

Une étude Total Economic Impact™
de Forrester commandée par Databricks
Avril 2020

Total Economic Impact™ de la plateforme Unified Data Analytics Platform de Databricks

Économies et avantages commerciaux
offerts par la plateforme Databricks

Table des matières

Synthèse	1
Principales conclusions	2
Cadre de référence et méthodologie du TEI	4
Le parcours client avec la plateforme Unified Data Analytics Platform de Databricks	5
Entreprises interrogées	5
Principaux défis	5
Exigences de la solution	7
Principaux résultats	7
Entreprise composite	10
Analyse des avantages	11
Profit supplémentaire	11
Amélioration de l'efficacité opérationnelle	13
Réduction des coûts de l'ancienne plateforme de data analysis	16
Avantages non quantifiés	18
Flexibilité	19
Analyse des coûts	20
Coûts de la plateforme Databricks, de la formation et du stockage	20
Coûts d'administration de Databricks	21
Coûts de formation sur Databricks	22
Bilan financier	24
Présentation de la plateforme Unified Data Analytics Platform de Databricks	25
Annexe A : Total Economic Impact	27
Annexe B : Notes de fin	28

Directeur du projet :
Edgar Casildo

À PROPOS DE FORRESTER CONSULTING

Forrester Consulting propose des services de conseil indépendants et objectifs, basés sur un travail de recherche, pour aider les dirigeants à réussir dans leurs entreprises. Qu'il s'agisse de courtes sessions de stratégie ou de projets sur mesure, les services de conseil de Forrester vous mettent en contact direct avec des data analysts qui mobilisent leur expertise pour répondre aux défis spécifiques de votre activité. Pour en savoir plus, rendez-vous sur forrester.com/consulting.

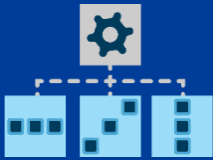
© 2020, Forrester Research, Inc. Tous droits réservés. Toute reproduction non autorisée est strictement interdite. Les informations fournies s'appuient sur les meilleures ressources disponibles. Les avis exposés reflètent un jugement à un moment donné et peuvent être amenés à évoluer. Forrester®, Technographics®, Forrester Wave, RoleView, TechRadar et Total Economic Impact sont des marques commerciales de Forrester Research, Inc. Toutes les autres marques commerciales appartiennent à leurs entreprises respectives. Pour obtenir des informations complémentaires, rendez-vous sur forrester.com.

Synthèse

Avantages



Augmentation des revenus :
9 895 520 dollars



Amélioration de l'efficacité
opérationnelle :
7 677 643 dollars



Réduction des coûts de
l'ancienne plateforme de data
analysis :
11 298 202 dollars

Forrester indique que : « Dans l'environnement hypercompétitif actuel, l'exploitation et l'application des données, de la Business Analytics et du machine learning (ML) s'imposent rapidement comme les prérequis à la réussite, afin de différencier les produits et les expériences clientsⁱ. » Tout comme Internet avant eux, l'IA et le machine learning représentent à la fois la plus grande menace et la plus grande opportunité pour les entreprises aujourd'huiⁱⁱ. Pour réussir, les entreprises devront tirer parti de l'IA et du ML pour optimiser leurs produits, leurs services et leurs opérations.

Les entreprises doivent être en mesure d'harmoniser efficacement les investissements en matière d'analyse et de ML avec les priorités de l'entreprise afin de garantir leur réussite. De même, elles doivent aligner leurs stratégies de gestion des données sur les priorités de l'entreprise et éliminer les contraintes ou les obstacles afin de garantir la réussite des équipes chargées de mettre en œuvre ces priorités.

Les entreprises doivent démocratiser les données (structurées et non structurées), faire évoluer les processus, restructurer les équipes, remettre en question les cultures et remodeler les infrastructures technologiques en vue de favoriser la réussite des équipes de données. Pour prendre des décisions en temps réel, les entreprises devront avoir accès aux données les plus à jour. Les données doivent également être ouvertes et accessibles via différents outils et systèmes au sein de l'entreprise ; elles ne doivent pas être enfermées dans des formats fermés ou des systèmes propriétaires inaccessibles à d'autres technologies et d'autres équipes. La formation des data engineers et data analysts pour leur permettre d'effectuer des travaux de data science améliorera la productivité en éliminant les goulots d'étranglement imputables à une dépendance excessive aux data scientists.

La mise en œuvre de tous ces composants à grande échelle est aujourd'hui hors de portée de la plupart des entreprises, compte tenu de la complexité et du coût de ces impératifs. C'est pourquoi les entreprises adoptent rapidement la plateforme de données unifiée et ouverte de Databricks, qui simplifie les données et l'IA pour le data engineering à grande échelle, la data science collaborative, le machine learning appliqué à l'ensemble du cycle de vie, et la Business Analytics. Databricks a chargé Forrester Consulting de mener une étude Total Economic Impact™ (TEI) auprès des clients actuels de Databricks et de déterminer le retour sur investissement (ROI) que les entreprises peuvent réaliser en déployant la plateforme Unified Data Analytics Platform de Databricks. Cette étude a pour but de fournir aux lecteurs un cadre de référence qui leur permet d'évaluer l'impact financier potentiel de l'utilisation de la plateforme Databricks dans leur entreprise.

Avant d'utiliser Databricks, les clients disposaient de différentes technologies de traitement des données, de Business Analytics et de ML, aussi bien sur site que dans le Cloud. Le cloisonnement des données, la difficulté d'utilisation des systèmes, les frictions opérationnelles et le manque d'évolutivité ont entravé la capacité des équipes chargées des données à collaborer efficacement pour maximiser la valeur de leurs données. Les équipes chargées des données passaient la plupart de leur temps à mettre en place et gérer des systèmes au lieu de stimuler l'innovation et les résultats opérationnels.

Principales conclusions

Avantages quantifiés. Les avantages suivants, quantifiés en valeur actuelle (VA) pondérée en fonction des risques, sont représentatifs de ceux des entreprises interrogées :



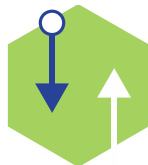
ROI
417 %



Avantages en VA
28,9 millions de dollars



VAN
23,3 millions de dollars



Délai de rentabilité
<6 mois

- › **Augmentation des revenus de 5 %.** Databricks a permis aux équipes de data scientists des clients interrogés de consacrer plus de temps à la création et l'amélioration des modèles ML. Databricks a également permis aux data scientists d'utiliser des modèles ML de pointe, dont les modèles de deep learning, qui ne leur étaient pas accessibles auparavant. En démocratisant l'accès aux données, les entreprises interrogées ont également permis à de nouveaux utilisateurs de créer un ensemble diversifié de nouveaux modèles ML et d'en tirer plus d'insights qu'auparavant. Cette combinaison de modèles ML plus nombreux et plus performants a entraîné une augmentation des revenus pour la plupart des entreprises interrogées.
- › **Amélioration de la productivité des équipes chargées des données de 25 % et 20 %, respectivement.** Databricks a permis à des équipes de data scientists et de data engineers de passer moins de temps à rechercher les données, à les nettoyer, ainsi qu'à créer et entretenir des pipelines d'extraction, de transformation et de chargement (ETL), et de passer plus de temps à développer et améliorer des modèles ML susceptibles d'améliorer les résultats. Databricks a également contribué à éliminer les obstacles techniques qui limitaient la collaboration entre les data analysts, les data scientists et les data engineers, permettant aux équipes chargées des données de travailler plus efficacement ensemble.
- › **Retrait des infrastructures sur site, permettant des économies de plusieurs millions de dollars par an.** En passant à Databricks, les entreprises interrogées ont pu mettre fin à leurs infrastructures sur site et annuler leurs licences de data analysis, désormais redondantes.

Avantages non quantifiés. Les entreprises interrogées ont constaté les bénéfices suivants, non quantifiés pour cette étude :

- › **Réduction des coûts d'exploitation.** Les différentes fonctionnalités offertes par Databricks ont permis aux entreprises interrogées de réduire leurs dépenses en matières premières et d'identifier plus rapidement les problèmes de fabrication, ce qui a permis de diminuer les coûts de réparation et les taux de retour.
- › **Meilleure sécurité.** Avant Databricks, les entreprises interrogées ne disposaient pas de normes de sécurité dans leurs environnements de data analysis. Databricks a fourni une protection native, ce qui a permis aux clients interrogés d'atténuer les problèmes de sécurité et de concentrer leurs ressources sur d'autres activités.

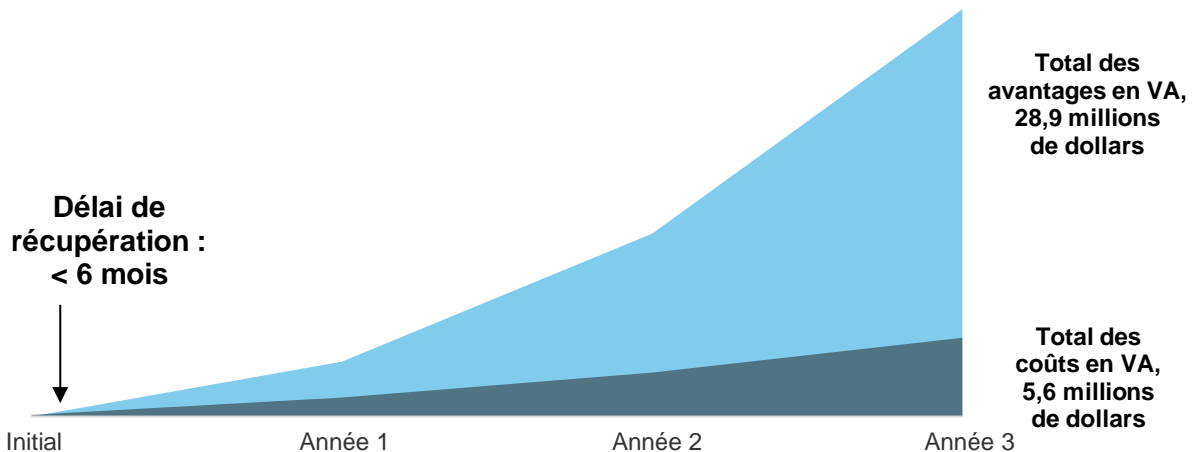
Coûts. Les entreprises interrogées ont constaté les coûts en VA pondérés en fonction des risques suivants :

- › **Les coûts de la plateforme Unified Data Analytics Platform de Databricks, de la formation et du stockage s'élèvent à 4,5 millions de dollars.** Il s'agit des coûts annuels d'utilisation et de formation que les utilisateurs paient à Databricks, ainsi que des coûts de stockage dans le Cloud d'un tiers.

