な指示にも従わず、そこから得られる情報の遅れが心配の種を増やしています。

そろそろ、ルールを決めなくてはなりません。

そこで登場するのが、データガバナンスです。

データガバナンスとは、組織の中でデータ管理を統制するためのルールや役割を指す言葉です。こうしたルールや役割は、データの利用を（セキュリティ、アクセス、コンプライアンスなどの面で）適切に、かつ（品質や運用性の面で）効果的にすることを目的としています。

データガバナンスは大きな見返りをもたらしますが、その導入には多大な労力が必要です。データガバナンスプロジェクトにはさまざまな要素があり、そのどれか1つに問題が起きただけでも、プロジェクトが脱線してしまうおそれがあります。本ガイドでは、データガバナンスの導入プロセスを説明するとともに、導入を成功させるために検討すべきポイントや避けるべき落とし穴について解説します。

ガバナンスの全体像を見据えた計画作り

データガバナンスの成否は、その大部分が計画段階で決まります。プロジェクトの計画がしっかりしていれば、実行段階で多少のつまずきがあったとしても、おそらく十分な成果が得られるでしょう。しかし、計画作りを避けたり手を抜いたりすれば、どれほど完璧に実行したとしても、期待通りの結果は得られません。

少し脅かしてしまったかもしれませんが、ここからは実践的な計画作りについてアドバイスしていきましょう。

的を絞って小さくスタート

初期のデータガバナンス導入は、パイロットプロジェクトとして扱います。

1つの業務部門、1つの部署、あるいは部署内の特定のセクションなどにデータセットを限定しましょう。導入単位が小さいほど導入がスムーズになります。

ただし、その単位内のデータが有意なもので、**関連性が高く、実際に活用されている**ことが条件です。また、以下の人物に関わってもらう必要があります。

- 1名以上のITインフラストラクチャー担当者
- 1名以上のデータチームメンバー
- 1名以上の該当分野の専門家
- 2名以上のビジネスユーザー

(1人の人物が上記の役割のいくつかを兼任していてもかまいません)

データガバナンス導入のリーダーを選定

次の職場の飲み会で、(実際の職場ではあり得ないかもしれませんが)その場の勢いに任せて標語を作るように言われたら次の一節をそのまま借用してください。

- 成功するのに
- 必要なのは
- リーダーだ。

美しい言葉だと思いませんか。そして、これは真実でもあります。

自社でデータガバナンスの導入を始める前に、ガバナンスを主導するのにふさわしい人物が誰なのか、じっくり検討してみてください。適切なリーダーには、以下の資質が求められます。

- データガバナンスの導入に積極的で、自社でガバナンスの導入を成功させる方法について学び、考えることを惜しまず取り組む。
- カリスマ性と優れたコミュニケーション能力のどちらか、または両方を備えている。
- 社内のさまざまな従業員やチームと良好な協力関係を築いている。



このすべてが必要な理由

当然ながら、技術部門や業務部門をはじめとする従業員全員がデータガバナンスの導入を求めているのなら、話は別です。しかし、そうでないのであれば、リーダーとなる人物には、データガバナンスの可能性を信じ、導入を技術面から指導できる能力だけでなく、データガバナンスの価値を従業員に効果的にアピールし、売り込む能力が求められます。

データガバナンスとは、従業員にとってビジネスプロセスに遅れを生じさせ、必要なデータへのアクセスを妨げる煩わしいポリシーだというように受け取られがちです。したがって、データガバナンス導入のリーダーは、従業員に対し、その認識とはまったく逆であり、データガバナンスがデータアクセスを**改善し、簡素化する**ものであることを広く納得してもらう必要があります。

そうでなければデータガバナンス戦略は支持や協力を得られず、失敗することになるでしょう。

データガバナンスマネージャーを部門横断型のチームでサポート

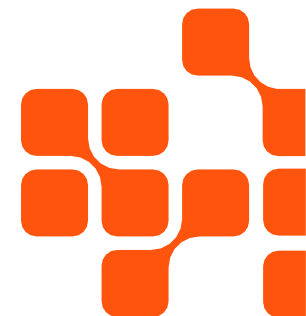
データガバナンスプロジェクトのリーダーが決まりました。

次に、導入プロセスの主要な段階でリーダーをサポートするチームを編成してみましょう。チームの具体的な構成は状況によって異なりますが、一般的には以下のメンバーが必要になります。

