

Databricks委託のForrester Total
Economic Impact™調査
2020年4月

Databricks 統合デー タ分析プラットフォームの Total Economic Impact™

Databricks プラットフォームが可能にする
コスト削減およびビジネス上の利益

目次

エグゼクティブサマリー	1
主な調査結果	1
TEIのフレームワークと調査手法	4
Databricks 統合データ分析プラットフォームのカスタマージャーニー	5
インタビューを受けた組織	5
主な課題	5
ソリューションの要件	7
主な結果	7
複合組織	9
利益の分析	11
増分利益	11
運営効率の向上	13
レガシーデータ分析プラットフォームのコスト削減	16
非定量化利益	18
柔軟性	19
コストの分析	20
Databricks プラットフォーム、トレーニング、 およびストレージのコスト	20
Databricks の管理コスト	21
Databricks のトレーニングコスト	22
財務状況の概要	24
Databricks 統合データ分析プラットフォーム：概観	25
付録 A：Total Economic Impact	27
付録 B：脚注	28

プロジェクトディレクター：
Edgar Casildo

FORRESTER CONSULTING について

Forrester Consulting は組織のリーダーがその組織を成功に導けるよう、独自の客観的調査に基づくコンサルティングを提供しています。お客様は、各事業課題に最適な専門知識と経験を有するリサーチアナリストを通じて、短期的な戦略に関するコンサルティングからカスタムメイドのプロジェクトにわたり、Forrester Consulting のサービスを利用することができます。詳細については、forrester.com/consulting をご覧ください。

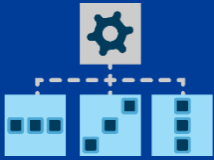
© 2020, Forrester Research, Inc. 不許複製・禁無断転載。本書を無断で複製することは固く禁じられています。本書の内容は、最適な情報源に基づいています。ここに記した見解はその時点でのものであり、最新の情報とは異なる場合があります。Forrester®、Technographics®、Forrester Wave、RoleView、TechRadar、および Total Economic Impact は Forrester Research, Inc. の商標です。その他の商標の所有権は各社に帰属します。詳細については、forrester.com をご覧ください。

エグゼクティブサマリー

利益



増収：
\$9,895,520



運営効率の向上：
\$7,677,643



レガシーデータ分析プラットフォームのコスト削減：
\$11,298,202

Forresterは、「競争の激しい今日のビジネス環境では、すべての機会においてデータ、ビジネス分析、および機械学習を利用かつ応用することによって製品と顧客体験の差別化を図ることが、即座に成功の要件となりつつある。」と報告しています。ⁱ それ以前のインターネットと同様に、AIと機械学習 (ML) は、今日の企業に対する最大の脅威と最高の機会の両方を象徴しています。ⁱⁱ 組織が成功するためには、製品、サービス、および業務にAIとMLを活用することが必要となるでしょう。

成功を推し進める会社には、ビジネス上の最優先事項として分析とMLへの投資を効果的に関係させることが求められます。同様に、データマネジメント戦略とビジネス上の優先事項を関係させると共に、データチームによるそのような優先事項の実行に対する制約や障害を取り除いて成功を確実にすることも必要です。

組織はデータ（構造化および非構造化データの両方）の大衆化を行い、プロセスを進化させ、チームを再構成し、組織文化に挑み、テクノロジースタックを再設計することで、データチームの成功を支持しなければなりません。組織がリアルタイムで決定するには、最新データへのアクセスが必要になります。さらに、データは組織内の様々なツールやシステムで開いたりアクセスしたりすることが可能で、他のテクノロジーまたはチームがアクセスできない閉鎖的なフォーマットや専用システム内に隔離されていないことが必要です。データ・エンジニア、及びアナリストがデータサイエンス関連の仕事ができるようにトレーニングして権限を付与することで、データサイエンティストのみを頼ることから生じるボトルネックを解消した生産性向上が見込めます。

これらの要件を満たすことは複雑でコストがかかるため、今日のほとんどの組織にとってこれらすべてを大規模に達成することは不可能です。これが、Databricks による開かれた統合データプラットフォームを組織が急速に採用している理由であり、それによって大規模なデータエンジニアリング、コラボレーション型データサイエンス、フルライフサイクルの機械学習、およびビジネス分析のためのデータと AI が単純化されます。

Databricks は Forrester Consulting に、Databricks の現在の顧客に対する Total Economic Impact™ (TEI) 調査の実施と、企業が Databricks 統合データ分析プラットフォームを採用することで認識する投資利益率 (ROI) の決定とを委託しました。本調査の目的は、Databricks プラットフォームによって企業にもたらされる可能性のある経済的影響を評価するためのフレームワークを読者に提供することです。

Databricksの導入前、顧客はオンプレミスおよびクラウド環境にわたる様々なデータプロセッシング、ビジネス分析、およびMLテクノロジーを所有していました。サイロ化されたデータ、使用が困難なシステム、運用上の摩擦、および拡張性の不足は、効果的に協力してデータからビジネスの価値を明らかにするというデータチームの能力を妨げていました。データチームは、彼ら自身が、ほとんどの時間をデータ主導の開発とビジネスの成果よりもシステムの設定と管理に費やしていることに気付いていました。

主な調査結果

定量的な利益。 以下のリスク調整後の現在価値 (PV) で示される定量的利益は、インタビューを受けた企業の経験を表しています。



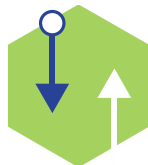
ROI
417%



利益 PV
\$2,890 万



NPV
\$2,330 万



回収期間
<6 カ月

- ▶ **5%の増収。**Databricks は、インタビューを受けた顧客のデータサイエンティストが、ML モデルの作成と改良にもっと多くの時間を使えるようにしました。Databricks はまた、以前には利用が難しかったディープラーニングモデルのような最先端の ML モデルをデータサイエンティストが使用できるようにしました。さらに、データへのアクセスの大衆化によって、インタビューを受けた組織は、新しいユーザーが新しい ML モデルの多様なセットを作成し、以前よりも深い洞察を導き出すのを目にしました。さらに多くの、そしてより良い ML モデルの組み合わせによって、インタビューを受けた多くの対象者の増収が促進されました。
- ▶ **データチームの生産性がそれぞれ 25%および 20%改善。**Databricks は、データサイエンティストとデータエンジニアのチームが、データの検索とクリーニング、抽出、変換、ロード (ETL) パイプラインの作成と維持に費やす時間を減らし、ビジネスの成果を推進する ML モデルの構築と改良に多くの時間を使えるようにしました。Databricks はまた、アナリスト、データサイエンティスト、およびエンジニアの間での協力を制限する技術的な障壁を取り除き、データチームがもっと効果的に協働できるよう手助けしました。
- ▶ **オンプレミスでのインフラの廃止により、年間数百万ドルを節約。**Databricks への移行により、インタビューを受けた組織は、オンプレミスでのインフラを廃止して現在の過剰なデータ分析ライセンスをキャンセルすることが可能になりました。

非定量的利益。インタビューを受けた組織は、この調査では定量化されていない以下のような利益を得ました。

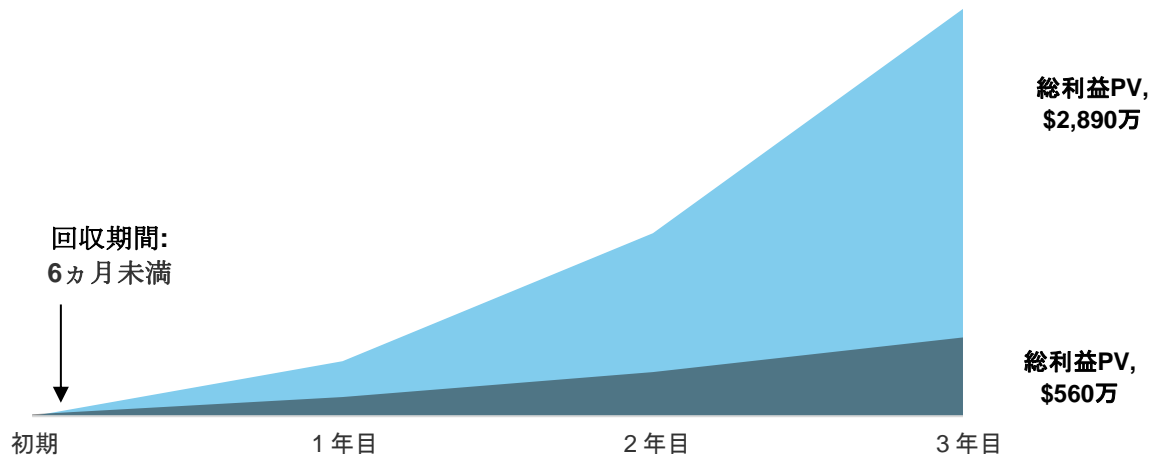
- ▶ **運用コストの削減。**Databricks が提供する様々な特徴は、インタビューを受けた組織の原材料への支出を減らし、製造上の問題の早期同定を可能にしたことから、保証修理費と返却率が低下しました。
- ▶ **セキュリティの向上。**Databricks の導入前、インタビューを受けた組織には、そのデータ分析環境全体にわたるセキュリティ標準がありませんでした。Databricks が提供したネイティブセキュリティ保護により、セキュリティに関する懸念は緩和され、インタビューを受けた顧客は他の活動のリソースに集中できるようになりました。