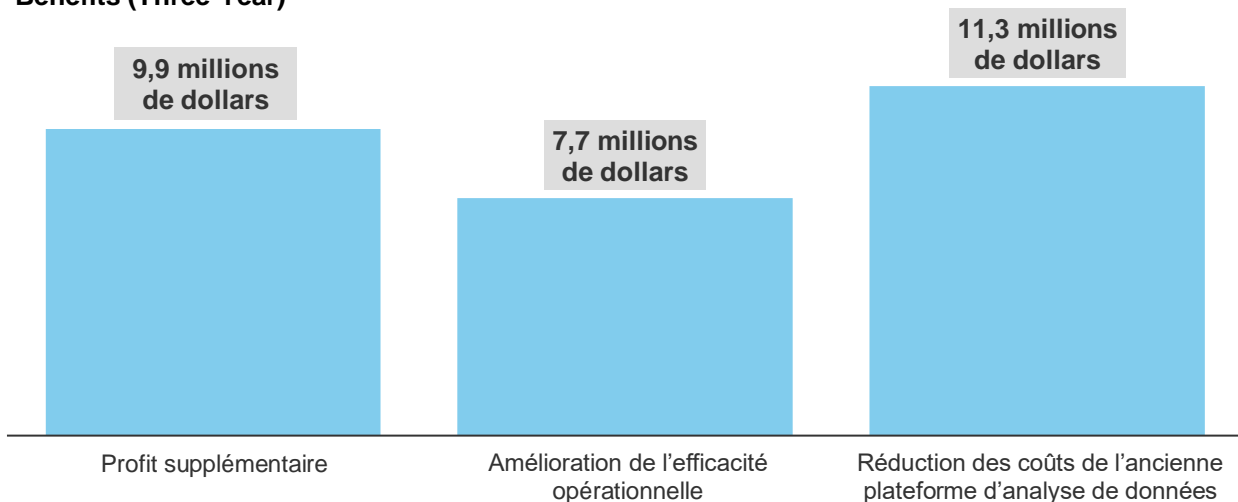
- › **Les coûts d'administration de Databricks s'élèvent à 815 400 dollars.** Il s'agit des salaires toutes charges comprises versés par l'entreprise pour gérer la mise en œuvre et l'exploitation de la plateforme.
- › **Les coûts de formation de la main-d'œuvre interne pour Databricks s'élèvent à 263 600 dollars.** Il s'agit des salaires toutes charges comprises de l'entreprise pour former les utilisateurs sur la plateforme.

Les entretiens de Forrester avec quatre entreprises clientes et l'analyse financière ultérieure ont montré qu'une entreprise composite analogue aux entreprises interrogées bénéficierait sur trois ans d'avantages se montant à 28,9 millions de dollars tout en assumant des coûts de 5,6 millions de dollars. Elle en retirerait donc une valeur actuelle nette (VAN) de 23,3 millions de dollars et un ROI de 417 %.

Bilan financier



Benefits (Three-Year)



La méthodologie TEI aide les entreprises à démontrer, justifier et concrétiser la valeur réelle des initiatives en matière de technologies de l'information, à la fois auprès de la direction et auprès des autres parties prenantes clés internes.

Cadre de référence et méthodologie du TEI

À partir des informations collectées dans ses enquêtes, Forrester a créé un cadre de référence Total Economic Impact™ (TEI) pour ces entreprises en ce qui concerne la mise en œuvre de la plateforme Unified Data Analytics Platform de Databricks.

L'objectif du cadre de référence est d'identifier les coûts, les avantages, la flexibilité et les facteurs de risque qui entrent en jeu dans la décision d'investissement. Forrester a utilisé une approche en plusieurs étapes pour évaluer l'impact que la plateforme Databricks peut avoir sur une entreprise :



CONTRÔLE PRÉALABLE

Nous nous sommes entretenus avec des intervenants de Databricks et les data analysts de Forrester pour recueillir des données relatives à Databricks.



ENTRETIENS AVEC DES CLIENTS

Nous avons enquêté auprès de quatre entreprises qui utilisent la plateforme pour obtenir des données concernant les coûts, les avantages et les risques.



ENTREPRISE COMPOSITE

Nous avons conçu une entreprise de référence d'après les caractéristiques des entreprises interrogées dans le cadre de l'enquête.



CADRE DE RÉFÉRENCE DU MODÈLE FINANCIER

Nous avons créé un modèle financier représentatif des entretiens à l'aide de la méthodologie TEI et pondéré en fonction des risques du modèle financier sur la base des problèmes et préoccupations des entreprises interrogées.



ÉTUDE DE CAS

Nous avons utilisé quatre éléments fondamentaux du TEI pour modéliser l'impact de Databricks : avantages, coûts, souplesse et risques. Compte tenu de la sophistication croissante des entreprises en matière d'analyses du ROI relatif à leurs investissements informatiques, la méthodologie TEI de Forrester offre un aperçu exhaustif de l'impact économique total des décisions d'achat. Veuillez vous reporter à l'annexe A pour des informations complémentaires sur la méthodologie TEI.

DÉCLARATIONS

Les lecteurs doivent être avisés de ce qui suit :

L'étude est commandée par Databricks et réalisée par Forrester Consulting. Elle n'est pas destinée à être utilisée comme analyse concurrentielle.

Forrester n'établit aucun postulat concernant le retour sur investissement potentiel que d'autres entreprises pourraient connaître. Forrester recommande vivement aux lecteurs d'utiliser leurs propres estimations dans les limites du cadre de référence fourni dans le rapport pour déterminer le caractère approprié ou non d'un investissement dans la plateforme Unified Data Analytics Platform de Databricks.

Databricks a relu l'étude et fourni des commentaires à Forrester, mais Forrester garde le contrôle éditorial de l'étude et de ses conclusions et n'accepte pas de modifications de l'étude qui contrediraient les conclusions de Forrester ou occulteraient le propos de l'étude.

Databricks a fourni les noms des clients pour les entretiens, mais n'y a pas pris part.

Le parcours client avec la plateforme Unified Data Analytics Platform de Databricks

AVANT ET APRÈS L'INVESTISSEMENT DANS LA PLATEFORME

Entreprises interrogées

Pour les besoins de cette étude, Forrester a mené des entretiens avec quatre clients de Databricks. Les entreprises interrogées présentent les caractéristiques suivantes :

SECTEUR	RÉGION	PERSONNE INTERROGÉE	ENVIRONNEMENT
Commerce de détail	Siège social dans la zone EMEA	Responsable de l'équipe de data science	<ul style="list-style-type: none">• 9 data scientists• 220 To de données• 450 utilisateurs de Databricks
Secteur pharmaceutique	Siège en Amérique du Nord	Manager, datalake et analyses	<ul style="list-style-type: none">• 100 data scientists• 300 To de données• 300 utilisateurs de Databricks
Équipements lourds	Siège en Amérique du Nord	Responsable de l'architecture, datalake et analyses	<ul style="list-style-type: none">• 10 data scientists• 120 To de données• 120 utilisateurs de Databricks
Médias	Siège en Amérique du Nord	Vice-président, data science	<ul style="list-style-type: none">• 200 data scientists• 10 To de données• 400 utilisateurs de Databricks

Principaux défis

Les clients interrogés savaient que leurs entreprises devaient être davantage axées sur les données pour obtenir de meilleurs résultats commerciaux. Leurs infrastructures technologiques existantes se sont cependant avérées être plus des obstacles que des atouts. Les entreprises interrogées ont cité les problèmes suivants concernant leurs infrastructures technologiques existantes :

- › **Le cloisonnement des données ralentissait le processus de développement, conduisait à des modèles ML moins précis et faisait perdre du temps aux équipes chargées des données.** Les data scientists et les data engineers avaient des difficultés à recueillir les données pertinentes pour lancer de nouveaux projets. Comme l'a souligné le responsable de l'architecture du secteur des équipements lourds : « Les données étaient dispersées dans toute l'entreprise. Il fallait connaître quelqu'un qui connaissait quelqu'un pour savoir où se trouvaient les données. Nous ne pouvions utiliser que ce que nous avons, et cela ne simplifiait pas le changement résultant de nos recherches. » Les données limitées et le temps relativement restreint que les data scientists pouvaient consacrer aux modèles ML ou à l'obtention d'insights réduisaient leur efficacité globale.

« Il y avait un manque de volonté de faire quoi que ce soit sur notre environnement précédent parce qu'il était très lent. »

Responsable de l'architecture, datalake et analyses, fabricant d'équipements lourds



› **Les infrastructures existantes ralentissaient les projets ou, au pire, les bloquaient.** L'infrastructure limitée et statique sur site empêchait les collaborateurs de lancer de nouveaux projets et ralentissait l'avancement des projets déjà lancés. Les environnements sur site des entreprises interrogées :

- **Ne proposaient pas la puissance de calcul nécessaire pour traiter des ensembles de données complets.** Le vice-président de l'équipe de data science de l'entreprise de médias a souligné : « Il nous était impossible d'analyser nos ensembles de données extrêmement volumineux. Nous étions obligés d'utiliser nos serveurs [existants] coûteux sur site ; nous avons besoin de quelque chose qui puisse s'adapter à nos besoins. » Les équipes chargées des données étaient contraintes de travailler avec des ensembles limités de données, ce qui les empêchait d'extraire des insights significatifs des essais ou des analyses.