- プロジェクトで必要となるデータ管理ソリューションの導入作業を率いることのできる、技術面で高度な導入スキルを備えた人物
- プロジェクトで使用するデータセット(またはそのデータセットの一部)に関する専門知識を持った人物

アドバイス

最初にプロジェクトの対象となる領域や部門を決め、その中から最適なチームメンバーを選ぶのが一般的ですが、逆に、優れた人材を確保できるという点からプロジェクトを実施する領域や部門を決定する方が合理的な場合もあります。



成果を測定する準備

こんな話を聞いたことはあるでしょうか。ある有名な弓の射手が、なぜいつもの中心を射抜けるのかと尋ねられて、次のように答えました。「簡単さ。最初に矢を放ってから、矢の周りに円を描くんだ！」

データガバナンスプロジェクトの「成功」の基準を、矢を放ってから円を描くように後から定義してはなりません。データガバナンスの導入が本当に目的を達成しているかを判断するには、計画段階で目的を定義する必要があります。

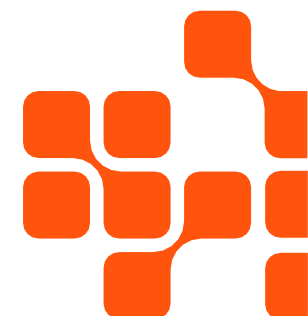
考えられる目標には、以下のようなものがあります。

- データの精度の向上
- データのアクセス性の向上
- データ管理の一元化
- 組織全体でのデータ共有の促進
- データに関するコミュニケーションの改善

ただし、目標を設定して終わりではありません。目標は重要ですが、たいていは曖昧で、測定が難しいものです。そこで次のステップとして、それぞれの目標にKPIを設定します。具体例として、米運輸省のデータガバナンスフレームワークの例を見てみましょう。

No.	目標	パフォーマンス測定
1	ニーズに基づいて定められた精度の基準をデータが満たしている	基準を満たしているデータの割合
2	データの意味を容易に理解できる	定義が明確であるデータ要素の割合
3	データに関する変更がデータのユーザー全員に伝わる	影響を受けるユーザーが、変更が加えられる30日前までには、その変更について認識している

設定するKPIがどのようなものでも、データガバナンスの導入を始める前にそれを測定して文書化し、それを基準として将来の測定値と比較できるようにします。



適切な支援ツールとソリューションを選択

データガバナンスの導入を成功させる上で極めて重要になるのが、適切なツールです。ただし、ご注意ください。どれほどソリューションの機能が豊富でも、それだけでデータガバナンスプロジェクト全体を管理することはできません。

適切なソリューションを見つけるには、データガバナンスの導入に関わる領域を特定し、領域ごとに要件をまとめることが鍵となります。次に、候補のツールを調査した上で、データガバナンスの領域を網羅し、かつスムーズな連携が可能な2~3つのソリューションを選びます。

関連する領域には、以下のようなものがあります。



データの品質 (最優先事項)

データの品質の維持は、データガバナンスにおける重要な目標の1つです。経営陣やビジネスユーザーによる重要な意思決定は、正確な情報に基づいて行う必要があります。

データ品質の維持にはさまざまな側面があります。例えば、データプロファイリング、データパイプライン監視、元のデータソースの評価、システム間を移動する過程での変換の評価(通常はデータリネージで対応します)、データクレンジングなどです。



ワークフロー/アクセス

データガバナンスの主な目的は、「ユーザーがアクセスすべき情報はアクセスを可能にし、アクセスすべきでない情報はアクセス不可にする」という一文に集約されます。

データガバナンスにおけるデータワークフロー/アクセスの領域は、この文章の前半部分、すなわち「ユーザーがアクセスすべき情報はアクセス可能にする」ことに対応します。

ユーザーが業務に必要なデータへのアクセスを求めた際は、最も包括的かつ正確な最新データがすぐに利用できる状態になっていなければなりません。手間のかかる権限設定や承認プロセスにユーザーの手をわずらわせてはなりません。

業務の流れをスムーズでストレスなく、効果的にするのが、理想的なワークフローソリューションなのです。



セキュリティ

データガバナンスにおけるセキュリティの領域は、先に述べたデータガバナンスの目的の後半部分、すなわち「ユーザーがアクセスすべきでない情報はアクセス不可にする」ことに対応します。