コスト。インタビューを受けた組織では、以下のようなリスク調整後の現在価値 (PV) コストが発生しました。

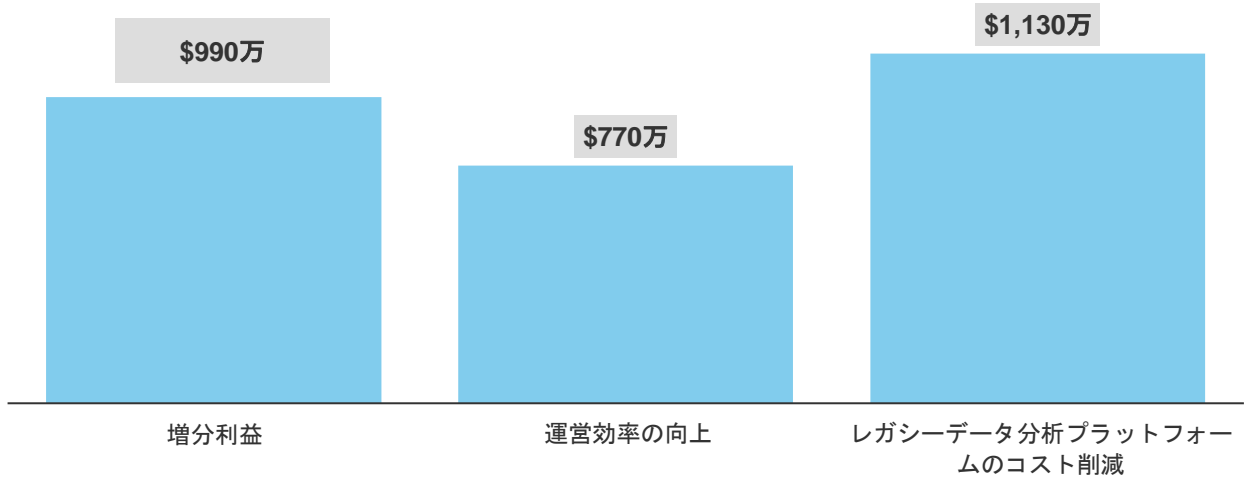
- ▶ **Databricks 統合データ分析プラットフォーム、トレーニング、およびストレージのコストは\$450 万。**これらには、ユーザーが Databricks に支払う年間の使用料とトレーニング料に加え、サードパーティーのクラウドストレージに対するコストが含まれています。
- ▶ **Databricks の管理コストは\$81 万 5,400。**これは、プラットフォームの実装と運用の管理にかかる組織の給与負担です。
- ▶ **Databricks トレーニングの内部人件費は\$26 万 3,600。**これはプラットフォーム上でユーザーをトレーニングするためにかかる組織の給与負担です。

Forrester¹による既存の顧客4社のインタビューとその後の財務分析により、これらの複合組織の3年間の利益は\$2,890万、コストは\$560万で、これらから算出された正味現在価格 (NPV) は\$2,330万、ROIは417%であったことが判明しました。

財務状況の概要



利益 (3年間)



TEIのフレームワークと調査手法

インタビューで得られた情報を基に、ForresterはDatabricks統合データ分析プラットフォームの導入を検討する組織のためにTotal Economic Impact™ (TEI) フレームワークを構築しました。

このフレームワークの目的は、投資の意志決定に影響するコスト、利益、柔軟性、およびリスク要因を特定することです。ForresterはDatabricksプラットフォームが組織にもたらす影響を、多段階アプローチを使用して評価しました。

TEI手法を使用することで、企業が上級管理職や他のビジネス上の主要な利害関係者に対して、ITイニシアティブの具体的な価値を実証し、正当化し、実現することができます。



デューデリジェンス

Databricks 関係者および Forrester アナリストにインタビューし、プラットフォームに関連するデータを収集しました。



顧客企業へのインタビュー

このプラットフォームを使用する組織4社のインタビューを行い、コスト、利益、リスクに関するデータを取得しました。



モデル組織

インタビュー対象の組織の特性に基づいてモデル組織を作成しました。



財務モデルのフレームワーク

TEI手法と、インタビュー調査対象の組織が持つ課題や懸念に基づいてリスク調整した財務モデルとを使用し、インタビュー結果から導かれる典型的な財務モデルを作成しました。



ケーススタディ

4つの基本要素である利益、コスト、柔軟性、およびリスクを採用してDatabricksの影響をモデル化しました。IT投資に関連するROI分析について、組織がますます高度なテクニックを駆使するようになりつつあることを考慮すると、ForresterのTEI手法は購買意思決定に際して総合的な経済的影響の全体像を把握するのに役立ちます。TEI手法の詳細については、付録Aを参照してください。

開示事項

読者は以下の点に注意してください。

本調査はDatabricksの依頼により、Forrester Consultingが実施しました。競合分析としての利用を意図するものではありません。

Forresterは、他の企業における潜在的なROIについては一切想定していません。Databricks統合データ分析プラットフォームへの投資の妥当性を判断する際には、本調査報告で提供されているフレームワークに読者自身の予測を適用することを強く推奨します。

Databricksは本調査の報告内容を確認した後、Forresterにフィードバックを提供しました。ただし、本調査の内容と結果についてはForresterが編集する権限を有し、Forresterの見解と矛盾する変更や調査の意味を曖昧にする変更は承認していません。

Databricksはインタビューを行う顧客名を提供しましたが、インタビューには参加していません。

Databricks統合データ分析プラットフォームの カスタマージャーニー

プラットフォームへの投資前と投資後の比較

インタビューを受けた組織

本調査でForresterはDatabricks顧客4社にインタビューを実施しました。
インタビューの対象となった顧客は以下の通りです。

業種	地域	インタビュー回答者	環境
小売	EMEA（欧州、中東、アフリカ）に本社	リードデータサイエンティスト	<ul style="list-style-type: none">9人のデータサイエンティスト220 TBのデータ450人のDatabricksユーザー
製薬	本社は北米	マネージャー、データレイクおよび分析	<ul style="list-style-type: none">100人のデータサイエンティスト300 TBのデータ300人のDatabricksユーザー
重機	本社は北米	主席建築士、データレイクおよび分析	<ul style="list-style-type: none">10人のデータサイエンティスト120 TBのデータ120人のDatabricksユーザー
メディア	本社は北米	副社長、データサイエンス	<ul style="list-style-type: none">200人のデータサイエンティスト10 TBのデータ400人のDatabricksユーザー

主な課題

インタビューを受けた顧客は、彼らの所属する組織が重要なビジネス上の成果を推進するためには、もっとデータ主導型になる必要があると認識していました。しかし既存のテクノロジースタックは、資産というよりは障害であることが証明されてきました。インタビューを受けた組織は、以前のテクノロジースタックの問題点を次のように挙げています。

- サイロ化されたデータが開発プロセスを遅くした結果、MLモデルの正確度は低く、データチームの時間が無駄になっていた。データサイエンティストとアナリストは、新規プロジェクトの開始に関連するデータを集めるために苦闘していました。重機製造業界の主席建築士は次のように話しました。「データは組織内に散らばっていました。データの場所を知るには、その場所がどこかを知っている人を見つける必要があったのです。当社で採用していたシステムは行き詰まっていて、私たちの調査に従った変更をより困難にしていました。」データが限られていることと、データサイエンティストがMLモデルに取り組んだり洞察を導き出したりするための時間が比較的少ないことで、彼らの全体的な能力は弱められていました。

「以前の環境はとても遅かったので、何をやる気にもなれませんでした。」

主席建築士、データレイクおよび分析、重機製造会社



▶ **現存のインフラは、良くてプロジェクトを遅延させ、最悪の場合はそれを妨害していました。** 限界があり、固定されたオンプレミスのインフラは、作業者の新規プロジェクト開始を妨げ、現存プロジェクトの進行を遅らせていました。インタビューを受けた組織のオンプレミス環境は以下のようなものでした。