- **Ne pouvaient s'adapter aux périodes de forte demande.** Le responsable de l'architecture du fabricant a indiqué : « Nous n'avons pas de réelle évolutivité ou élasticité intégrée pour répondre à nos besoins pendant notre haute saison. Pendant ces périodes, nous étions préoccupés. Est-ce que nos serveurs [sur site] allaient être capables de survivre ou de traiter les volumes de données qui arrivaient. »

Le manager de l'entreprise pharmaceutique s'est fait l'écho de ces préoccupations : « Il existe plusieurs cas d'utilisation qui n'ont pas leur place dans un environnement statique en cluster. Par exemple, nous obtenons des ensembles de données sur une base hebdomadaire ou mensuelle. Idéalement, nous pourrions lancer un ou plusieurs clusters pour traiter les ensembles de données le plus rapidement possible, avant de transmettre les insights à nos utilisateurs. Mais c'est impossible avec un cluster statique. »

› **Avaient des difficultés à suivre l'augmentation des coûts sur site et dans le Cloud.** Le responsable du data lake et des analyses de la société pharmaceutique a indiqué : « Le coût de notre cluster [sur site], tant du point de vue de l'infrastructure que des licences, a augmenté de façon spectaculaire. » D'autres responsables des données interrogés ont indiqué qu'ils devaient continuellement acheter de nouveaux serveurs sur site pour répondre aux besoins croissants de leur entreprise. Les besoins toujours croissants en matière d'infrastructure et de maintenance liés à cette croissance se sont avérés décourageants pour certaines entreprises.

Les entreprises qui sont passées au Cloud ont continué à voir leurs coûts de calcul et de stockage augmenter. Un responsable a souligné que les coûts de stockage de son entreprise ont augmenté de 16 % par an.

› **Manquaient d'outils de collaboration efficaces pour stimuler les résultats commerciaux.** Les équipes de data science, d'ingénierie et de data analysis ne disposaient pas auparavant d'outils de collaboration modernes dans leurs piles de développement, et devaient donc souvent travailler indépendamment sur des projets. Le responsable de l'équipe de data science du retailer a expliqué : « Avant, quelqu'un développait quelque chose et on espérait que cela marcherait. Puis, quand les choses tournaient mal, personne ne s'occupait de corriger les erreurs. » La création de modèles ML

« Nos coûts d'infrastructure et de licences [sur site] augmentaient de façon assez spectaculaire à mesure que nos besoins en stockage augmentaient, mais nos besoins en puissance de calcul ne changeaient pas vraiment. La gestion de cette infrastructure croissante était assez décourageante. »

Manager, data lake et analyses, entreprise pharmaceutique



efficaces qui permettent d'obtenir des résultats nécessite une collaboration entre les data scientists, les data engineers, les développeurs d'applications et les intervenants commerciaux. S'ils ne disposent pas des bons outils pour collaborer, les équipes de data scientists ne peuvent créer efficacement des modèles, ni les adapter par la suite. De plus, sans documentation adéquate, les équipes de data science risquent de créer des modèles que personne ne comprend, ce qui les oblige à passer beaucoup de temps à comprendre les modèles ou à les refaire.

Exigences de la solution

Les entreprises interrogées recherchaient une solution capable de :

- › Fournir une infrastructure technologique moderne pour favoriser la collaboration et stimuler les résultats commerciaux.
- › Dégager du temps pour les data scientists et les data analysts afin qu'ils puissent se concentrer sur des activités à plus forte valeur ajoutée.
- › Permettre une transition vers le Cloud.
- › Maintenir la fiabilité et la performance des données grâce aux politiques de gouvernance améliorées.

Principaux résultats

Les entretiens ont révélé que les résultats clés de l'investissement dans la plateforme Databricks sont :

- › **La démocratisation de l'accès aux données a entraîné différents résultats.** La possibilité pour les collaborateurs dans toute l'entreprise de tirer parti de la plateforme Databricks pour fournir un ensemble varié de perspectives et de compétences. Ces nouveaux utilisateurs ont permis d'augmenter les revenus, de réduire les coûts et d'améliorer les performances.

- **L'accélération de la mise sur le marché.** L'adoption de Databricks a permis aux entreprises interrogées de s'attaquer aux goulots d'étranglement dans leurs environnements existants. Le responsable de l'équipe de data science du retailer a souligné : « Nos collaborateurs du service marketing peuvent désormais développer eux-mêmes des logiciels et des outils, et les mettre en production, sans avoir à faire appel à des développeurs ou des data scientists. [Cela nous a permis] d'accélérer les délais de mise sur le marché tout en réduisant les ressources nécessaires. »

- **L'augmentation des revenus grâce à l'identification de nouvelles opportunités.** L'équipe marketing du retailer a identifié la possibilité d'augmenter les ventes en prédisant les retours sur les articles épuisés. Le responsable de l'équipe de data science a déclaré que l'équipe marketing « avait développé une solution par ses propres moyens grâce aux Notebooks Databricks. »

Le responsable de l'architecture s'est fait l'écho de ces points et a déclaré : « Nous accélérons les délais de mise sur le marché parce que je suis maintenant en mesure de faire des choses qu'il était difficile de réaliser auparavant. Et cela fonctionne, de manière performante et adaptée aux besoins.

« Les bases de données réduisent les frictions, démocratisent le contrôle et permettent à la data science de faire de la R&D sans aucun blocage. »

Vice-président, data science, médias



« Nous avions auparavant un nombre restreint de data analysts qui savaient comment tout fonctionnait ; tout passait donc par eux. Mais maintenant, tout le monde apprend. Tous les collaborateurs de l'entreprise veulent utiliser Databricks. »

Responsable de l'équipe de data science, retail



Je peux traiter plus de données que je ne pouvais le faire auparavant. Je peux traiter des ensembles de données plus volumineux et plus complexes qu'il m'était impossible de traiter auparavant. J'ai ouvert des perspectives dont je ne disposais pas auparavant, et je rencontre aussi moins de problèmes. Ma production est mise sur le marché plus rapidement. Mes idées sont plus rapides. Je suis en mesure d'aborder d'autres idées et d'autres problèmes que je n'aurais pas eu le temps d'aborder par le passé, parce que j'avais juste le temps de [terminer] les bases sur lesquelles je travaillais. »

- **Forts de leurs nouvelles compétences, les collaborateurs ont pu découvrir des opportunités supplémentaires tout en réduisant la charge de travail des équipes de data science.**

Le responsable de l'équipe de data science a expliqué qu'un membre de l'équipe marketing avait identifié un moyen d'optimiser certaines activités par un facteur de 25 %. « Ils n'avaient aucune expérience en matière de data science ou de data engineering. Ils se sont simplement attelés à la tâche. Ils ont créé un article de blog avec un bel exemple illustrant comment tout le monde pourrait faire de même. »

Le responsable de l'architecture du fabricant d'équipements lourds a déclaré que l'adoption de Databricks a permis un changement de mentalité au sein de son entreprise. « Notre équipe d'assistance à la clientèle essaie de trouver comment utiliser les informations dont nous disposons afin de mieux prendre en charge nos clients et nos machines sur le terrain. À cette fin, elle crée des modèles permettant d'identifier les premiers signes de problèmes afin d'alerter nos concessionnaires pour qu'ils effectuent des réparations de manière proactive. Notre équipe d'ingénieurs se demande comment utiliser nos informations pour améliorer le processus de fabrication. Le service des pièces détachées réfléchit à l'automatisation de ses processus manuels. » Bien qu'il soit encore trop tôt pour mesurer l'impact de certaines de ces initiatives, le responsable de l'architecture a indiqué que le service de tarification des pièces a déjà généré des millions de dollars de revenus supplémentaires en optimisant uniquement le prix d'une partie relativement faible du catalogue global de l'entreprise.

- › **L'amélioration de la collaboration et de l'expérimentation a conduit à de meilleurs produits ML, qui génèrent des revenus et réduisent les coûts.** La plateforme Databricks a permis aux clients interrogés d'accélérer leurs cycles de développement et d'affiner leurs modèles ML au fil du temps. Le responsable de l'équipe de data science du retailer a indiqué : « Databricks nous permet de réduire le temps nécessaire pour développer un modèle ML, vérifier qu'il fonctionne, puis demander à quelqu'un d'autre de l'examiner et de l'améliorer, et de le suivre au fil du temps. » Grâce à ces moyens supplémentaires, les entreprises interrogées ont pu créer ou optimiser les applications suivantes dans leurs entreprises :

- L'entreprise pharmaceutique a créé un moteur d'IA pour identifier les clients potentiels sur lesquels les commerciaux devraient concentrer leurs efforts, et quand le faire. L'entreprise a également découvert qu'elle payait trop pour les matières premières, ce qui lui a permis de renégocier ses contrats et d'économiser des dizaines de millions de dollars.

« Nous avons constaté d'énormes avantages dans toute l'entreprise. Les gens dépendent moins des développeurs et des data analysts, et sont plus autonomes. Ils peuvent mener leurs projets par leurs propres moyens. »

Responsable de l'équipe de data science, retail



- Le retailer a augmenté ses ventes grâce à l'amélioration de son moteur de recherche et la création de moteurs de recommandation pour différents types de clients. Il a également pu réduire ses coûts d'exploitation en réduisant le volume de retours et en identifiant les commandes frauduleuses.
 - Le fabricant d'équipements lourds a amélioré sa capacité à identifier les problèmes de fabrication, réduisant ainsi les temps d'arrêt pour ses clients et les coûts de réparation.
- › **La création d'un plus grand nombre de modèles ML.** L'accès aux bibliothèques ML, les fonctions de suivi des expérimentations via l'inclusion native de MLflow (une plateforme open source de gestion du cycle de vie ML) et les Notebooks collaboratifs ont aidé les entreprises interrogées à accélérer la création de modèles ML. En démocratisant l'accès à la plateforme Databricks, davantage de personnes au sein des entreprises ont pu contribuer à la création de modèles et d'algorithmes ML.
- › **Des environnements plus stables et plus performants.** Les personnes interrogées indiquent que les processus sur la plateforme Databricks se terminent bien plus rapidement qu'avec leurs solutions précédentes. Par exemple, le responsable de l'architecture a noté une diminution de 40 % du temps de traitement et une diminution de 97 % du temps nécessaire à l'exécution complète des processus Apache Spark (un moteur de calcul distribué polyvalent et open source utilisé pour le traitement et l'analyse de grandes quantités de données). Le responsable de l'architecture a également observé une réduction du nombre de travaux incorrects, passant d'un taux d'échec de 0,4 à 1,0 % avant Databricks à un taux d'échec de 0,1 % avec Databricks. Le vice-président de l'équipe de data science a noté que l'exécution complète des processus prenaient une semaine avant Databricks, contre une heure et demie avec Databricks. L'amélioration des performances a permis aux data scientists et aux data analysts de réaliser davantage d'expérimentations dans le cadre de cycles rapides, et a permis aux entreprises de mieux comprendre leurs modèles.
- › **La réduction des infrastructures et la simplification de l'administration.** En passant à Databricks, les entreprises interrogées ont pu mettre fin à leurs infrastructures sur site, annuler leurs licences logicielles redondantes, et réallouer leurs ressources informatiques. L'administration de la plateforme s'est avérée beaucoup plus facile qu'avec les environnements précédents. Les entreprises étaient en mesure de passer à des environnements DevOps, ce qui pouvait simplifier encore plus l'administration de leur environnement. Le responsable du datalake et de l'analyse pour l'entreprise pharmaceutique a déclaré : « [Databricks] a énormément facilité les choses. [Personne] n'a envie de passer son temps à s'occuper des finances et les gens se plaignent des mises à niveau ou de la baisse des performances. [Avec Databricks], nous avons adopté les DevOps : chaque équipe a son propre cluster qu'elle paie. Elle gère la taille de son cluster et le fait évoluer quand elle le veut. »