最初のデータガバナンス導入の規模が小さくても、将来的に大量のデータを扱うようになった場合を念頭に置くようにしてください。その際には、データを分類し、適切なアクセスレベルを自動で判断する仕組みが必要になるでしょう。また、アクセスレベルの確認や承認を人間が行わなければならない場合でも、少なくとも初期の作業は自動化する必要があります。



コンプライアンス/プライバシー

金融、医療、保険をはじめ、どの業界に属していても、遵守すべき規制は存在するものです。データガバナンスの大きな役割の1つは、コンプライアンスを確保するためのポリシーを策定し、コンプライアンスを遵守している事実を証明することにあります。このプロセス、すなわちポリシーを策定して維持し、そのポリシーを遵守したことを証明する作業を容易にできるソリューションを見つけましょう。

どの分野でも、機密データや個人データは細心の注意を払って取り扱う必要があります。データガバナンスソリューションには、このようなデータを容易かつ正確に特定し、必要なデータマスキングやアクセスポリシーを適用して、その実施状況を証明できる機能が求められます。



データの民主化/ユーザーエンパワーメント/データ検出

難しそうな言葉が並んでいますが、意味はシンプルに、データのユーザーが各自で必要な情報を見つけられるようにするという考え方です。

通常、このようにユーザーに力を与えるには、データカタログが不可欠です。データカタログは、データの検出や調査のための「信頼できる唯一の情報源」として機能します。データガバナンスプラットフォームにはたいいてい、データカタログが組み込まれており、データガバナンスの設定作業の一環として、ガバナンスの対象となるすべてのメタデータをそのカタログに登録します。

ただし、企業内の全データについてこのプロセスを手作業で行うと、数ヶ月はかかります(最初のプロジェクトであれば短期間で済むかもしれませんが、データガバナンスを組織全体に拡大したときのことを考える必要があります)。ここで重要になるのが自動化です。これにより、構築の段階では、組み込みの自動データリネージツールを使えば、昼休み中にデータカタログ全体を設定するといったことが可能になります。さらに、カタログのメンテナンス段階では、定期的な更新作業を通じて、最近のメタデータの変更や追加を適用してデータカタログやガバナンスプラットフォームを最新の状態に保つことが可能になります。

アドバイス

ツールを評価する際には一般的に、自動化機能に注目します。スキャン、マッピング、プロファイリング、ポリシー適用など、自動化できる機能が多いほど、チームメンバーの負担を軽減できる可能性が高まります。ただし、自動化の質はどの程度か、正確な結果がどれだけ得られているかといった点を確認するようにしましょう。



データを深く掘り下げる

これで、データガバナンスの取り組みの全体的な計画が完成しました。次は、対象となる個々のデータシステムに焦点を当て、実践的かつ具体的なデータガバナンス導入企画を策定しましょう。

データの動きを把握

10代の頃にベビーシッターをしたことはあるでしょうか。子供が起きているときと寝ているときでは、どちらの方が楽でしたか。

動き回っている対象を追跡する方が、仕事ははるかに難しくなります。そして、企業の業務システムにおいては、データは常に動き回っている存在なのです。

しかし、データがどのように動いているのかを理解すれば、追跡は簡単になります。このレポートにとって重要なデータは何でしょうか。そのデータはどこから来るのでしょうか。そのデータはどのようにしてレポートに反映されるのでしょうか。その過程でどのような変換が行われるのでしょうか。

Cloudera Octopai Data Lineage のような自動データリネージツールは、直感的な操作でデータ環境全体のデータ経路を可視化できるため、このような場面では極めて有用です。

データの動きが予測可能になれば、ガバナンスも可能になります。

データの使われ方をデータリネージから把握

これで、データがどこへ流れていくのかは理解できました。では、誰がそのデータを利用し、誰がそのデータへのアクセスを求めているのでしょうか。

データ利用状況のマッピングとデータリネージは、データガバナンスプロジェクトを成功させる上で不可欠なものです。データの種類やレベルが異なれば、そのデータを利用するユーザーの種類やレベルも異なります。例えば、規制データへのアクセスを必要とするユーザーは通常、業務データへのアクセスを必要とするユーザーより数が少なく、上位の役職に就いています。また、規制データに直接アクセスして変更を加えるようなユーザーは、規制データの閲覧のみを必要とするユーザーよりもさらに数が少なくなります。