- **フルデータセットを処理するコンピュータの力不足。**メディア会社のデータサイエンスの副社長は次のように説明しました。「我が社の巨大なデータセットを分析してくれる場所はどこにもありませんでした。私たちはオンプレミスの高価な（レガシー）サーバーで身動きできなくなっていて、ニーズに合うように設計してくれる何かを必要としていました。」データチームはデータの一部に対して仕事をするように強制されていたので、実験や分析から意味のある洞察を導き出す能力が妨げられていました。
- **要求度の高い時間に合わせるように見積もれない。**製造会社の首席建築士は次のように話しました。「我が社には、繁忙期のニーズを満たすように実装された真の拡張性または弾力性はありませんでした。このような期間に、私たちは我が社の（オンプレミス）サーバーが生き延びるか否か、または入ってくる大量のデータを処理できるかを心配していました。」製薬会社のマネージャーも、このような懸念を述べました。「固定されたクラスター環境では意味をなさない使用例がいくつかあります。例えば我が社では週1回または月1回ベースで取り込まれるデータセットを取得しています。理想は、1つまたは複数のクラスターを回転させてデータセットを可能な限り迅速に処理し、ユーザーに洞察を提供することですが、固定クラスターではそれができません。」

▶ **上昇するオンプレミスとクラウドのコストへの対応への苦闘。**製薬会社のデータレイクおよび分析マネージャーは次のように話しました。「我が社の（オンプレミス）クラスターのコストは、インフラおよびライセンス両方の見地から劇的に上昇しています。」インタビューを受けた他のデータエグゼクティブは、成長するビジネスニーズについていくため、新しいオンプレミスサーバーを常に購入し続けなければならないことに言及しました。この成長に伴って常に増え続けるインフラと保守のニーズは、一部の組織をひるませることを証明しました。

クラウドに移行した組織は、コンピュータとストレージのコスト増大を経験しました。あるエグゼクティブは、組織のストレージコストが年間16%増大したことに気付いたのです。

▶ **ビジネスの成果を推進するための効果的なコラボレーションツールの不足。**以前、データサイエンス、エンジニアリング、および分析のチームは、開発スタック内に最新のコラボレーションツールを有していなかったため、プロジェクト上でデータチームが独立して作業することがよくありました。小売業のリードデータサイエンティストは次のように説明しました。「以前には、誰かが何かを構築して、それがうまく作動したら、そこで終わりでした。ですから問題が起きたとき、誰もエラー修正のフォローアップをしなかったのです。」ビジネスの成果を推進する効果的なMLモデルの作成には、データサイエンティスト、エンジニア、アプリ開発者、およびビジネス上のステークホルダーの協働が必要です。協働に適したツールがないと、データサイエンティストのチームは、モデルを効果的に作成してそれを反復するこ

「我が社の（オンプレミス）インフラとライセンスのコストはストレージの必要性が上がるにつれて劇的に増大しますが、コンピュータのニーズはそれほど変化しません。増大しつつあるインフラへの恐れを何とかしようとしているところです。」

マネージャー、データレイクおよび分析、製薬



とができません。さらに、適切な文書化がないと、データサイエンスのチームは誰も理解できないモデルを有するリスクがあり、そのモデルを理解するため、またはそれについて言及するために多くの時間を費やすことになります。

ソリューションの要件

インタビューを受けた組織は、以下を可能とするソリューションを求めていました。

- ▶ 最新のテクノロジスタックを提供して、協働を助長し、ビジネスの成果を推進する。
- ▶ データサイエンティストとアナリストの時間を開放して、価値の高い活動に集中できるようにする。
- ▶ クラウドへの移行を可能にする。
- ▶ 改善された管理方針を通して、データをクリーンで高性能に保つ。

主な結果

インタビューにより、Databricksへの投資により得られる以下の主な結果が明らかになりました。

- ▶ **大衆化されたデータへのアクセスにより様々なビジネスの成果が推進される。** 組織内の全員が、様々な展望とスキルのセットをもたらすDatabricksプラットフォームを活用できるようになり、これらの新規ユーザーによって収益は上がり、コストは下がり、業績は改善されました。
 - **迅速な市場展開。** Databricksの採用により、インタビューを受けた組織は以前の環境でのボトルネックに対処できました。小売業者のリードデータサイエンティストは次のように話しました。「今ではマーケティング担当者自身がソフトウェアとツールを構築し、データサイエンティストや開発者の助けなしにそれらを作成できます。（それにより）市場への投入時間が短縮され、必要なリソースも少なく済みます。」
 - **新しい機会の特定による増収。** 小売業のマーケティングチームは、売り切れた物品の返品を予想することによって、販売を増大させる機会を特定しました。リードデータサイエンティストはマーケティングチームについて次のように話しました。「Databricksのノートブックから、完全に彼らだけでソリューションを構築しました。」

主席建築士もこの点を繰り返し、次のように話しました。

「以前には難しかったことができるようになったので、市場への投入が迅速になりました。本当に役立っています。規模と業績を伴った多くの結果を得ています。以前には処理しきれなかった多くのデータが処理可能になりました。以前には取り扱いが非常に難しかった大きなデータセットも取り扱えます。以前にはなかった可能性が広がり、抱えていた問題も少なくなりました。アウトプットが市場へ迅速に投入され、アイデアもすぐに実現します。以前には基本的作業を（終わらせるために）苦労していたので、別のアイデアを取り入れたり、解決する時間がなかった問題がありましたが、今ではこれらにも取り組むことができるようになりました。」

「Databricksは摩擦を減らし、管理を大衆化し、データサイエンスが何の障害もなくR&Dを実施できるようにしました。」

副社長、データサイエンス、メディア



「以前には、すべてがどのように作動しているかを理解しているアナリストの数は限られていたので、すべて彼らを通す必要がありました。しかし今では、基本的に全員が学習しています。社内の全員がDatabricksによって何かをすることを望んでいます。」

リードデータサイエンティスト、小売業



- 熟練した従業員が新しい機会を発見する一方で、データサイエンティストチームの負担は減少した。リードデータサイエンティストは、組織内のあるマーケティング担当者が運用を25%最適化する方法を特定したと説明しました。「彼にはデータサイエンスまたはデータエンジニアリングの経験は何もありませんでした。ただそれを取り上げただけです。そして、それをどのように行うかについての素晴らしい例をブログの記事にしました。」

重機製造会社の実務責任者は、Databricksの採用によって組織全体のマインドセットがシフトしたと話しました。「我が社のカスタマーサポートチームは、「この分野におけるお客様と機械をサポートするため、私たちが持っている情報をいかにより良く利用するか」を把握しようとしています。そのために彼らは、問題を早期に捉え、販売業者に積極的な修理を行うよう警告するモデルを作成しています。我が社のエンジニアリングチームは「私たちが持っている情報をどのように製造過程の改善に使用するか」について問いかけています。パーツ部門は「この主導プロセスをどのように自動化できるか」と問いかけています。これらのイニシアチブの幾つかについて、その影響を測るのは早すぎますが、実務責任者の報告によると、パーツ価格部門は組織全体のカタログ中で比較的わずかなパーセンテージの価格を最適化しただけで、既に数百万ドルの追加収益が得られました。

- ▶ コラボレーションと実験の手腕の改善がより良い ML 製品を生み出し、それによって収益が上がりコストは減少する。Databricks プラットフォームによって、インタビューを受けた顧客の開発サイクルは速まり、ML モデルの改良に時間をかけられるようになりました。小売業のリードデータサイエンティストは次のように話しました。「Databricks の使用によって、ML モデルの完成、動作チェック、そして当該モデルをレビューして改良する担当者の動員にかかる時間が短縮される一方で、そのモデルを時間をかけて追跡できるようにもなりました。」このような能力の追加により、インタビューを受けた組織は以下に挙げる社内アプリケーションを作成または最適化できました。

- 製薬会社が AI エンジンを作成して、セールス担当者はどの見込み客に営業努力を集中させるべきか、それをいつ行うべきかを特定しました。この組織はまた、原材料への支払いが過分であることを見出し、契約を再交渉して数億ドルを節約できました。
- 小売業者は、検索エンジンを改良し、様々な顧客の個性に対する推奨エンジンを作成することによって売り上げを増やしました。また返品を減らし、不正な注文を特定することによって、運用コストを下げることができました。
- 重機製造会社では、製造上の問題の特定、顧客へのダウンタイムの減少、および保証修理費の低下に対する実施能力が改善されました。

「私たちは組織全体での巨大な利益を認識しています。社員が開発者またはアナリストに頼ることが減り、自分で解決する機会が増えました。社員は自分たちでプロジェクトを構築できます。」