« Nous devons auparavant attendre deux semaines pour qu'un processus se termine avant de pouvoir analyser un modèle ML et l'adapter. Ce travail prend désormais une heure. Nous pouvons l'adapter plusieurs fois par jour, ce qui nous permet de vraiment comprendre les résultats de notre recherche, l'affiner et créer un nouveau modèle sans interruption. »

Vice-président, data science, médias



- » **Une expertise inestimable dans le domaine de la biotechnologie et de la génétique.** Le responsable du datalake et de l'analyse de l'entreprise pharmaceutique a déclaré : « Databricks est vraiment utile dans le domaine de la santé parce que l'entreprise emploie des programmeurs Spark et des généticiens. Au lieu de passer tout notre temps à essayer de comprendre quelque chose par nous-mêmes, nous pouvons nous tourner vers Databricks. L'entreprise comprend ce que nous essayons de faire avec le séquençage de l'ADN parce qu'elle a investi dans le domaine de la biotechnologie. »

Entreprise composite

À partir des entretiens, Forrester a établi un cadre de référence pour ce TEI, une société composite et l'analyse de son ROI illustrant les domaines affectés sur le plan financier. L'entreprise composite est représentative des quatre sociétés interrogées par Forrester. Elle est utilisée pour présenter l'analyse financière sous forme agrégée dans la section suivante. L'entreprise de référence synthétisée par Forrester à partir de ses entretiens avec les clients présente les caractéristiques suivantes :

Description de l'entreprise composite. L'entreprise composite est une entreprise mondiale dans le secteur des produits et des services B2C et B2B. Elle a son siège aux États-Unis et génère un chiffre d'affaires annuel de 5 milliards de dollars.

L'entreprise dispose actuellement d'importants moyens de data analysis sur site : 100 data scientists, 200 data engineers et 450 To de données. Elle dispose d'un ensemble de modèles ML qui prennent en charge ses moteurs de recommandation et son moteur de recherche.

Caractéristiques du déploiement. L'entreprise composite passe progressivement à la plateforme Databricks sur une période de trois ans, en transférant chaque année 25 % de ses données, des data scientists et des data engineers. L'entreprise intègre également 100 nouveaux utilisateurs chaque année : notamment des spécialistes du marketing, des business analysts, des développeurs d'applications et des chefs de produits.

Dans le cadre de la migration, l'entreprise composite adopte Delta Lake (une couche de stockage open-source qui intègre des transactions ACID dans Apache Spark™ et des charges de travail du Big Data) pour se conformer aux normes de gouvernance et assurer la performance de ses données. L'entreprise utilise les fonctionnalités de Databricks pour affiner ses modèles ML existants et en créer de nouveaux. Parmi les exemples de nouveaux modèles ML figurent un moteur d'optimisation des prix, l'analyse de la détection des fraudes, l'optimisation des processus de fabrication et la maintenance prédictive.



Hypothèses clés

5 milliards de dollars de chiffre d'affaires

100 data scientists

200 data engineers

525 utilisateurs de

Databricks à l'Année 3

Analyse des avantages

DONNEES QUANTIFIEES SUR LES AVANTAGES APPLIQUEES A L'ENTREPRISE COMPOSITE

Total des avantages

RÉF.	AVANTAGE	ANNÉE 1	ANNÉE 2	ANNÉE 3	TOTAL	VALEUR ACTUELLE
Atr	Profit supplémentaire	843 750 dollars	3 375 000 dollars	8 437 500 dollars	12 656 250 dollars	9 895 520 dollars
Btr	Amélioration de l'efficacité opérationnelle	1 594 219 dollars	3 188 438 dollars	4 782 656 dollars	9 565 313 dollars	7 677 643 dollars
Ctr	Réduction des coûts de l'ancienne plateforme de data analysis	1 840 625 dollars	4 417 500 dollars	7 951 500 dollars	14 209 625 dollars	11 298 202 dollars
	Total des avantages (pondéré en fonction des risques)	4 278 594 dollars	10 980 938 dollars	21 171 656 dollars	36 431 188 dollars	28 871 365 dollars

Profit supplémentaire

La plateforme Unified Data Analytics Platform de Databricks a permis aux clients interrogés d'augmenter leurs revenus en leur fournissant les outils nécessaires pour créer de meilleurs modèles ML et de meilleures analyses, plus rapidement qu'auparavant. Avec Databricks, les équipes de data scientists et de data engineers passent moins de temps à nettoyer les données, ainsi qu'à créer et entretenir des pipelines d'ETL, et passent plus de temps à développer et améliorer des modèles ML susceptibles d'améliorer les résultats. Les entreprises interrogées ont généré des profits supplémentaires grâce à :

- › **L'augmentation des ventes via l'amélioration de l'expérience client (CX).** En augmentant le nombre et l'efficacité de ses moteurs de recommandation pour les différents types d'acheteurs, le retailer a pu augmenter la fréquence des commandes. Il a calculé que chaque pourcentage d'augmentation de la fréquence des commandes se traduisait par une augmentation de 2,8 % du chiffre d'affaires. Le retailer a également pu identifier les problèmes qui entraînaient la perte des clients. Par exemple, il a constaté que plus d'un tiers des abandons de paniers d'achat étaient dus à la création forcée de comptes. En éliminant l'obligation de créer un compte, le retailer a pu augmenter son chiffre d'affaires de 4 %.
- › **L'optimisation des prix.** Le responsable de l'architecture du fabricant d'équipements lourds a indiqué : « Notre équipe chargée de la tarification des pièces est passée d'un processus manuel à un processus plus automatisé à l'aide de Databricks. Elle a estimé que l'optimisation des prix a déjà permis de réaliser un bénéfice de plusieurs millions de dollars, bien qu'elle ne se soit pas encore

Le tableau ci-dessus présente le total de tous les avantages dans les domaines listés ci-dessous, ainsi que les valeurs actuelles (VA) sur la base d'un taux d'actualisation de 10 %. Sur une durée de trois ans, l'entreprise composite prévoit que le total des avantages pondéré en fonction des risques atteigne une valeur actuelle de plus de 28 millions de dollars.

occupée du prix de toutes les pièces. Nous pensons donc que cette opportunité dépassera largement les quelques millions de dollars par an. »

- › **L'augmentation des moyens des commerciaux.** L'entreprise pharmaceutique a créé un assistant d'IA qui a optimisé les activités des commerciaux en hiérarchisant les clients. « Il existe toujours plus de prescripteurs que de commerciaux. Nous avons besoin d'un moyen de hiérarchiser les personnes avec lesquelles nous nouons le dialogue... En créant un assistant d'IA pour les représentants des ventes, nous les aidons à mieux optimiser leur temps, ce qui se traduit par des revenus plus élevés. »
- › **Les modèles ML sont mis en production plus rapidement qu'auparavant.** Bien que les personnes interrogées aient eu du mal à quantifier la vitesse à laquelle elles allaient commercialiser leurs produits grâce à Databricks, elles ont convenu que le délai de commercialisation avait un impact sur les avantages financiers qu'elles imputaient à Databricks. Les entreprises interrogées ont pu concrétiser les avantages financiers de leurs modèles ML plus rapidement qu'auparavant. Elles ont également pu développer davantage de modèles ML au cours de la même période, entraînant des avantages financiers supplémentaires. Au lieu de créer un seul modèle ML entraînant une augmentation des revenus en un an, une entreprise était en mesure de lancer plusieurs modèles ML au cours de la même période.

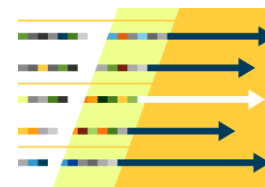
En modélisant cet avantage, Forrester a fait les hypothèses suivantes :

- › L'entreprise composite augmente son chiffre d'affaires en améliorant son taux de conversion, en optimisant ses tarifs et ses offres par la création de nouveaux modèles ML et l'optimisation des modèles ML existants.
- › Au fur et à mesure que les utilisateurs de Databricks collaborent et affinent leurs modèles ML, et découvrent de nouveaux insights, les revenus augmentent grâce aux composés de Databricks.
- › Comme Databricks permet à l'entreprise composite de développer et de publier davantage de modèles ML plus rapidement qu'elle ne le pouvait auparavant, l'entreprise composite comptabilise ces augmentations de revenus plus tôt.

L'amélioration du profit variera en fonction de :

- › La possibilité pour les utilisateurs de Databricks de créer de nouveaux modèles ML et d'extraire des informations de leurs données.
- › La possibilité pour les utilisateurs de Databricks d'affiner leurs modèles ML.
- › La marge d'exploitation d'une entreprise.
- › Le degré de hiérarchisation des initiatives générant une augmentation du chiffre d'affaires.

Pour tenir compte de ces risques, Forrester a réduit cet avantage de 10 %, ce qui a donné une valeur actuelle (VA) pondérée en fonction des risques de 9 895 520 dollars sur trois ans.



La création et l'optimisation plus rapides des modèles ML à l'aide de Databricks entraînent un profit supplémentaire.

Le risque d'impact est le risque que l'investissement ne réponde pas aux besoins technologiques ou métiers de l'entreprise, ce qui entraînerait des bénéfices totaux globalement inférieurs. Plus l'incertitude est grande, plus la fourchette de résultats possibles est large pour l'estimation des bénéfices.