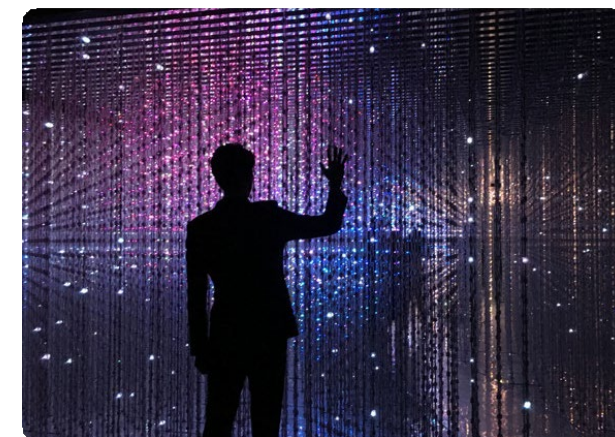
このような場面では、部門横断型のチームが真価を発揮します。なぜなら、それぞれのユーザーが自身の担当領域でどのようにデータを利用しているのかを正確に定義できるからです。

最初のデータクレンジングを実施

汚れた湖に深く潜っても良いことはありません。データを深く掘り下げる場合も同じです。もちろん、データストアを新鮮な湧き水のようにクリーンにすることは、データガバナンスの最終目標の1つと言えますが、当然ながら時間がかかります。それでも、明らかなゴミをあらかじめ取り除いておけば、データガバナンスの導入ははるかに簡単になるはずで

す。冗長なデータセットは捨ててしまいましょう。古くなったレポートも不要です。

では、何がゴミで何がゴミではないのか見極めるにはどうすればいいのでしょうか。データリネージマッピングソリューションを使用して可視化データを作成し、部門横断型のチームで確認してみてください。関連性もなくなり、破棄しても構わないデータ資産を見つけ出しましょう。こうした作業によって、データガバナンスの負担が軽減されるはずで



データガバナンスフレームワークの構築

ついにここまで来ました！すべての準備が完了した今、いよいよポリシーや基準、ルールの策定に取り掛かりましょう。

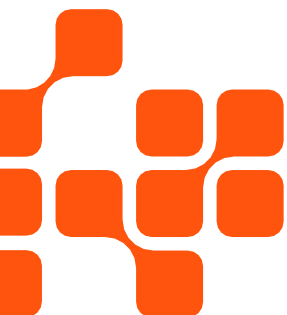
データとその利用に関するポリシーを策定

データガバナンスポリシーとは、組織内でのデータ利用がどうあるべきかを大局的な観点から定めたものです。

以下は、米運輸省のデータガバナンスフレームワークにおける代表的な例の一部です。

No.	目標	パフォーマンス測定
1	FHWA データは一貫性を保っている必要がある	すべての戦略的な FHWA データのモデリング、命名、定義付けは、組織全体で基準に従って一貫性のある形で行われなければなりません。経営陣はデータの共有を推進し、正当な理由なく冗長なデータを保持しないよう努める必要があります。さらに、データを最初に生成するビジネスチュワードは、FHWA データを必要とする下流のプロセスや業務部門の情報ニーズを把握しておく必要があります。
2	FHWA データは許容できる十分な品質でなければならない	FHWA のミッションを達成するには、質の高いデータが不可欠です。データチュワードは、FHWA データが意図された目的および用途に対して正確かつ適切であることを保障する責任を負います。データプロバイダーは、FHWA データの収集、処理、レポートに関するすべての報告要件を遵守し、データ品質法のあらゆる要件を満たすことを担当します。データ品質基準は、業務責任者によって定義された FHWA データについて承認した信頼性レベルに基づいて、能動的に管理および適用されるものとします。

ポリシーは、一般のビジネスユーザーでも理解できる言葉で記述するようにしてください。データサイエンティストしか意味を理解できないようでは、データガバナンスの導入に成功したと言えます。もう一度やり直してみましょう。



ポリシーを策定したら、そのポリシーを実際に適用するための基準、手順、ルール作りに入ります。

データの命名と保存に関する基準と手順を確立

例えば、自分が米国に住んでいて、欧州にできた親友から、「4/3/2023」に30歳になるという内容のメッセージを受け取ったとしても、4月3日にサブライズパーティーを企画してはいけません。それでは1ヶ月遅れになってしまいます。

個人レベルであれば、米国と欧州の日付表記を混同しても、ちょっとした笑い話や誤解で済むでしょう(あるいは、賞味期限を勘違いしておなかを壊すくらいでしょうか)。しかし、自社のデータベースで日付表記が一貫性を欠いていれば、深刻な影響が極めて広い範囲に及ぶおそれがあります。