リードデータサイエンティスト、
小売業



- さらに多くの ML モデル作成が可能。ML ライブラリーへのアクセス、MLflow（ML のライフサイクルを管理するオープンソースプラットフォーム）に元から組み込まれている実験の追跡機能、およびコラボレーション型ノートブックによって、インタビューを受けた組織による ML モデルの作成を促進します。さらに、プラットフォームへのアクセスの大衆化によって、組織内のさらに多くの人々が ML モデルとアルゴリズムに貢献できるようになりました。
- さらに安定した効率の良い環境を提供。Databricks プラットフォーム上でのプロセスは、インタビューを受けた顧客が以前のソリューションに費やしていた時間に比べ、ごくわずかな時間で完了しました。例えば主席建築士は、コンピュータ時間が 40%減少、および Apache Spark（オープンソースであり、大量のデータを処理及び分析するために用いられる多目的分散コンピューティングエンジン）のプロセス完了までの時間が 97%減少したことを認識しました。主席建築士はまた、Databricks 導入前には 0.4%から 1.0%であった作業の失敗率が、Databricks の導入によって 0.1%に低下したことも見出しました。一方、データサイエンス会社の副社長は、Databricks 導入前には完了までに 1 週間を要したプロセスが、Databricks の導入によって 1.5 時間になったことを認識しました。このような成績の向上は、データサイエンティストおよびアナリストがスプリントサイクル中にさらに多くの実験を行うことを助け、また組織がそれらのモデルに対してより良い判断を行うことを助けました。
- インフラを減らして管理を単純にする。Databricks への移行により、インタビューを受けた組織は、オンプレミスでのインフラを廃止して余分なソフトウェアライセンスをキャンセルし、IT リソースを再配置することが可能になりました。以前の環境に比べると、プラットフォームの管理は大幅に容易になったことが証明されました。環境の管理をさらに単純にするため、組織は DevOps 環境に移行できます。製薬会社のデータレイクと分析を担当するマネージャーは次のように述べました。「(Databricks によって) 生活はとても容易になりました。(誰も) 自分の時間すべてを財務とアップグレードやパフォーマンスの低下に不満を漏らす社員への対応に費やしたいとは思わないでしょう。(Databricks 上で) 誰もが費用を負担して自分のクラスターを持つ DevOps を採用しました。これにより各自が好きな時にサイズの管理とアップグレードを実行できます。」
- バイオテクノロジーと遺伝学の分野に貴重な専門的知識を提供。製薬会社のデータレイクと分析を担当するマネージャーは次のように話しました。「Databricks には Spark プログラマーと遺伝学者が所属しているため、ヘルスケア分野での大きな助けになります。何かを自分で解決するために私たちの時間すべてを使う代わりに、Databricks へ助けを求めることができます。彼らはバイオテクノロジー分野に投資しているため、私たちが DNA シークエンシングによって何をしようとしているかを理解できるのです。」

複合組織

インタビュー調査に基づき、Forrester は TEI フレームワーク、複合組織、関連する ROI 分析を構築し、財政的に影響する領域を具体的に示しました。複合組織とは Forrester がインタビュー調査した 4 社を代表する組織であり、次のセクションでは財務分析および分析の総合結果を表すために使用されています。Forrester が顧客のインタビューから抽出した複合組織は、次の特性を備えています。

「以前には、ML モデルの分析と反復を行う前のプロセスを完了するまで 2 週間待たなくてはなりませんでしたが。今では同じ作業が 1 時間で済みます。私たちはモデルを 1 日に何回も反復できます。それによって研究の成果を真に理解し、それを微調整し、障害なしに新しいモデルを作成することが可能です。」

副社長、データサイエンス、メディア



主な前提条件

\$50 億の収益

100 人のデータサイエンティスト

200 人のデータエンジニア

3 年目までに 525 人の

Databricks ユーザー

複合組織の説明。複合組織は、B2CおよびB2B商品の取り扱いとサービスを営むグローバル企業です。本社は米国にあり、年間収益は\$50億です。

この組織は現在、大規模なオンプレミスデータ分析フットプリント、100人のデータサイエンティスト、200人のデータエンジニア、450 TBのデータを有しています。一連のMLモデルを有しており、それが推奨するエンジンと検索エンジンを支持しています。

デプロイメントの特徴。この複合組織は、毎年25%のデータ、データサイエンティスト、データエンジニアを移行させるやり方で、3年をかけて徐々にDatabricksプラットフォームへ移行しています。加えて、この組織は毎年、マーケティング担当者からビジネスアナリスト、アプリケーション開発者、プロダクトマネージャーなどにわたる100人の新規ユーザーを獲得しています。

移行の一部として、その複合組織はDelta Lake（ACIDトランザクションをApache Spark™とビッグデータワークロードに持ち込むオープンソースのストレージレイヤー）を採用し、それによって管理標準を実行してデータを高性能に保っています。この組織は、Databricksの手腕を既存のMLモデルの改良と新規モデルの作成に利用しています。新規MLモデルの例には、価格最適化エンジン、不正検出分析、製造工程の最適化、および保守の予想が含まれます。

利益の分析

コンポジットに適用した定量的利益のデータ

総利益

参照	利益	1年目	2年目	3年目	合計	現在価値
Atr	増分利益	\$843,750	\$3,375,000	\$8,437,500	\$12,656,250	\$9,895,520
Btr	運営効率の向上	\$1,594,219	\$3,188,438	\$4,782,656	\$9,565,313	\$7,677,643
Ctr	レガシーデータ分析プラットフォームのコスト削減	\$1,840,625	\$4,417,500	\$7,951,500	\$14,209,625	\$11,298,202
	総利益（リスク調整後）	\$4,278,594	\$10,980,938	\$21,171,656	\$36,431,188	\$28,871,365

増分利益

Databricks統合データ分析プラットフォームは、より良いMLモデルを作成し、分析を以前よりも迅速にするツールを提供することで、インタビューを受けた顧客の収益を上げることを可能にしました。Databricksにより、分析とデータサイエンスのチームは、データのクリーニング、およびETLパイプラインの作成と維持に費やす時間を減らし、ビジネスの成果を推進するMLモデルの構築と改良に多くの時間を使えるようになりました。インタビューを受けた組織は、以下を通して増分利益を得ました。

- ▶ **顧客体験 (CX) の改善を通じた売り上げの増大。** さまざまな購買者の個性を考慮した推奨エンジンの数を増やし、効果を上げることで、小売業者への注文頻度が上昇しました。注文頻度が1パーセント上がるごとに、収益は2.8%増える計算になりました。小売業者はまた、売り上げを失う結果となる、顧客が不満を感じる点を特定することもできました。例えば、カートをキャンセルする理由の3分の1超は、アカウント作成の強制だということが分かりました。この小売業者は、アカウント作成の条件を除くことによって収益を4%上げられる可能性があります。
- ▶ **価格の最適化。** 製造会社の主席建築士は次のように話しました。「我が社のパーツ価格チームは、手動のプロセスから Databricks を用いてさらに自動化されたプロセスに移行しました。彼らは、価格の最適化によって数百万ドルの利益がもたらされたと推定していますが、長期間検討したいパーツについてはまだほとんど価格設定を行っていません。私たちはこの機会が年間数百万ドルよりもっと重要だと考えています。」
- ▶ **セールス担当者の能力を高める。** 製薬会社は、セールス担当者が連絡を取るべき人々とその時期について優先順位をつけることで彼らの努力を最適化する AI アシスタントを創出しました。「連絡可能な処方者の数は、セールス担当者の数よりも常に多いのです。連絡を取る人を

上表は、下に掲載された領域全体にわたる総利益と、10%の割引適用後の現在価値 (PV) を示します。3年間にこの複合組織にもたらされると推定される総利益のリスク調整後の現在価値は \$2,800 万超です。

優先させる方法が必要なのですが... セールス担当者の AI アシスタントを創出することによって、セールス担当者が時間をさらに最適化して収益を上げられるように援助しています。」

- ▶ **結果的に ML モデルは生産を以前よりも迅速にする。**インタビューを受けた顧客は、Databricks によって市場への投入がどれほど迅速になったかを定量化するのに苦闘していましたが、市場へ至る時間が金銭的利益に影響することを Databricks によって認識したということには同意していました。インタビューを受けた組織は ML モデルの金銭的利益を以前よりも迅速に認識できるようになりました。加えて、同じ時間内でさらに多くの ML モデルを作成できるようになり、その結果として金銭的利益も増えました。組織は、年間の増収を推進する 1 つの ML モデルを作成する代わりに、同じ期間に複数の ML モデルをリリースすることが可能です。