Profit supplémentaire : tableau de calcul

RÉF.	INDICATEUR	CALCUL	ANNÉE 1	ANNÉE 2	ANNÉE 3
A1	Chiffre d'affaires annuel	Entreprise composite	5 000 000 000 dollars	5 000 000 000 dollars	5 000 000 000 dollars
A2	Pourcentage de l'entreprise/du chiffre d'affaires affecté par Databricks	Entreprise composite	25 %	50 %	75 %
A3	Augmentation du chiffre d'affaires imputable à Databricks	Hypothèse	1,50 %	3,00 %	5,00 %
A4	Augmentation du chiffre d'affaires imputable à Databricks	$A1 \cdot A2 \cdot A3$	18 750 000 dollars	75 000 000 dollars	187 500 000 dollars
A5	Marge d'exploitation	NYU Stern	5 %	5 %	5 %
At	Profit supplémentaire	$A4 \cdot A5$	937 500 dollars	3 750 000 dollars	9 375 000 dollars
	Pondération des risques	↓10 %			
Atr	Profit supplémentaire (pondéré en fonction de risques)		843 750 dollars	3 375 000 dollars	8 437 500 dollars

Amélioration de l'efficacité opérationnelle

Avant de migrer vers la plateforme Databricks, les équipes de data analysis, de data science et de data engineering passaient la plupart de leur temps sur des activités à faible valeur ajoutée autres que la création de modèles ML et l'obtention d'insights à partir des données. Les data analysts, les data engineers et les data scientists consacraient la majeure partie de leur temps :

- › **Rassemblement et nettoyage des données.** Les équipes de data analysis et de data science pouvaient passer des semaines à préparer leurs expérimentations sur les ensembles de données. Elles passaient beaucoup de temps à rechercher les données dans les silos et à les nettoyer.
- › **Création d'environnements de test et de production.** Comme les équipes de data science et de data engineering travaillaient de manière cloisonnée, il était difficile de collaborer sur un projet. Les data scientists et les data engineers utilisaient des langages de programmation et des méthodologies différents, et ils ne se comprenaient pas entre eux. Par conséquent, la création d'un nouvel environnement de test ou de production nécessitait de nombreux allers-retours.

« Ma production est mise sur le marché plus rapidement. Mes idées sont plus rapides. Je suis en mesure d'aborder d'autres idées et d'autres problèmes que je n'aurais pas eu le temps d'aborder par le passé, parce que j'avais juste le temps de terminer le travail de base. »

Responsable de l'architecture, datalake et analyses, fabricant d'équipements lourds



Le vice-président de l'équipe de data science de l'entreprise de médias a souligné les difficultés rencontrées par son équipe :
« Historiquement, nous faisons toute notre R&D dans notre environnement, nous rédigeons un cahier des charges, puis nous l'envoyons à l'ingénierie pour qu'elle crée un environnement de développement. Cela nécessitait beaucoup d'allers-retours car l'ingénierie ne comprenait pas vraiment ce que nous lui demandions de mettre en œuvre. C'était un paradigme de développement vraiment lent et fastidieux. »

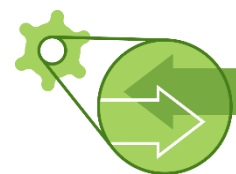
Comme les data scientists devaient consacrer beaucoup de temps aux tâches ci-dessus, ils avaient relativement peu de temps à consacrer aux activités à valeur ajoutée. Les data scientists passaient une petite fraction de leur temps à peaufiner leurs modèles ML, collaborer avec les intervenants ou travailler sur de nouveaux projets. Pour de nombreux clients interrogés, leur infrastructure ne faisait qu'aggraver ce problème. En raison de la lenteur des processus et du manque d'automatisation, les data scientists ne pouvaient réaliser qu'une seule expérimentation sur un cycle de deux semaines.

Les entreprises interrogées utilisent Delta Lake pour maintenir la fiabilité et la performance de leurs données. En appliquant des normes de qualité des données et en centralisant leurs données, les entreprises interrogées ont réduit le temps et les efforts nécessaires pour lancer de nouveaux projets. L'accès à un plus grand volume de données, de meilleure qualité, a amélioré la qualité des modèles ML.

Après avoir migré vers la plateforme Databricks et géré leurs données sur Delta Lake, les entreprises interrogées ont fait état de gains de productivité au sein de leurs équipes de data science. Les data analysts et les data scientists ont pu passer moins de temps à rechercher des données, à les nettoyer ou à analyser les résultats, et ils ont pu consacrer plus de temps aux activités génératrices de résultats pour l'entreprise. L'amélioration de la collaboration et de la gestion du cycle de vie ML a aidé les data scientists à créer et améliorer leurs modèles ML plus rapidement qu'auparavant. Les fonctionnalités et les capacités supplémentaires de la plateforme ont permis aux équipes de travailler beaucoup plus rapidement. Au lieu d'effectuer une seule expérimentation sur un cycle de deux semaines, les équipes de data science étaient en mesure d'effectuer plusieurs analyses par jour, ce qui leur permettait de mieux cerner les résultats. Cela a donné lieu à de meilleurs modèles ML.

Les langages supplémentaires pris en charge par la plateforme ont favorisé la collaboration entre les data analysts, les data scientists et les data engineers. Le vice-président de l'équipe de data science de l'entreprise de médias a indiqué : « Databricks nous a permis de mieux nous intégrer à l'équipe d'ingénieurs que notre ancienne solution. Nous sommes mieux intégrés à leur infrastructure technologique, nous travaillons sur les langages qu'ils sont prêts à prendre en charge, et nous avons des référentiels en commun. »

Les entreprises interrogées ont noté les gains de temps suivants permis par la plateforme Databricks :



25 % d'augmentation
de la productivité des
data scientists.

CATEGORIE	AVANT DATABRICKS	APRES DATABRICKS
Création et maintenance de pipelines d'ETL	Plusieurs jours	Quelques minutes
Configuration de clusters de serveurs pour l'expérimentation de nouveaux produits de données	De plusieurs semaines à plusieurs mois	Quelques heures
Vérification et nettoyage des données	Plusieurs jours	Quelques heures
Formation des modèles/réglage des hyperparamètres et validation des modèles	Limité ou inexistant	De quelques heures à une journée
Mise en production des modèles, suivi, gestion et recyclage des modèles	Plusieurs cycles de deux semaines sur plusieurs mois	De quelques jours à quelques semaines, selon le niveau de libre-service activé
Passer le produit de données du développement à la production	De plusieurs semaines à plusieurs mois	De quelques jours à quelques semaines, selon le niveau de libre-service activé

Tous ces facteurs ont permis aux entreprises interrogées de mettre au point de meilleurs modèles ML, plus rapidement et avec moins de ressources qu'auparavant. Pour l'entreprise composite, Forrester émet les hypothèses suivantes :

- › Les data scientists et les data engineers passent 75 % de leur temps à rechercher des données et à les nettoyer, à créer des environnements de test et à faire passer les modèles en production.
- › Forrester part du principe que les data engineers et les data scientists réduisent le temps consacré aux activités susmentionnées de 25 % et 20 %, respectivement. Les data scientists y parviennent principalement en réduisant le temps passé à rechercher et nettoyer les données. Les data engineers réduisent en premier lieu le temps passé à nettoyer les données et à créer des environnements de test et de production.

L'augmentation de la productivité variera en fonction des facteurs suivants :

- › La façon dont les data engineers et les data scientists passent actuellement leur temps.
- › La qualité des données.
- › La rémunération toutes charges comprises d'un data engineer et d'un data scientist.

Pour tenir compte de ces risques, Forrester a réduit cet avantage de 5 %, ce qui a donné une valeur actuelle (VA) pondérée en fonction des risques de 7,7 millions de dollars sur trois ans.



20 % d'augmentation de la productivité des ingénieurs

Amélioration de l'efficacité opérationnelle : tableau de calcul

RÉF.	INDICATEUR	CALCUL	ANNÉE 1	ANNÉE 2	ANNÉE 3
B1	Data scientists affectés par Databricks	Entreprise composite	25	50	75
B2	Pourcentage du temps des data scientists affecté par l'adoption de Databricks	Sur la base des entretiens avec les clients	75 %	75 %	75 %
B3	Gain de temps pour les data scientists grâce à Databricks	Sur la base des entretiens avec les clients	25 %	25 %	25 %
B4	Salaire des data scientists	Moyenne du secteur	150 000 dollars	150 000 dollars	150 000 dollars
B5	Data engineers affectés par Databricks	Entreprise composite	50	100	150
B6	Pourcentage du temps des data engineers affecté par Databricks	Sur la base des entretiens avec les clients	75 %	75 %	75 %
B7	Pourcentage du temps des data engineers affecté par l'adoption de Databricks	Sur la base des entretiens avec les clients	20 %	20 %	20 %
B8	Salaire des data engineers	Moyenne du secteur	130 000 dollars	130 000 dollars	130 000 dollars
Bt	Amélioration de l'efficacité opérationnelle	$(B1*B2*B3*B4)+(B5*B6*B7*B8)$	1 678 125 dollars	3 356 250 dollars	5 034 375 dollars
	Pondération des risques	↓5 %			
Btr	Amélioration de l'efficacité opérationnelle (pondérée en fonction des risques)		1 594 219 dollars	3 188 438 dollars	4 782 656 dollars

Réduction des coûts de l'ancienne plateforme de data analysis

Avant d'adopter Databricks, la plupart des entreprises interrogées géraient de vastes environnements de data analysis sur site. Les entreprises interrogées disposaient soit de centaines ou de milliers de serveurs de base utilisant des solutions open source, soit de serveurs coûteux utilisant des logiciels propriétaires, soit les deux dans certains cas. Pour continuer à répondre à leurs besoins organisationnels, les clients interrogés ont dû continuer à acheter des serveurs, entraînant une augmentation des frais de gestion. Un responsable a décrit la

difficulté de gérer son ancien environnement : « Vous devez sans arrêt acheter de nouvelles machines ou ajouter de nouvelles instances pour accompagner l'augmentation de vos besoins. Vous étendez votre infrastructure et achetez de nouvelles licences. C'est difficile à gérer. Vous devez vous occuper du financement [sur l'ensemble des différents départements], des demandes d'approvisionnement et de la mise à niveau des équipements. » Ces difficultés ne font qu'augmenter à mesure que l'environnement continue de s'étendre.