データエコシステムのあらゆる要素を常にスムーズに連携させるには、データの命名や保存に関する規則を確立し、遵守しなければなりません。つまり、データガバナンスポリシーの実践的な活用方法の1つが、このような規則を確立することなのです。

このような規則や基準は通常、さまざまなデータカテゴリごとに以下のような形で定義できます。

No.	情報の特性	特性の説明
1	名前と属性	変数名と属性は、あらゆるシステムで一意でなければなりません。名前は静的な場合もあれば、システム実行時に動的に決定される場合もあります。
2	コンテナの形式	FHWAのコンテンツデータは、想定される文字種、形式、最小/最大フィールド長、およびその他すべての形式固有の特性を反映する形で、正確に文書化されなければなりません。
3	コンテンツの長さ	一意に定義されたすべての変数が、妥当なデータ長を指定する必要があります。すべてのマッピング変数は、特定のコンテンツ長に従わなければなりません。

また、これらのデータカテゴリのいずれか(「名前」など)については、さらに細かく指定することもできます。

No.	情報の特性	特性の説明
1	名前の大文字/小文字表記は統一すること	同じ文脈内では、大文字/小文字表記を統一してください(例: キャメルケース、パスカルケース、スネークケース)
2	名前には一般的な略語を使用すること	名前には、フルネームではなく一般的に使われている略語を使用してください(例: 「Identifier」ではなく「ID」を使用)
3	名前には複数形ではなく単数形を使用すること	名前の中で用語として使用される名詞は、概念自体が複数形である場合(例: 「Tool Suite Total License Cost」)を除き、単数形にしてください(例: 「Vehicle Depreciation Rate」)

(出典: 米連邦政府 Centers for Medicare and Medicaid Services の『Data Naming Quick Reference Guide』)

データの利用と共有に関するルールを定義

このフィールド/行/列/テーブル/カテゴリのデータを閲覧を誰に許可しますか。利用、変更、共有を誰に許可するのでしょうか。

また、データの閲覧、利用、共有にあたって満たすべき条件はありますか。例えば、特定の日までしか利用できない、特定の合意書への署名と記録が完了している場合のみ利用できるといった条件はあるのでしょうか。

Connected Automated Driving (CAD) のナレッジベースで作成された『Data Sharing Framework』には、データ共有の条件が極めて明確に示されています。

この推奨事項は、データを収集するプロジェクトの実施中や終了後のデータ共有をスムーズにするために、合意書に何を盛り込むかについてガイダンスを提供する目的で策定されたものです。なお、この推奨事項は自動運転のパイロットプログラムにも適用されています。データ共有という点では、以前のFOTと現在の自動運転パイロットプログラムとの間に大きな違いはありません。

01

作業内容の記述を含む資金提供合意書

02

コンソーシアム合意書

03

同意書を含む参加者との合意書

04

外部データプロバイダーとの合意書

前述のルールと手順を実施するための役割と責任を割り当て

組織内のデータ利用に関するルールと手順が明確に定義されれば、取り組みは半分終わったも同然です。しかし、残りの半分の方がしばしば難しくなりがちです。具体的には、人々にルールや手順に従ってもらい、忘れられたり無視されたりしないようにするための取り組みです。法の執行を必要としない社会がほとんど存在しない理由はここにあります。

例えば、データを共有する前に同意書が揃っていることを確認する責任は誰が負うのでしょうか。あるいは、そのような状況を追跡するデータ管理システムを構築する責任は誰にあるのでしょうか。

新たに取り込まれたデータがデータ品質基準を満たしているかどうかを確認するのは誰でしょうか。自動チェックが失敗したときの通知を誰が受け取るべきなのでしょうか。

データの利用者が必要とするデータにアクセスできない場合、その問題を解決するには誰に問い合わせればいいのでしょうか。

このような役割やその他のすべてについて、内容を明確にし、担当者を割り当てる必要があります。

もちろん、最終的な目標は、データの利用者全員がデータガバナンスのルールや手順を理解して、自然にそのルールや手順に従うようになり、強制的な措置が最小限で済むようにすることです。これを達成するための主要な方法の1つが、データガバナンスの手順を日常的なデータ利用のルーチンに組み込むことです。ほとんどの人が、「食器を使ったら洗う」とか、「家に入ったら靴を脱ぐ」といった生活上のルーチンを習慣化しているように、データガバナンスを行う組織にとっては、「データ資産を作成したらカタログに登録する」といったルーチンを定着させることが理想です。