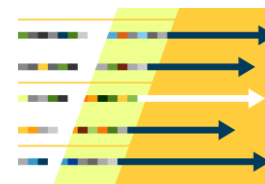
このような利益をモデル化するため、Forrester は以下を想定しました。

- ▶ 複合組織は、新規 ML モデルの作成と既存の ML モデルの最適化を通して転換率を改善し、価格を最適化し、売り込むことによって収益が上がります。
- ▶ Databricks のユーザーが協働と ML モデルの改良と新しい見識の発見を続けるに従い、Databricks を取り入れたことによる収益が上がります。
- ▶ Databricks により、複合組織はさらに多くの ML モデルの開発とリリースを以前よりも迅速に行うことが可能になるので、複合組織は増収を早期に認識できます。

増収は以下に関連して変動する可能性があります。

- ▶ 新規 ML モデルの作成とデータから洞察を導き出す Databricks ユーザーの能力。
- ▶ ML モデルを改良する Databricks ユーザーの能力。
- ▶ 組織の営業利益。
- ▶ 組織が増収のイニシアチブをどれだけ優先しているか。

これらのリスクを考慮するために、Forrester はこの利益を 10% 下方修正し、リスク調整後の 3 年間の現在価値 (PV) を \$9,895,520 としました。



Databricks の利用で ML モデルの作成と最適化が迅速になり、増分利益が得られます。

影響リスクとは、組織の業務ニーズまたは技術的ニーズが投資によって満足されず、全体的な利益が低下するリスクです。不確実性が高いほど、利益の潜在的な見積もり範囲が広がります。

増分利益：計算表

参照	評価項目	計算法	1年目	2年目	3年目
A1	年間収益	コンポジット	\$5,000,000,000	\$5,000,000,000	\$5,000,000,000
A2	Databricksに影響を受けた組織/収益のパーセント	コンポジット	25%	50%	75%
A3	Databricksによりもたらされた増収	仮定	1.50%	3.00%	5.00%
A4	Databricksによりもたらされた増収	$A1 \times A2 \times A3$	\$18,750,000	\$75,000,000	\$187,500,000
A5	営業利益	NYU Stern	5%	5%	5%
At	増分利益	$A4 \times A5$	\$937,500	\$3,750,000	\$9,375,000
	リスク調整	↓10%			
Atr	増分利益（リスク調整済み）		\$843,750	\$3,375,000	\$8,437,500

運営効率の向上

Databricksプラットフォームへの移行前、分析、データサイエンス、およびエンジニアリングのチームは、MLモデルの作成やデータから洞察を導き出すといったこと以外の、価値の低い活動にほとんどの時間を費やしていました。データサイエンティスト、アナリスト、およびエンジニアは、ほとんどの時間を以下のようなことに費やしていました。

- データの採集とクリーニング。** 分析とデータサイエンスのチームは、データセットの実験準備のためだけに数週間かかることがありました。これらのデータチームは、サイロ全体からデータを探したり、データをクリーニングしたりすることに非常に多くの時間を費やしていたのです。
- テストおよび生産環境の構築。** データサイエンスとデータエンジニアリングのチームはサイロ内で作業していたため、プロジェクト上のコラボレーションは困難でした。データサイエンティストとエンジニアは異なるプログラミング言語と方法を用いており、他方が行っていることを理解できませんでした。その結果、新しいテスト環境または生産環境を構築するための作業には何度もやり取りを繰り返すことが必要だったのです。

メディア会社のデータサイエンスの副社長は、彼の会社のチームが直面する困難について次のように説明しました。「私たちはこれまでにR&Dのすべてを社内環境で行ってきました。必要な書類を作成し、それをエンジニアリングチームへ送って開発環境を構築してもらって

「私のアウトプットは市場へ迅速に投入され、アイデアもすぐに実現します。以前には基本的作業を（終わらせるために）苦労していたので、別のアイデアを取り入れたら、解決する時間がなかった問題がありましたが、今ではこれらにも取り組むことができるようになりました。」

主席建築士、データレイクおよび分析、重機製造会社



ました。彼らは私たちがしてほしいことを本当には理解していなかったもので、何度もやり取りを繰り返しました。非常に遅くてやっかいな開発パラダイムでした。」

データサイエンティストは上記のようなタスクに非常に多くの時間を使う必要があったため、付加価値のある活動に費やす時間は比較的少なかったのです。データサイエンティストが、MLモデルの反復、ステークホルダーとのコラボレーション、または新規プロジェクトのための作業に使う時間はごくわずかでした。インタビューを受けた顧客の多くにとって、彼らのインフラはこの問題を悪化させたただけでした。遅い処理時間と自動化の欠如は、2週間のスプリントサイクルの間にデータサイエンティストが行いうる実験はわずかに1つであることを意味します。

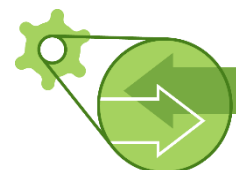
インタビューを受けた組織は、クリーンで高性能なデータを保つため、Delta Lakeを活用していました。データの品質基準に厳しく従い、データを集中化することによって、インタビューを受けた組織は新規プロジェクトの開始に必要な時間と努力を減らしました。さらに、もっと多くの高品質なデータへアクセスすることによってMLモデルの質が改善されました。

インタビューを受けた組織は、Databricksプラットフォームへ移行してDelta Lake上でデータを管理するようになってから、データサイエンスチーム全体で生産性が上がったと報告しました。アナリストとデータサイエンティストは、データの検索、そのクリーニング、または結果の取捨選択を少ない時間でできるようになり、ビジネスの成果を推進する活動に多くの時間を使えるようになりました。また、コラボレーションの改善とMLライフサイクルの管理機能は、データサイエンティストがMLモデルを以前よりも迅速に作成および改良するのに役立ちました。プラットフォームのさらなる特色と機能により、チームはさらに迅速に作業できるようになりました。データサイエンスチームは、1つの実験を2週間のスプリントサイクルにわたって行う代わりに、複数の実験を1日で行うことができ、それによって結果のより良い理解が得られました。その結果、より良いMLモデルが得られました。

プラットフォーム上でサポートされる追加言語が、アナリスト、データサイエンティスト、およびエンジニアの協働に役立ちました。メディア会社のデータサイエンスの副社長は次のように話しました。

「Databricksは、我が社が古いソリューションを使用していた頃に比べて、さらにエンジニアリングチームを統合してくれました。私たちは彼らのテクノロジスタックによってさらに統合を強化し、サポートしてくれる言語で作業し、リポジトリを共有しています。」

インタビューを受けた組織は、Databricksプラットフォームによって以下に挙げるような時間の節約ができたことを認識しました。



25%の上昇がデータサイエンティストの生産性にみられます。

カテゴリー	DATABRICKS導入前	DATABRICKS導入後
ETLパイプラインの作成と維持	数日間	数分間
新規データ製品の実験のための、サーバークラスターの設計	数週～数カ月間	数時間
データの検証とクリーニング	数日間	数時間
モデルトレーニング/ハイパーパラメータ調整およびモデル検証	不在に制限される	数時間～1日
モデルの生産開始、モニター、管理、およびモデルの再トレーニング	数ヶ月にわたり、2週間のスプリントサイクルが複数回	可能なセルフサービスのレベルに応じて数日～数週間
データ製品を開発から生産へ移行	数週～数カ月間	可能なセルフサービスのレベルに応じて数日～数週間

これらの因子すべてにより、インタビューを受けた組織のすべてが、有意義でより良いMLモデルを以前よりも迅速かつ少ないリソースで作成できるようになりました。この複合組織に対し、Forresterは以下のように想定しました。

- データサイエンティストとデータエンジニアは、75%の時間をデータの検索、データのクリーニング、テスト環境の構築、およびモデルの生産への移行に費やします。
- Forresterは、データサイエンティストとエンジニアが上記の活動に費やす時間を、控えめに見積もってもそれぞれ25%および20%少なくできると想定します。データサイエンティストは、主にデータの検索とクリーニングに使っていた時間を減らすことで、これらを達成します。一方でデータエンジニアは、主にデータのクリーニングと、テストおよび生産環境の構築に使っていた時間を減らします。

生産性の上昇は以下に基づいて変動します。

- データサイエンティストとデータエンジニアが現在どのように時間を使っているか。
- データ品質。
- データサイエンティストとデータエンジニアの人件費

これらのリスクを考慮するために、Forresterはこの利益を5%下方調整し、調整された3年間の総額のPVをおよそ\$770万としました。



20%の上昇がエンジニアの生産性にみられます

運営効率の向上: 計算表

参照	評価項目	計算法	1年目	2年目	3年目
B1	Databricks に影響を受けたデータサイエンティスト	コンポジット	25	50	75
B2	Databricks の採用により影響を受けたデータサイエンティストの時間のパーセント	顧客企業へのインタビューに基づく	75%	75%	75%
B3	Databricks により節約されたデータサイエンティストの時間	顧客企業へのインタビューに基づく	25%	25%	25%
B4	データサイエンティストの給与	業界平均	\$150,000	\$150,000	\$150,000
B5	Databricks に影響を受けたデータエンジニア	コンポジット	50	100	150
B6	Databricks に影響を受けたデータエンジニアの時間のパーセント	顧客企業へのインタビューに基づく	75%	75%	75%
B7	Databricks の採用により影響を受けたデータエンジニアの時間のパーセント	顧客企業へのインタビューに基づく	20%	20%	20%
B8	データエンジニアの給与	業界平均	\$130,000	\$130,000	\$130,000
Bt	運営効率の向上	$(B1*B2*B3*B4)+(B5*B6*B7*B8)$	\$1,678,125	\$3,356,250	\$5,034,375
	リスク調整	↓5%			
Btr	運営効率の向上 (リスク調整後)		\$1,594,219	\$3,188,438	\$4,782,656