L'adoption du Cloud n'a guère contribué à infléchir la courbe des coûts. Un responsable nous a confié que les coûts liés au Cloud augmentaient de plus de 16 % chaque année pour répondre à la demande croissante.

Grâce à Databricks, les clients interrogés ont pu commencer à mettre fin à leurs infrastructures sur site et à réduire ou annuler les licences et services de tiers. Databricks a également permis de réduire les coûts d'administration. Les ingénieurs n'ont plus à se soucier de la maintenance ou de la mise à niveau de la plateforme. Plusieurs personnes interrogées ont encore réduit les exigences administratives en fournissant des portails en libre-service et en passant aux DevOps. « Nous adoptons les DevOps et nous passons à un système de type « vous le possédez, vous le dimensionnez et vous gérez vos propres coûts ». Nous utilisons des interfaces d'API et des outils d'orchestration pour mettre en place de nouvelles instances avec les bonnes permissions, appliquer le tagging et journaliser toutes les activités pertinentes. »

Réduction des coûts étendue au Cloud. Les personnes interrogées ont indiqué avoir constaté que les coûts du Cloud étaient restés stables ou avaient augmenté moins rapidement qu'auparavant, malgré l'augmentation substantielle de la taille de leurs data lakes. L'évolutivité automatique de Databricks, associée à sa stabilité et à ses performances nettement supérieures, a permis de réduire les coûts de l'infrastructure dans le Cloud.

Les responsables en charge des données interrogés ont noté des améliorations significatives en matière de stabilité et de traitement après le passage à Databricks :

- › Le responsable de l'architecture du fabricant d'équipements lourds a déclaré avoir observé une augmentation de 40 à 50 % des processus spécifiques. Il a également constaté une réduction de 97 % du temps nécessaire à l'achèvement des processus Spark, par rapport à la solution précédente dans le Cloud, qui utilisait Spark en open source. Le responsable de l'architecture a également décrit les problèmes liés à l'ancien environnement : « Certains processus ne fonctionnaient tout simplement pas. Ils fonctionnaient pendant deux ou trois heures, puis plantaient le serveur. [Avec Databricks,] ils s'exécutent et se terminent en quelques secondes. » Enfin, le nombre d'échecs est passé de « un sur 100 à 250 travaux [avant Databricks] à un sur mille [avec Databricks] ».
- › Le vice-président de l'équipe de data science a noté que l'exécution complète des processus est passée d'une semaine à une heure et demie avec Databricks. D'après lui : « Un processus prenait avant une semaine de traitement et une semaine d'activités manuelles. Grâce aux données disponibles par programme dans un data lake centralisé et à la fonctionnalité de gestion des clusters de Databricks, nous sommes en mesure d'automatiser l'ensemble du processus du début jusqu'à la fin. »

« Nos coûts liés au Cloud sont restés stables d'une année sur l'autre, malgré la croissance exponentielle et l'utilisation du Cloud. »

Vice-président, data science, médias



75 % des anciens environnements sont désactivés sur trois ans avec Databricks

Pour l'entreprise composite, Forrester émet les hypothèses suivantes :

- › Le coût de maintenance de ses infrastructures sur site est de 7 millions de dollars par an. Cela comprend les dépenses en matériel pour les serveurs mis hors service, la connectivité réseau, le refroidissement, le stockage, les coûts de main-d'œuvre administrative interne et les coûts des logiciels non liés à la solution de data analysis.
- › Les coûts d'infrastructure augmentent de 20 % d'une année sur l'autre pour répondre aux besoins croissants.

Les coûts d'infrastructure sur site varient en fonction des facteurs suivants :

- › La taille de l'environnement de data analysis sur site d'une entreprise.
- › Les coûts des licences de data analysis.

Pour tenir compte de ces risques, Forrester a réduit cet avantage de 5 %, ce qui a donné une valeur actuelle (VA) pondérée en fonction des risques de 11 298 202 dollars sur trois ans.

Réduction des coûts de l'ancienne plateforme de data analysis : tableau de calcul

RÉF.	INDICATEUR	CALCUL	ANNÉE 1	ANNÉE 2	ANNÉE 3
C1	Coût annuel de l'infrastructure sur site pour la solution existante de data analysis	Entreprise composite	7 000 000 dollars	8 400 000 dollars	10 080 000 dollars
C2	Coûts de licence et de support de la plateforme existante de data analysis	Entreprise composite	750 000 dollars	900 000 dollars	1 080 000 dollars
C3	Pourcentage de l'environnement mis hors service	Entreprise composite	25 %	50 %	75 %
C4	Coûts évités pour le matériel et les licences	$(C1+C2)*C3$	1 937 500 dollars	4 650 000 dollars	8 370 000 dollars
Ct	Réduction des coûts de l'ancienne plateforme de data analysis	C4	1 937 500 dollars	4 650 000 dollars	8 370 000 dollars
	Pondération des risques	↓5 %			
Ctr	Réduction des coûts de l'ancienne plateforme de data analysis (pondérée en fonction des risques)		1 840 625 dollars	4 417 500 dollars	7 951 500 dollars

Avantages non quantifiés

Les entreprises interrogées ont constaté les avantages suivants, qui n'ont pu être financièrement quantifiés pour cette étude. Les avantages non quantifiés de la plateforme Unified Data Analytics Platform de Databricks, évalués par Forrester, comprennent :

› **Réduction des coûts d'exploitation.** Les nouveaux modèles ML créés sur la plateforme Databricks, associés aux nouveaux insights issus de la centralisation de leurs données, ont permis aux entreprises interrogées de réduire leurs dépenses en matières premières, d'identifier plus rapidement les problèmes de fabrication, et de réduire les coûts de réparation et les taux de retour.

- L'entreprise pharmaceutique a créé un moteur d'IA pour identifier sur qui les commerciaux devraient concentrer leurs efforts, et quand le faire. L'entreprise a également découvert qu'elle payait trop pour les matières premières, ce qui lui a permis de renégocier ses contrats et d'économiser des dizaines de millions de dollars.
- Le retailer a augmenté ses ventes grâce à l'amélioration de son moteur de recherche et la création de moteurs de recommandation pour différents types de clients. Il a également pu réduire ses coûts d'exploitation en réduisant le volume de retours et en identifiant les commandes frauduleuses.
- Le fabricant d'équipements lourds a amélioré sa capacité à identifier les problèmes de fabrication, réduisant ainsi les temps d'arrêt pour ses clients et les coûts de réparation.

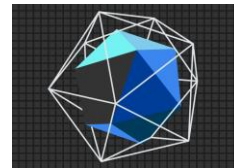
› **Meilleure sécurité.** Avant d'utiliser Databricks, les entreprises interrogées ne disposaient pas de normes de sécurité dans leur environnement ni des ressources internes nécessaires pour appliquer les normes de sécurité. Databricks intègre une protection de sécurité native.

Flexibilité

La valeur de la flexibilité est clairement spécifique à chaque client et sa mesure varie donc d'une entreprise à l'autre. Il existe de nombreux scénarios de mise en œuvre de la plateforme Databricks, dans lesquels le client peut identifier d'autres possibilités d'utilisation et opportunités métier, notamment :

- › **Utilisation de nouvelles bibliothèques ML.** Databricks fournit des bibliothèques ML et des frameworks prêts à l'emploi, permettant aux data scientists d'expérimenter facilement les dernières techniques d'apprentissage machine.
- › **Adoption de nouveaux langages de programmation.** Databricks prend en charge un large éventail de langages de programmation modernes (notamment R, Python, Scala et SQL), ce qui permet aux équipes de manipuler les données à l'aide de nouveaux langages de programmation dans le même Notebook.

La flexibilité peut également être quantifiée lors de son évaluation dans le cadre d'un projet spécifique (voir description détaillée en annexe A).



Meilleures sécurité et gouvernance des données

La flexibilité, telle que définie par le TEI, représente un investissement dans une capacité ou une fonctionnalité supplémentaire qui peut se révéler être un avantage commercial pour d'autres investissements à venir. Elle offre à une entreprise le « droit » ou la possibilité de lancer des initiatives à l'avenir, mais pas l'obligation de le faire.

Analyse des coûts

DONNEES DE COÛTS QUANTIFIEES APPLIQUEES A L'ENTREPRISE COMPOSITE

Total des coûts							
RÉF.	COÛT	INITIAL	ANNÉE 1	ANNÉE 2	ANNÉE 3	TOTAL	VALEUR ACTUELLE
Dtr	Coûts de la plateforme Databricks, de la formation et du stockage	0 dollar	971 250 dollars	1 732 500 dollars	2 913 750 dollars	5 617 500 dollars	4 503 916 dollars
Etr	Coûts d'administration de Databricks	136 500 dollars	273 000 dollars	273 000 dollars	273 000 dollars	955 500 dollars	815 411 dollars
Ftr	Coûts de formation sur Databricks	4 234 dollars	101 606 dollars	105 840 dollars	105 840 dollars	317 520 dollars	263 593 dollars
	Total des coûts (pondéré en fonction des risques)	140 734 dollars	1 345 856 dollars	2 111 340 dollars	3 292 590 dollars	6 890 520 dollars	5 582 920 dollars

Coûts de la plateforme Databricks, de la formation et du stockage

Pour utiliser la plateforme Unified Data Analytics Platform de Databricks, les entreprises ont engagé des coûts de consommation appelés unités Databricks (DBU). La DBU est une unité de traitement par heure, facturée à la seconde. Les entreprises interrogées ont payé pour l'intégrer et former leurs data analysts, data engineers et data scientists, et pour démocratiser l'utilisation de la plateforme auprès des spécialistes du marketing, des intervenants et d'autres collaborateurs au sein des entreprises.

Les entreprises ont également encouru des coûts de stockage de la part de leurs prestataires Cloud tiers.

Pour l'entreprise composite, Forrester a émis les hypothèses suivantes :

- › Les coûts d'utilisation augmentent chaque année à mesure que de nouvelles données et business units passent à la plateforme Databricks.