導入

ポリシーの整備と手順の計画は完了しました。理論はここまでにして、次はいよいよ実践の段階です。

とはいえ、データガバナンスプロジェクトを構想段階から実用段階へと進めるにあたっては、心構えと実務の両面で以下の準備が整っていることを確認する必要があります。

あらゆる経験から学ぶ

このデータガバナンスプロジェクトを成功させるために多大な労力を費やしてきましたが、何らか

の問題が起きることも想定しておく必要があります。そして問題の発生は悪いことではありません。将来に向けた学びの機会となります。

データガバナンスの導入にあたり、何かうまくいかないことが起きた場合は、すべての関係者(特にデータガバナンスプロジェクトのコーディネーター)がその問題に注意を向け、「何が問題だったのか」、「次回に向けてどう修正すればいいのか」と問いかける必要があります。

プロジェクトの規模が小さく、対象が適切に限定されている場合は、現実的にあらゆる不測の事態

から学び、(可能な範囲で)完璧な状態に近づけることができます。そうすれば、対象を他の事業領域に広げる際にも、データガバナンス導入の仕組み作りの段階ですでに大きな失敗のほとんどを経験し、分析と修正を完了していることになるため、導入プロセスをはるかにスムーズかつ効率的に進められます。



データガバナンスを成功させましょう！



自社のデータがどこにあるか把握しているでしょうか。

データガバナンスの導入が完了していれば、その答えを知るだけでなく、はるかに多くのことが可能になります。

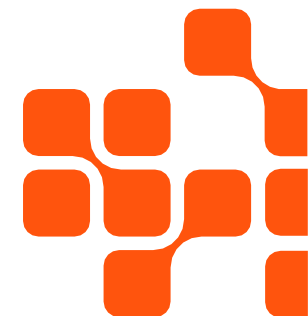
例えば、ミッションクリティカルな意思決定に利用できる高品質のデータを入手できます。

また、必要なときに必要なデータにシームレスにアクセスできる権限をデータのユーザーに付与できます。

どのような監査にも容易に対応し、合格できる安全かつコンプライアンスに準拠したデータシステムも利用できます。

Cloudera Octopai Data Lineage のデータインテリジェンスプラットフォームは、データガバナンスの導入と維持において、あらゆる組織に価値ある支援を提供します。エンドツーエンドの自動データリネージツールとして、データ環境全体を背景として、あらゆるレガシーシステムやクラウドベースのデータシステムに接続してマッピングし、視覚化することができます。また、データカタログの自動作成と自動更新により、組織内のデータ作成者とユーザーの全員が共通認識を持てるようになり、データに関するコミュニケーションやコラボレーションのための組み込みのプラットフォームが提供されます。

データ管理ツールを利用しましょう。
これで、データガバナンスに取り組む準備は万全です。



Cloudera について

Cloudera は、データ、分析、AI を活用できる唯一の本格的なハイブリッドプラットフォームです。クラウドのみを提供する他のベンダーと比べて100倍以上のデータを管理できるため、グローバル企業はどのようなパブリッククラウドやプライベートクラウド上でも、あらゆる種類のデータを価値ある信頼性の高い洞察に変換できます。また、Cloudera のオープンデータレイクハウスは、拡張性と安全性の高いデータ管理とポータブルなクラウドネイティブ分析を提供し、お客様がプライバシーを維持し、責任ある信頼性の高い方法でAIを導入しながら、生成AIモデルをデータに適用できるように支援します。データを活用して不可能を可能にする能力を、現在そして将来にわたって提供する Cloudera は、金融サービス、保険、メディア、製造業界の大手グローバル企業や政府機関から信頼されています。

詳しくは、jp.cloudera.com をご覧頂くか、[LinkedIn](#) または [X](#) で Cloudera をフォローしてください。Cloudera および関連するマークは、Cloudera, Inc. の商標または登録商標です。その他すべての会社名および製品名は、それぞれの所有者の商標です。

CLouDERA

Cloudera 株式会社 | お問い合わせ先: info-jp@cloudera.com | jp.cloudera.com

© 2025 Cloudera, Inc. All rights reserved. Cloudera および Cloudera ロゴは、米国その他の国における Cloudera Inc. の商標または登録商標です。その他のすべての商標は、それぞれの企業に帰属します。本書の内容は予告なく変更されることがあります。7912-0012025年8月08日