レガシーデータ分析プラットフォームのコスト削減

Databricksの採用前、インタビューを受けた組織のほとんどが広範なオンプレミスのデータ分析環境を管理していました。インタビューを受けた組織はまた、オープンソースソリューションを実行する商用サーバーもしくは知的財産権下にあるソフトウェアを実行する高額なサーバーのいずれか、または場合によりその両方を数百から数千所有していました。組織的なニーズを満たし続けるため、インタビューを受けた顧客はさらに多くのサーバーを購入し続け、管理用の間接費も増えていました。あるエグゼクティブは、以前の環境の管理の困難さについて次のように説明しました。「ニーズが増え続けるので、常に新しい機械を購入したり新たなインスタンスを付加したりしなければなりません。インフラを拡大したり新しいライセンスを購入したりすれば、管理は難しくなります。資金を（様々な部門にわたって）処理し、要求を出し、設備をアップグレードしなくてはなりません。」環境が成長を続けるに従い、このような問題は困難さを増すだけです。

クラウドへの移行は、コストのカーブの下降にはわずかにしか役立ちません。あるエグゼクティブは、増え続ける要求に合わせるため、クラウドにかかるコストが毎年16%超上昇したことを認識しました。

Databricksにより、インタビューを受けた顧客はオンプレミスのインフラの廃止を始めて、サードパーティーからのライセンスとサービスを減らしたりキャンセルしたりできるようになりました。さらに、Databricksは管理費削減にも役立ちました。エンジニアはもうプラットフォームの保守やアップグレードの心配をする必要がなくなったのです。インタビューを受けた顧客のうち数社は、さらにセルフサービスポータルを提供してDevOpsへ移行することにより、管理の必要性を減らしました。「我が社ではDevOpsを採用し、「所有して、サイズを調整して、コストを管理する」システムに移行しています。APIラッパーと組織化ツールを使用して、適切な許可を有し、タグ付けを強制し、すべての関連動作を記録する新規インスタンスをセットアップします。」

クラウドにも及ぶコスト削減。インタビューを受けた顧客は、データレイクのサイズが大幅に大きくなっているにも関わらず、クラウドにかかるコストが変わらないかまたは迅速に以前よりも低下していることを見出しました。非常に良い性能と安定性を備えたDatabricksの自動スケールリングにより、クラウドインフラのコストは低下しました。

また、インタビューを受けたデータエグゼクティブは、Databricksへの移行後に安定性とコンピューットの著しい改善を認識しました。

- ▶ 重機製造会社の主席建築士は、特定のプロセスで40%から50%の上昇を認めたと話しました。彼らはさらに、Spark プロセスの完了に必要な時間が、オープンソースのSpark を用いた以前のクラウドソリューションに比べて97%減少したことを経験しました。この主席建築士はまた、以前の環境での問題についても説明しました。「どうしても実行できないプロセスがいくつかありました。2~3時間の実行後に、サーバー上でクラッシュしてしまうのです。(Databricksでは)実行されてから数秒以内に完了します。」最後に、失敗した作業の数は、100~250作業ごとに1(Databricks導入前)から1,000作業ごとに1(Databricks導入後)に減少しました。
- ▶ データサイエンス会社の副社長は、以前のソリューションでは完了までに1週間を要したプロセスが、Databricksでは1.5時間になったことを認識しました。彼は次のように話しました。「以前、作業はコンピューターに1週間、手動プロセスに1週間かかりました。プログラムによって使用可能な、中央化されたデータレイク内のデータと、Databricksのクラスター管理の機能性によって、私たちは全てのプロセスを端から端まで自動化できます。」

この複合組織に対し、Forresterは以下のように想定しました。

- ▶ オンプレミスのインフラの年間維持費は\$700万。これには、サーバーのデコミッション、ネットワークング、冷却、ストレージ、内部管理の件費にかかるハードウェア関連のコスト、およびデータ分析ソリューションに関連しないソフトウェアのコストが含まれます。
- ▶ インフラのコストは、増え続けるニーズを満たすために前年比で20%上昇。

オンプレミスのインフラのコストは以下によって変動します。

- ▶ 組織のオンプレミスデータ分析環境のサイズ。
- ▶ データ分析ライセンス費用。

「クラウドにかかる我が社のコストは、指数関数的成長とクラウドコンピューットの使用にも関わらず、前年と変わりませんでした。」

副社長、データサイエンス、メディア



75%のレガシー環境が、3年を超えるDatabricksの使用によって廃止されました

これらのリスクを考慮するために、Forresterはこの利益を5%下方修正し、リスク調整後の3年間の総額のPVを \$11,298,202としました。

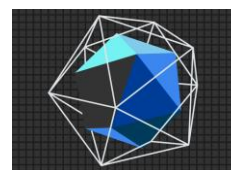
レガシーデータ分析プラットフォームのコスト削減：計算表

参照	評価項目	計算法	1年目	2年目	3年目
C1	レガシーデータ分析ソリューションのためのオンプレミスインフラの年間コスト	コンポジット	\$7,000,000	\$8,400,000	\$10,080,000
C2	レガシーデータ分析プラットフォームのライセンスとサポートのコスト	コンポジット	\$750,000	\$900,000	\$1,080,000
C3	環境のデコミッションのパーセント	コンポジット	25%	50%	75%
C4	ハードウェアとライセンスのコスト回避	$(C1+C2)*C3$	\$1,937,500	\$4,650,000	\$8,370,000
Ct	レガシーデータ分析プラットフォームのコスト削減	C4	\$1,937,500	\$4,650,000	\$8,370,000
	リスク調整	↓5%			
Ctr	レガシーデータ分析プラットフォームのコスト削減（リスク調整後）		\$1,840,625	\$4,417,500	\$7,951,500

非定量化利益

インタビューを受けた組織は、この調査では金銭的に定量化できない以下のような利益を得ました。Databricks統合データ分析プラットフォームに対してForresterが評価した非定量化利益には以下が含まれます。

- 運用コストの削減。** Databricks プラットフォーム上で作成された新規 ML モデルと、データの中央化によって得られた新しい見識との組み合わせは、インタビューを受けた組織が原材料、製造上の問題の早期同定にかかるコストを低下させ、保証修理費と返品率を減少させました。
 - 製薬会社が AI エンジンを作成して、セールス担当者は誰に営業努力を集中させるべきか、それをいつ行うべきかを特定しました。この組織はまた、原材料への支払いが過分であることを見出し、契約を再交渉して数億ドルを節約できました。
 - 小売業者は、検索エンジンを改良し、様々な顧客の個性に対する推奨エンジンを作成することによって売り上げを増やしました。また返品を減らし、不正な注文を特定することによって、運用コストを下げることもできました。
 - 重機製造会社では、製造上の問題を特定して顧客へのダウンタイムを減少させ、かつ保証修理費を低下させる実施能力が改善されました。



より良いセキュリティとデータ管理

TEI の定義によれば、柔軟性とは将来の追加投資で事業利益に還元可能な機能または能力の拡大への投資を意味します。これにより、組織は将来の推進事項を開始する「権利」または機能を得ることができますが、それは義務ではありません。

- › **セキュリティの向上。** Databricks の使用前、インタビューを受けた組織には、複数の環境を網羅するセキュリティ標準、およびセキュリティ標準の実行に必要なインターネットリソースがありませんでした。Databricks はネイティブセキュリティ保護を提供しました。

柔軟性

柔軟性の価値は明らかに顧客ごとに異なり、その価値の尺度も組織によって異なります。顧客がDatabricksプラットフォームを実装するシナリオは複数存在し、以下に示すように追加の使用やビジネスチャンスを実現する場合があります。

- › **新しい ML ライブラリーを活用する。** Databricks は、すぐに使える一般的な ML ライブラリーとフレームワークを提供して、データサイエンティストが最新の機械学習技術により容易に実験できるようにします。
- › **新しいプログラミング言語を採用する。** Databricks は広範な最新のプログラミング言語（R、Python、Scala、および SQL を含む）をサポートして、チームが同じノートブックの中で新しいプログラミング言語を使用してデータを探索できるようにします。