Forrester reconnaît que ces coûts varieront en fonction des facteurs suivants :

- › L'utilisation de la plateforme Databricks.
- › La quantité de données transférées dans le Cloud.
- › Le nombre de personnes formées sur la plateforme.

Pour tenir compte de ces risques, Forrester a augmenté cet avantage de 5 %, ce qui a donné une valeur actuelle (VA) pondérée en fonction des risques de 4 503 916 dollars sur trois ans.

Le tableau ci-dessus présente le total de tous les coûts dans les domaines répertoriés ci-dessous, ainsi que les valeurs actuelles (VA) sur la base d'un taux d'actualisation de 10 %. Sur une durée de trois ans, l'entreprise composite prévoit que les coûts totaux, pondérés en fonction des risques, atteindront une valeur actuelle de plus de 5,5 millions de dollars.

Le risque de mise en œuvre est le risque qu'un investissement donné s'écarte des critères initiaux ou attendus et entraîne des coûts plus élevés que prévu. Plus l'incertitude est grande, plus la fourchette des résultats possibles est large pour estimer les coûts.

Coûts de la plateforme Databricks, de la formation et du stockage : tableau de calcul

RÉF.	INDICATEUR	CALCUL	INITIAL	ANNÉE 1	ANNÉE 2	ANNÉE 3
D1	Coûts de la plateforme Databricks et de la formation			800 000 dollars	1 400 000 dollars	2 400 000 dollars
D2	Coûts du stockage dans le Cloud			125 000 dollars	250 000 dollars	375 000 dollars
Dt	Coûts de la plateforme Databricks, de la formation et du stockage	D1+D2	0 dollar	925 000 dollars	1 650 000 dollars	2 775 000 dollars
	Pondération des risques		↑5 %			
Dtr	Coûts de la plateforme Databricks, de la formation et du stockage (pondérés en fonction des risques)		0 dollar	971 250 dollars	1 732 500 dollars	2 913 750 dollars

Coûts d'administration de Databricks

Avant d'adopter pleinement la plateforme Unified Data Analytics Platform de Databricks, les entreprises interrogées ont effectué une validation de principe pour s'assurer que la plateforme répondait à leurs besoins. Les entreprises interrogées ont ensuite affecté des data engineers à l'administration de la plateforme Databricks ; notamment :

- › La gestion de Delta Lake, y compris la mise en œuvre initiale des politiques de gouvernance et la migration continue des données vers la plateforme.
- › L'aide à la formation des nouveaux collaborateurs.

Pour la modélisation de ces coûts administratifs, Forrester suppose que deux data engineers se consacrent à l'administration de la plateforme Databricks.

Les frais d'administration peuvent varier en fonction des facteurs suivants :

- › La quantité de main-d'œuvre et ses compétences existantes.
- › La taille et la portée d'un déploiement.
- › La quantité de données transférées dans le Cloud.
- › Les besoins de l'entreprise.
- › Les exigences en matière de gouvernance des données.

Pour tenir compte de ces risques, Forrester a augmenté cet avantage de 5 %, et a obtenu une valeur actuelle (VA) pondérée en fonction des risques de 815 411 dollars sur trois ans.



Six mois :
Temps total de mise en œuvre et de déploiement



Deux ETP
consacrent 100 % de leur temps à l'administration de la plateforme Databricks.

Tableau de calcul des coûts d'administration de Databricks

RÉF.	INDICATEUR	CALCUL	INITIAL	ANNÉE 1	ANNÉE 2	ANNÉE 3
E1	Data engineers dédiés à la validation de principe de Databricks et sa mise en œuvre	Entreprise composite	2	0,00	0,00	0,00
E2	Mois nécessaires à la mise en œuvre de Databricks		6	0,00	0,00	0,00
E3	Data engineers nécessaires à la maintenance de Databricks			2,00	2,00	2,00
E4	Salaire des data engineers		130 000 dollars	130 000 dollars	130 000 dollars	130 000 dollars
Et	Coûts d'administration de Databricks	$(E1 * E2 / 12 * E4) + (E3 * E4)$	130 000 dollars	260 000 dollars	260 000 dollars	260 000 dollars
	Pondération des risques	↑5 %				
Etr	Coûts d'administration de Databricks (pondérés en fonction des risques)		136 500 dollars	273 000 dollars	273 000 dollars	273 000 dollars

Coûts de formation sur Databricks

En plus des coûts de formation calculés ci-dessus, Forrester a calculé les coûts de main-d'œuvre internes associés à la formation des collaborateurs de l'entreprise composite sur la plateforme Databricks. Dans le calcul de ce coût, Forrester suppose que :

- › Dans le cadre de la validation de principe initiale, un groupe de sept ETP composé de data engineers, de data scientists et d'autres utilisateurs, sont formés sur la plateforme Databricks.
- › 25 data scientists, 50 data engineers et 100 utilisateurs supplémentaires sont formés sur la plateforme Databricks au cours des Années 2 et 3.
- › Les utilisateurs consacrent en moyenne 12 heures à des séminaires animés par un instructeur, à des déjeuners et des séances d'apprentissage, ainsi qu'à des cours auto-guidés.
- › Un salaire mixte moyen de 100 000 dollars par an.

Les coûts de formation varient en fonction des facteurs suivants :

- › Le nombre d'ETP formés par an.
- › Les compétences existantes, qui augmenteront ou diminueront les besoins de formation.
- › La nécessité de former d'autres utilisateurs que les data engineers et les data scientists.
- › Les salaires des ETP.

Pour tenir compte de ces risques, Forrester a augmenté cet avantage de 5 %, et a obtenu une valeur actuelle (VA) pondérée en fonction des risques de 263 593 dollars sur trois ans.

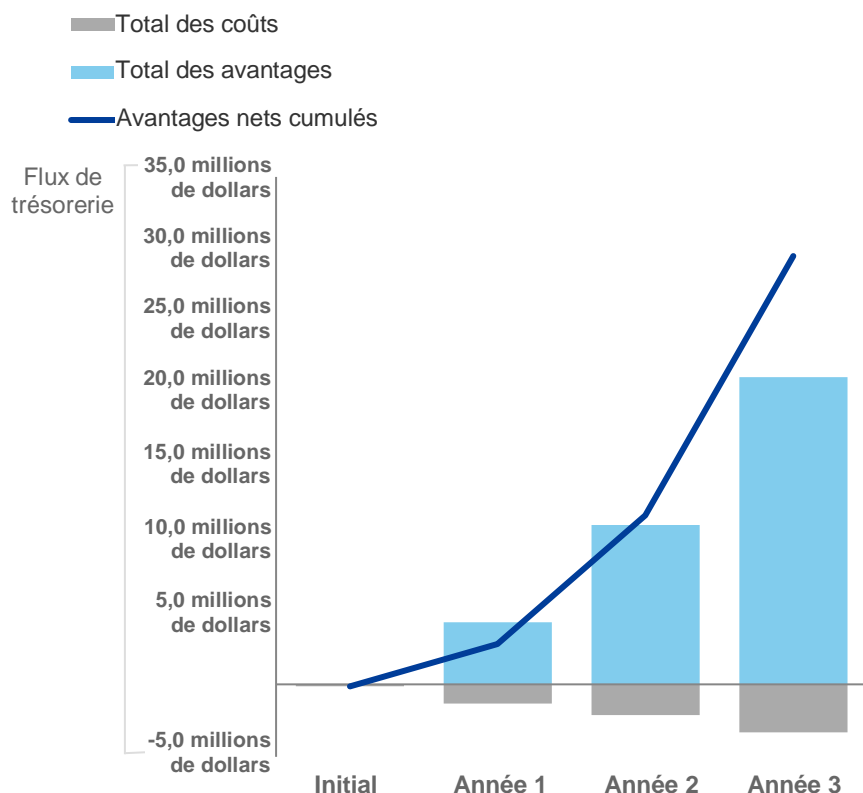
Coûts de formation sur Databricks : tableau de calcul

RÉF.	INDICATEUR	CALCUL	INITIAL	ANNÉE 1	ANNÉE 2	ANNÉE 3
F1	ETP formés sur Databricks	Entreprise composite	7	168	175	175
F2	Heures consacrées à la formation	Hypothèse	12	12	12	12
F3	Salaire pondéré d'un ETP	Hypothèse	48 dollars	48 dollars	48 dollars	48 dollars
Ft	Coûts de formation sur Databricks	$F1 \cdot F2 \cdot F3$	4 032 dollars	96 768 dollars	100 800 dollars	100 800 dollars
	Pondération des risques	↑5 %				
Ftr	Coûts de formation sur Databricks (pondérés en fonction des risques)		4 234 dollars	101 606 dollars	105 840 dollars	105 840 dollars

Bilan financier

INDICATEURS CONSOLIDÉS SUR TROIS ANS ET PONDÉRÉS EN FONCTION DES RISQUES

Graphique des flux de trésorerie (pondérés en fonction des risques)



Les résultats financiers calculés dans les sections Avantages et Coûts peuvent être utilisés pour déterminer le retour sur investissement (ROI), la valeur actuelle nette (VAN) et le délai de rentabilité pour l'investissement de l'entreprise composite. Forrester suppose un taux d'actualisation annuel de 10 % pour cette analyse.



Ces valeurs de ROI, VAN et de délai de rentabilité pondérées en fonction des risques sont déterminées en appliquant des facteurs de pondération des risques aux résultats bruts de chaque section Bénéfices et Coûts.

Analyse des flux de trésorerie (estimations pondérées au risque)

	INITIAL	ANNÉE 1	ANNÉE 2	ANNÉE 3	TOTAL	VALEUR ACTUELLE
Total des coûts	(140 734 dollars)	(1 345 856 dollars)	(2 111 340 dollars)	(3 292 590 dollars)	(6 890 520 dollars)	(5 582 920 dollars)
Total des avantages	0	4 278 594 dollars	10 980 938 dollars	21 171 656 dollars	36 431 188 dollars	28 871 365 dollars
Bénéfices nets	(140 734 dollars)	2 932 737 dollars	8 869 598 dollars	17 879 066 dollars	29 540 668 dollars	23 288 445 dollars
ROI						417 %
Délai de récupération (mois)						<6

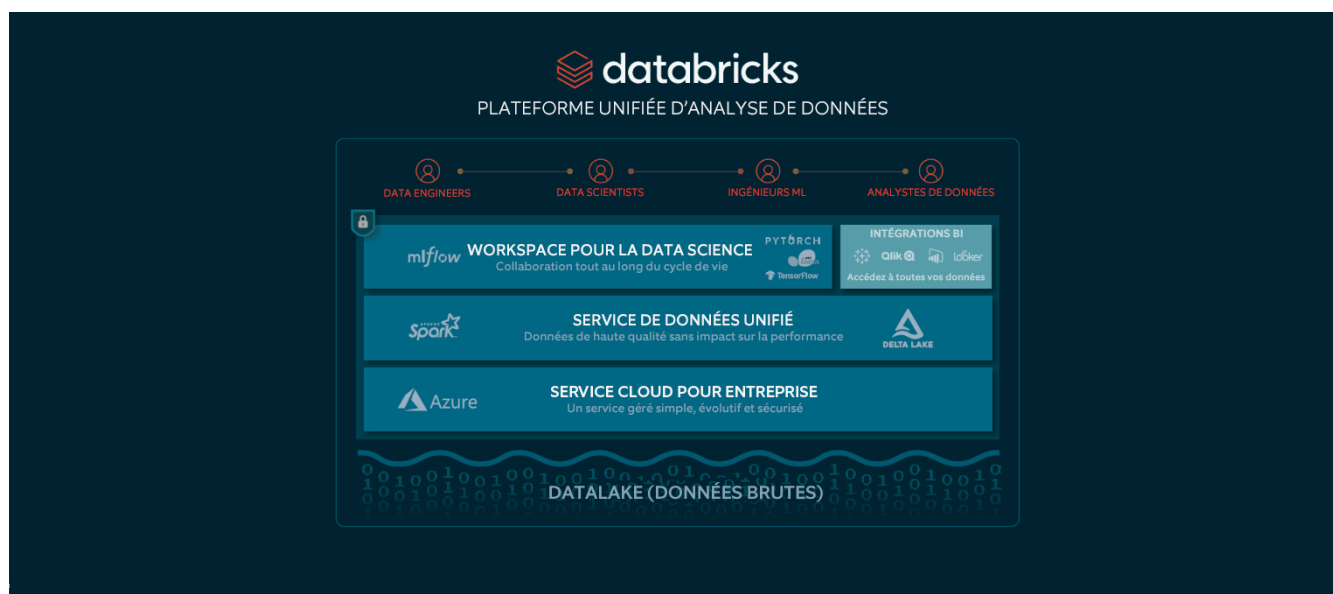
Présentation de la plateforme Unified Data Analytics Platform de Databricks

Les informations ci-dessous sont fournies par Databricks. Forrester n'a pas validé ces informations et ne cautionne pas Databricks ou ses offres.

Accélération de l'innovation axée sur les données avec Databricks

Databricks simplifie les données et l'IA afin que les équipes chargées des données puissent résoudre les problèmes les plus difficiles qui soient. Databricks est une solution open source et Cloud first. Le produit combine des données et des analyses dans une plateforme unifiée et ouverte qui permet aux équipes chargées des données de collaborer et d'innover plus rapidement que jamais. Des milliers d'entreprises à travers le monde, dont Comcast, Shell, Starbucks et Regeneron, utilisent Databricks pour préparer leurs données en vue de les analyser, de prendre des décisions axées sur les données dans toute l'entreprise et déployer rapidement le machine learning. Databricks est une entreprise financée par le capital-risque et fondée par les créateurs des projets open source Apache Spark™, Delta Lake et MLflow.

La plateforme Unified Data Analytics Platform de Databricks



Workspace pour la data science

Workspace pour la data science est un environnement collaboratif permettant aux professionnels d'exécuter tous les processus analytiques à partir d'un seul et même endroit et de gérer les modèles ML tout au long du cycle de vie.

- Notebooks collaboratifs : les Notebooks Databricks prennent en charge Python, R, SQL et Scala en mode natif afin que les professionnels puissent travailler avec les langages et les bibliothèques de leur choix pour découvrir, visualiser et partager des informations avec les intervenants.
- Runtime de Machine Learning : accès en un clic à des clusters ML préconfigurés reposant sur une version évolutive et fiable des frameworks ML les plus populaires, avec AutoML intégré et optimisations pour des performances sans pareil.
- Managed MLflow : articulé sur MLflow, une plateforme open source de Databricks, Managed MLflow permet de gérer des modèles ML de manière sécurisée, fiable et évolutive, de l'expérimentation jusqu'à la production.

Service de données unifié

Le Service de données unifié de Databricks fournit une plateforme fiable et évolutive pour vos pipelines de données, vos data lakes et vos plateformes de données. Gérez intégralement le parcours de vos données afin de pouvoir absorber, traiter, stocker et exposer les données sur l'ensemble de votre entreprise.

- **Delta Lake pour Databricks** : Delta Lake améliore la fiabilité, les performances et la gestion du cycle de vie des données. Fini les travaux incomplets qu'il faut nettoyer et recommencer, les données suspectes ajoutées à votre data lake ou les difficultés à supprimer des données pour cause de changements de conformité.
- **Databricks Runtime** : Databricks Runtime est un moteur de traitement de données distribué reposant sur une version hautement optimisée d'Apache Spark, pouvant atteindre des performances 50 fois supérieures. Développez des pipelines, programmez des travaux et formez facilement des modèles en libre-service de manière économique.
- **BI Reporting sur Delta Lake** : BI Reporting sur Delta Lake intègre des analyses à votre data lake. Connectez-vous directement à vos données les plus complètes et les plus récentes dans votre data lake avec Delta Lake et SparkSQL, et utilisez vos outils préférés pour la visualisation et le reporting de BI en vue d'obtenir des insights plus rapidement.

Enterprise Cloud Service

Enterprise Cloud Service intègre une sécurité native, une administration simple à l'échelle de l'entreprise et une automatisation pour la plateforme Unified Data Analytics Platform dans différents Clouds.

- **Sécurité au niveau de l'entreprise** : avec la fédération des identités, le chiffrement et les contrôles d'accès natifs, vous disposez des outils nécessaires pour sécuriser vos données. Vous pouvez créer un environnement d'analyse sûr pour vos utilisateurs via le réseau intégré, les données et l'isolement des processus avec différentes options de conformité.
- **Administration simple** : les pistes d'audit, les journaux et les rapports de facturation et d'utilisation fournissent une visibilité opérationnelle totale. Créez des workspaces individuels pour chaque équipe, définissez des politiques régissant les limites d'utilisation et analysez les activités pour garantir la conformité.
- **Prêt pour la production** : l'approche API first garantit des processus CI/CD transparents et l'automatisation sur une infrastructure avec mise à l'échelle automatique. Respectez les SLA avec une application facile et la supervision de l'infrastructure.

Annexe A : Total Economic Impact

Total Economic Impact est une méthodologie élaborée par Forrester Research qui améliore les processus décisionnels d'une entreprise en matière de technologies et aide les fournisseurs à communiquer la proposition de valeur de leurs produits et services à leurs clients. La méthodologie TEI aide les entreprises à démontrer, justifier et concrétiser la valeur réelle des initiatives en matière de technologies de l'information, à la fois auprès de la direction et auprès des autres parties prenantes clés internes.

L'approche Total Economic Impact



Les avantages représentent la valeur apportée à l'entreprise par le produit. La méthodologie TEI met en balance la mesure des avantages et la mesure des coûts, ce qui permet l'examen complet de l'impact de la technologie sur l'ensemble de l'entreprise.



Les coûts tiennent compte de toutes les dépenses nécessaires pour obtenir la valeur attendue ou les avantages du produit. La catégorie de coûts du TEI correspond aux coûts différentiels dans l'environnement existant pour déterminer les coûts récurrents associés à la solution.



La flexibilité représente la valeur stratégique pouvant être obtenue pour un investissement futur supplémentaire complétant l'investissement initial déjà réalisé. La possibilité de tirer parti de cet avantage présente une VA qui peut être estimée.



Les risques mesurent l'incertitude des estimations des avantages et des coûts en considérant : 1) la probabilité que les estimations correspondent aux projections d'origine et 2) la probabilité que les estimations soient suivies dans le temps. Les facteurs de risque du TEI reposent sur une « distribution triangulaire ».

La colonne Investissement initial présente les coûts engagés à « l'instant 0 » ou au début de l'année 1, et non actualisés. Tous les autres flux de trésorerie sont actualisés au taux d'actualisation en fin d'année. Les calculs de la VA sont effectués pour chaque estimation de coûts et d'avantages totaux. Les calculs de la VAN figurant dans les tableaux de synthèse correspondent à la somme de l'investissement initial et des flux de trésorerie actualisés chaque année. Il est possible que les calculs des sommes et de la valeur actuelle des tableaux Avantages totaux, Coûts totaux et Flux de trésorerie ne s'additionnent pas exactement, en raison de l'arrondissement des nombres.



Valeur actualisée (VA)

Valeur actuelle ou en cours des estimations de bénéfices et de coûts (actualisées) à un taux d'intérêt donné (taux d'actualisation). La VA des coûts et des bénéfices entre dans la valeur actuelle nette totale des flux de trésorerie.



Valeur actuelle nette (VAN)

Valeur actuelle ou en cours des futurs flux de trésorerie nets (actualisés) à un taux d'intérêt donné (taux d'actualisation). La VAN positive d'un projet indique normalement que l'investissement est recommandé, mais d'autres projets peuvent présenter des VAN supérieures.



Retour sur investissement (ROI)

Rentabilité attendue d'un projet, exprimée en pourcentage. Le ROI est calculé en divisant les avantages nets (avantages moins coûts) par les coûts.



Taux d'actualisation

Taux d'intérêt utilisé dans l'analyse des flux de trésorerie pour prendre en compte la valeur temps de l'argent. Les entreprises utilisent généralement des taux d'actualisation compris entre 8 et 16 %.



Délai de rentabilité

Seuil de rentabilité d'un investissement. Stade auquel les avantages nets (avantages moins coûts) équivalent à l'investissement ou au coût initial.

Annexe B : Notes de fin

ⁱ Source : « Build An Insights-Driven Business », Forrester Research, Inc., 10 décembre 2019

ⁱⁱ Source : « The Future Of Machine Learning Is Unstoppable », Forrester Research, Inc., 25 avril 2019