柔軟性は、特定のプロジェクトの一環として評価すれば定量化できます（詳細は付録Aに記載）。

コストの分析

コンポジットに適用された定量化コストデータ

総コスト

参照	コスト	初期	1年目	2年目	3年目	合計	現在価値
Dtr	Databricks プラットフォーム、トレーニング、およびストレージのコスト	\$0	\$971,250	\$1,732,500	\$2,913,750	\$5,617,500	\$4,503,916
Etr	Databricks の管理コスト	\$136,500	\$273,000	\$273,000	\$273,000	\$955,500	\$815,411
Ftr	Databricks のトレーニングコスト	\$4,234	\$101,606	\$105,840	\$105,840	\$317,520	\$263,593
総コスト（リスク調整後）		\$140,734	\$1,345,856	\$2,111,340	\$3,292,590	\$6,890,520	\$5,582,920

Databricksプラットフォーム、トレーニング、およびストレージのコスト

Databricks統合データ分析プラットフォームの使用にあたり、組織はDatabricksユニット (DBU) と呼ばれる消費コストを負担することになります。DBUは1時間あたりの処理能力の単位で、1秒あたりの使用量として請求されます。加えて、インタビューを受けた組織は、アナリスト、データサイエンティスト、データエンジニアの新人研修とトレーニング、ならびにマーケティング担当者、ビジネス上のステークホルダー、および組織全体の他の人々へのプラットフォーム使用の大衆化に対して支払いを行いました。

さらに、この組織はサードパーティーのクラウドプロバイダーによるストレージの費用を負担しました。

この複合組織に対し、Forresterは以下のように想定しました。

- ▶ さらに多くのデータとビジネスユニットが Databricks プラットフォームへ移行するにつれて、使用量のコストは毎年上昇。

Forresterは、これらのコストが以下に基づいて変動することを認識しています。

- ▶ Databricks プラットフォームの使用量。
- ▶ クラウドへ移行したデータの量。
- ▶ プラットフォームのトレーニングを受けた人数。

これらのリスクを考慮するために、Forresterはこのコストを5%上方修正し、リスク調整後の3年間の現在価値 (PV) を\$4,503,916としました。

上表は、以下に挙げる領域全体にわたる総コストと、10%の割引適用後の現在価値 (PV) を示します。3年間にこの複合組織にもたらされると推定される総コストのリスク調整後の現在価値は\$550万超です。

実装リスクとは、提案された投資が初期の、または予想される要件から外れ、予想より高いコストとなるリスクです。不確実性が高いほど、コストの潜在的な見積もり範囲が広がります。

Databricks プラットフォーム、トレーニング、およびストレージのコスト：計算表

参照	評価項目	計算法	初期	1年目	2年目	3年目
D1	Databricks プラットフォームおよびトレーニングのコスト			\$800,000	\$1,400,000	\$2,400,000
D2	クラウドストレージのコスト			\$125,000	\$250,000	\$375,000
Dt	Databricks プラットフォーム、トレーニング、およびストレージのコスト	D1+D2	\$0	\$925,000	\$1,650,000	\$2,775,000
	リスク調整			↑5%		
Dtr	Databricks プラットフォーム、トレーニング、およびストレージのコスト（リスク調整後）		\$0	\$971,250	\$1,732,500	\$2,913,750

Databricksの管理コスト

Databricks統合データ分析プラットフォームを完全に導入する前、インタビューを受けた組織はこのプラットフォームが彼らのニーズに合うかを確認するため、概念実証 (PoC) を行いました。その後、インタビューを受けた組織は、以下を含むDatabricksプラットフォームの管理にデータエンジニアを任命しました。

- › Delta Lake の管理。これには管理方針の最初の実行、および継続的に行われるプラットフォームへのデータの移行が含まれます。
- › 新入社員へのトレーニングの支援。

このような管理費のモデル化において、Forresterは2人のデータエンジニアがDatabricksプラットフォームの継続的管理に任命されると推測します。

管理費は以下に基づいて変動する可能性があります。

- › 人員数と現在持っている知識や技能。
- › 展開の規模と範囲。
- › クラウドへ移行したデータの量。
- › ビジネス要件。
- › データ管理要件。

これらのリスクを考慮するために、Forresterはこのコストを5%上方修正し、リスク調整後の3年間の現在価値 (PV) を\$815,411としました。



6 カ月：
導入および展開の総期間



2 人の FTE が Databricks
プラットフォームの継続
的な管理に 業務時間の
100%を費やします。

Databricks の管理コスト：計算表

参照	評価項目	計算法	初期	1年目	2年目	3年目
E1	Databricks の PoC と実行の専任データエンジニア	コンポジット	2	0.00	0.00	0.00
E2	Databricks のセットアップに必要な月数		6	0.00	0.00	0.00
E3	Databricks の保守に必要なデータエンジニア			2.00	2.00	2.00
E4	データエンジニアの給与		\$130,000	\$130,000	\$130,000	\$130,000
Et	Databricks の管理コスト	$(E1 \cdot E2 / 12 \cdot E4) + (E3 \cdot E4)$	\$130,000	\$260,000	\$260,000	\$260,000
	リスク調整	↑5%				
Etr	Databricks の管理コスト（リスク調整後）		\$136,500	\$273,000	\$273,000	\$273,000

Databricks のトレーニングコスト

上で算出したトレーニング費用に加え、Forresterは複合組織の従業員を Databricks プラットフォーム上でトレーニングする際の内部人件費を算出しました。この費用の算出において、Forresterは以下のように想定しました。

- 最初の PoC の一部として、7 人の FTE（データエンジニア、データサイエンティスト、およびその他のユーザーを含む）がプラットフォーム上でトレーニングされます。
- 25 人のデータサイエンティスト、50 人のデータエンジニア、およびその他 100 人のユーザーが、2 年目と 3 年目に Databricks プラットフォーム上でトレーニングされます。
- ユーザーは、平均 12 時間をインストラクターによるセミナー、昼食および学習セッション、ならびに自己学習コースに費やします。
- 年間の混合給与は平均で \$100,000 です。

トレーニングコストは次の要因に基づいて変化します：

- 年間にトレーニングされる FTE の人数。
- トレーニングの必要性を高めたり低めたりする、現在持っている知識や技能。
- データサイエンティストおよびデータエンジニア以外のユーザーをトレーニングするか否か。
- FTE の給与。

これらのリスクを考慮するために、Forresterはこのコストを5%上方修正し、リスク調整後の3年間の現在価値(PV)を\$263,593としました。

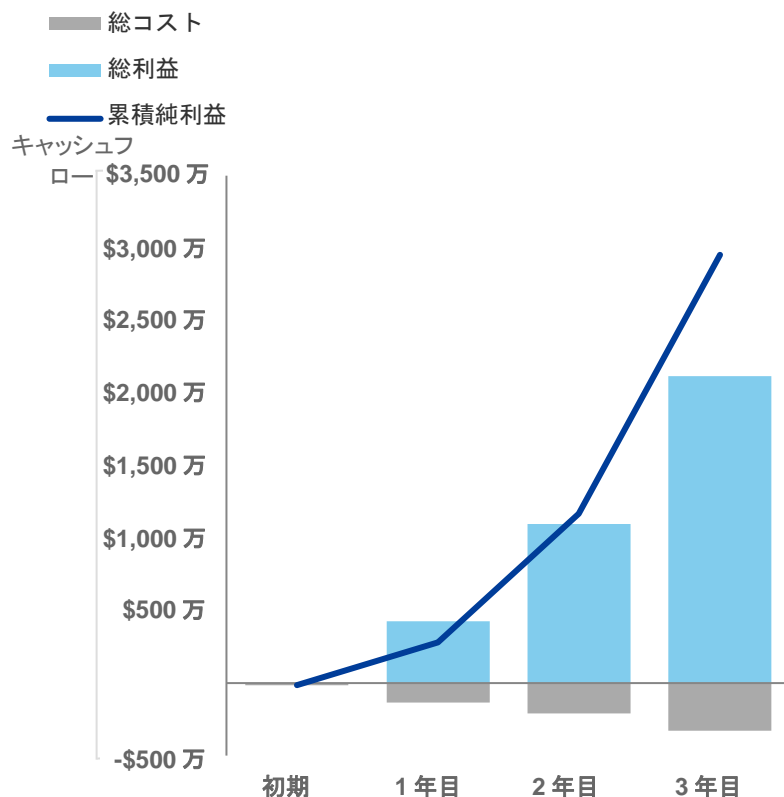
Databricks のトレーニングコスト：計算表

参照	評価項目	計算法	初期	1年目	2年目	3年目
F1	Databricks のトレーニングを受ける FTE	コンポジット	7	168	175	175
F2	トレーニング時間数	仮定	12	12	12	12
F3	FTE の混合給与	仮定	\$48	\$48	\$48	\$48
Ft	Databricks のトレーニングコスト	$F1 * F2 * F3$	\$4,032	\$96,768	\$100,800	\$100,800
	リスク調整	↑5%				
Ftr	Databricks のトレーニングコスト（リスク調整後）		\$4,234	\$101,606	\$105,840	\$105,840

財務状況の概要

リスク調整後の3年連結評価

キャッシュフローチャート（リスク調整後）



「コスト」と「利益」のセクションで計算された経済的影響を使用して、このモデル組織の投資に対するROI、NPVおよび回収期間を決定できます。Forresterは、この分析で年10%の割引率を想定しています。



これらのリスク調整後のROI、NPV、回収期間の値は、「利益」と「コスト」のセクションの未調整結果にリスク調整因子を適用することで決定されます。

キャッシュフロー分析（リスク調整後推定）

	初期	1年目	2年目	3年目	合計	現在価値
総コスト	(\$140,734)	(\$1,345,856)	(\$2,111,340)	(\$3,292,590)	(\$6,890,520)	(\$5,582,920)
総利益	\$0	\$4,278,594	\$10,980,938	\$21,171,656	\$36,431,188	\$28,871,365
純利益	(\$140,734)	\$2,932,737	\$8,869,598	\$17,879,066	\$29,540,668	\$23,288,445
投資収益率(ROI)						417%
回収期間（月）						<6

Databricks統合データ分析プラットフォーム：概観

以下の情報は、Databricksが提供したものです。Forresterはいかなる主張も検証しておらず、Databricksまたはその提供物を保証するものではありません。

Databricksはデータ主導の改革を促進します

Databricksは、データとAIを単純化して、データチームが世界の難題を解決できるようにします。オープンソースとクラウド第一の方針により、Databricksは データ、分析、およびビジネスを開かれた統合プラットフォームに集積させました。それによりデータチームは今までにない速さで協働と開発を実行できます。Comcast、Shell、StarbucksおよびRegeneronを含む世界中の数千の組織がDatabricksを利用して、社内のデータすべてを分析可能な状態にし、社内全体でのデータ主導の決定を進め、機械学習を迅速に展開しています。Databricksはベンチャー事業であり、Apache Spark™、Delta Lake、およびMLflowを含むオープンソースプロジェクトの当初の考案者によって設立されました。

Databricks統合データ分析プラットフォーム



データサイエンスワークスペース

データサイエンスワークスペースは、作業者が一か所ですべての分析プロセスを実行し、ライフサイクル全体にわたってMLモデルを管理する協働的環境です。

- コラボレーション型ノートブック：Databricksのノートブックは、Python、R、SQL、およびScalaを元からサポートしているため、作業者は自身が選択した言語とライブラリーを用いて発見や可視化を協働して行い、ステークホルダーと見識を共有できます。
- 機械学習ランタイム：予め構成されたMLクラスターにワンクリックでアクセスでき、そこではビルトインAutoMLと不一致動作の大規模な最適化により、最も一般的なMLフレームワークが拡張性と信頼性をもって配布されます。
- 管理済みの MLflow：DatabricksのオープンソースプラットフォームであるMLflowに加えて、管理済みのMLflowは、企業のセキュリティ、信頼性、および規模を基に、MLモデルの実験から生産までの管理を助けます。

統合データサービス

Databricks統合データサービスは、データパイプライン、データレイク、およびデータプラットフォームのために、信頼性と拡張性のあるプラットフォームをご提供します。データジャーニーを完全に管理することで、組織全体を通じたデータの取り込み、処理、保管、公開が可能になります。

- DatabricksのためのDelta Lake：Delta Lakeは、データの信頼性、性能、およびライフサイクル管理を改善します。もう、クリーンアップ、データレイクに加わった疑わしいデータ、またはコンプライアンスの変更によるデータ削除の困難さのために、未完の作業をロールバックする必要はありません。
- Databricksランタイム：Databricksランタイムは、最高で50xの性能が得られるApache Sparkの高度に最適化されたバージョンに構築された、配布済みのデータプロセッシングエンジンです。容易なセルフサービスとコスト削減型の動作により、パイプライン、スケジュール作業の構築、およびモデルのトレーニングを行います。
- Delta Lake上のBIレポート：Delta Lake上のBIレポートは、データレイク上でビジネス分析を行います。データレイク内にある完成された直近のデータのほとんどをDelta LakeおよびSparkSQLに直接接続し、好みのBI可視化およびレポートツールを使用して、さらにタイムリーなビジネスの洞察を行います。

企業向けクラウドサービス

企業向けクラウドサービスは、ネイティブセキュリティ、組織全体のシンプルな管理、および複数のクラウドにわたる統合データ分析プラットフォームの規模での自動化を提供します。

- 企業セキュリティ：ネイティブでのアイデンティティの連合、暗号化、およびアクセスコントロールにより、データの安全性に必要なツールが取得できます。ビルトインネットワーク、データ、および様々なコンプライアンスのオプションによる隔離作業の使用で、ユーザーに対して安全な分析環境を創出できます。
- シンプルな管理：監査証跡、ログ、および請求書作成と使用量レポートによって、運用を完全に可視化します。各チームに対して個別のワークスペースを作成し、使用量制限についての方針を設定し、確実に準拠するよう活動を分析します。
- 生産準備完了：APIファーストのアプローチにより、シームレスなCI/CDと、インフラの自動スケーリングの世界規模での自動化を保証します。使いやすいアプリケーションとインフラの監視により、SLAを維持します。

付録A : Total Economic Impact

Total Economic Impactは、Forrester Researchが開発した手法であり、企業の技術関連の意志決定プロセスを強化し、ベンダーが製品やサービスの価値をクライアントに提案するための支援を行います。TEI手法を使用することで、企業が上級管理職や他のビジネス上の主要な利害関係者に対して、ITイニシアティブの具体的な価値を実証し、正当化し、実現することができます。

Total Economic Impact 手法



利益とは、製品がビジネスにもたらす価値のことです。TEI手法では、利益の測定とコストの測定に同じ重みを与えることで、組織全体に与える技術の恩恵を徹底的に評価することが可能になります。



コストでは、提案されている製品の価値または製品の利益をもたらすために必要なすべての支出が考慮されます。TEIでのコスト区分には、ソリューションに関連して継続して発生するコストのための既存環境上の増分コストが含まれます。



柔軟性は、すでに行われた導入時の投資を基礎とした将来の追加投資で得られる戦略的価値を表します。その利益を獲得できる能力を入手することは、見積もり可能な現在価値を持つこととなります。



リスクとは、利益とコストの見積もりの不確かさを測定したもので、1) 見積もりが初期の予想に見合う可能性と、2) 時間の経過と共に見積もりが追跡される可能性が考慮されます。TEIではリスク因子は「三角分布」に基づいています。

初期投資の欄には、「時間0」または1年目の始まりに発生するコストが記載されます。これらのコストには割引率は適用されません。その他すべてのキャッシュフローには、年末の割引率が適用されます。現在価値 (PV) は、それぞれの総コストおよび利益の見積もりに対して計算されます。サマリーテーブルの正味現在価値 (NPV) は、初期投資と各年の割引後のキャッシュフローの合計になります。総利益、総コスト、キャッシュフローの各表の合計金額および現在価値については、四捨五入のため合計値が合わないことがあります。



現在価値 (PV)

特定の利率（割引率）を使用した場合の（割引された）コストおよび利益の現在価値。コストおよび利益のPVは、キャッシュフローの総正味現在価値 (NPV) に適用されます。



正味現在価値 (NPV)

特定の利率（基準割引率）における、（割引された）将来の正味キャッシュフローの現在の価値。プロジェクトの正味現在価値 (NPV) の値が正であれば、他のプロジェクトのNPVがそれより高くない限り、通常は投資すべきであると考えられます。



投資収益率 (ROI)

プロジェクトがもたらすと予想される利益をパーセント値で表した数値。ROIは、純利益（粗利益からコストを引いた値）をコストで割ることによって求められます。



割引率

キャッシュフロー分析において金銭の時間的価値を考慮するために使用する利率。通常、企業は8%~16%の割引率を使用します。



回収期間

投資した金額および投資による利益が同額になるポイント。これは純利益（粗利益から費用を引いた値）が初期投資またはコストに見合うタイミングです。

付録B：脚注

ⁱ 出典：“Build An Insights-Driven Business” Forrester Research, Inc., December 10, 2019

ⁱⁱ 出典：“The Future Of Machine Learning Is Unstoppable,” Forrester Research, Inc., April 25, 2019