



# L'effet transformateur des marketplaces de données

LIVRE BLANC





# SOMMAIRE

<b>Introduction</b>	<b>3</b>
Des entreprises submergées par la data (et non par leur valeur ajoutée).....	3
<b>Les leçons apprises des data lakes</b>	<b>4</b>
Un référentiel unique peut-il tout régir ?.....	5
<b>Du data lake à la marketplace des données</b>	<b>5</b>
<b>Fonctionnalités de la marketplace des données</b>	<b>7</b>
Enregistrement de données externes.....	10
Extraction automatique des métadonnées et des informations de profilage.....	10
Analyser les données d'utilisation.....	11
<b>Faire vivre une marketplace des données</b>	<b>11</b>
Gestionnaires data.....	13
Utilisateurs des données.....	13
<b>Conclusion</b>	<b>14</b>

Pour valoriser toutes les données que vous collectez en permanence, vous avez besoin d'un nouveau concept qui tire les leçons apprises de l'usage des datawarehouses et des datalakes et qui les transcende en y ajoutant de nouvelles fonctionnalités cruciales. Nous appelons ce nouveau concept la « marketplace des données ».

## INTRODUCTION

Dans certaines entreprises, le data lake s'est littéralement asséché. Dans d'autres, il déborde, il est stagnant ou pollué. Rares sont les entreprises qui peuvent crier victoire en termes de gestion, d'analyse et « d'opérationnalisation » de leurs jeux de données. En effet, leur volume croît de façon exponentielle, que les données proviennent de sources existantes ou nouvelles.

Pour valoriser toutes les données que vous collectez en permanence, vous avez besoin d'un nouveau concept qui tire les leçons apprises de l'usage des datawarehouses et des datalakes et qui les transcende en y ajoutant de nouvelles fonctionnalités cruciales. Nous appelons ce nouveau concept la « marketplace des données ». Ce livre blanc explique :

- 1 Pourquoi les data warehouses et les data lakes ne parviennent pas à relever le défi actuel lié à l'infrastructure data.
- 2 En quoi consiste la marketplace des données et en quoi ce concept diffère des approches antérieures.
- 3 Comment la marketplace des données garantit la disponibilité des données pour ceux qui en ont besoin et comment ces données deviennent un actif avec une valeur ajoutée croissante.
- 4 À quoi ressemble un environnement avec une marketplace des données opérationnelle.

### Des entreprises submergées par la data (et non par leur valeur ajoutée)

Les entreprises n'ont jamais eu autant de données à gérer. L'émergence du Big Data a entraîné une hausse du volume, de la vitesse et de la diversité des données. Ceci fait naître à la fois des opportunités et des défis. Le nombre de nouvelles applications sur site, dans le Cloud et proposées par les fournisseurs de SaaS ne cesse d'augmenter, tout comme les données issues de ces applications. Ajoutez-y la data provenant des logs de serveurs, des données machine, des dispositifs IoT et d'autres domaines digitaux qui génèrent de nouveaux types de signaux, si vous êtes en mesure de les capturer. Les nouvelles technologies de la BI, de l'IA et du machine learning génèrent des informations métier uniquement si elles sont alimentées avec les données adéquates. En même temps, elles génèrent elles-mêmes d'énormes volumes de data.

Pour tirer parti de toutes ces données, les entreprises doivent disposer des systèmes de données, des architectures et des outils nécessaires pour en assurer le suivi, les transformer, faire connaître leur existence, puis les diffuser auprès des utilisateurs, quand et où ils en ont besoin. Malheureusement, c'est trop rarement le cas. La data n'arrive pas à temps pour permettre une prise de décision rapide et efficace. En tant que tel, un volume accru de données n'apporte pas d'avantages supplémentaires. En réalité, les données non identifiables ou inutilisables représentent plutôt une hausse des coûts, et non une source de valeur ajoutée.

“ Les organisations se rendent compte que la simple insertion d'un nombre massif et diversifié de données dans Hadoop ou dans un data lake ne va pas créer, comme par magie, des insights pertinents. Elles doivent y ajouter l'intégration, la transformation, l'enrichissement, la sécurité et la gouvernance. La distribution de données connectées entre différentes sources sur site et dans le Cloud n'est pas anodine, en particulier lorsqu'elle implique des volumes de données importants, des modèles de data complexes et une vitesse d'ingestion élevée. Une intégration insuffisante des data d'entreprise débouche souvent sur de mauvaises décisions métier. Elle réduit la satisfaction des clients, affaiblit la compétitivité et ralentit l'innovation. Au final, les revenus sont restreints et la valeur ajoutée reste faible, voire nulle, pour l'entreprise. »

NOEL YUHANNA,  
FORRESTER RESEARCH

Les outils disponibles jusqu'à récemment n'étaient pas conçus pour cette ère du Big Data. Ou ils ne tenaient tout simplement pas leurs promesses, à savoir pouvoir gérer ou « opérationnaliser » ce Big Data.

Le data warehouse et le data lake offrent tout deux des avantages spécifiques, qu'il faut intégrer à l'avenir du stockage et de la gestion de la data. Toutefois, utilisés séparément, ils ne constituent pas, et ne peuvent pas être, la réponse aux besoins des entreprises du 21<sup>e</sup> siècle en termes de données.

Les data warehouses ont créé un modèle canonique et unifié, devenu la base pour le reporting, la BI et la Data Discovery. Certes, leurs fonctionnalités de gestion de la data adaptées aux entreprises ainsi que l'attention particulière qu'ils portent à la gouvernance restent pertinentes. Cependant, ils ne sont généralement pas capables de délivrer les nouvelles données suffisamment rapidement, ni de générer des vues détaillées sur le Big Data.

Les data lakes sont nés dans l'idée de tirer, pour les entreprises, une valeur ajoutée du Big Data, nom donné au volume considérable de données issues de nouvelles sources et qui ne pouvaient être ni gérées ni analysées au moyen d'un data warehouse. Cette idée a ensuite été étendue, par certains, à la vision d'un super data warehouse qui pouvait faire office de référentiel unique pour régir le tout. Les data lakes ont été mis en œuvre pour nous permettre de stocker d'importants volumes de données, dans le but de les analyser de façon différente et ainsi récolter des signaux métier exploitables. Le problème est que les entreprises qui ont mis en place des data lakes parallèlement à leurs data warehouses manquent généralement d'un contexte métier clair pour la data au sein du data lake. Au mieux, la data est donc inutilisable et, au pire, son utilisation est risquée.

## LES LEÇONS APPRISSES DES DATA LAKES

L'expérience acquise dans l'utilisation des data lakes, malgré leurs lacunes, nous permet de tirer plusieurs leçons.

Le data lake représentait un pas en avant. En effet, il offrait les avantages suivants :

- **Volume.** Stockage de volumes importants de data à un prix abordable.
- **Diversité.** Stockage de nombreux types de données différents.
- **Schéma en lecture.** Nouveaux paradigmes pour l'exploitation de données permettant la collecte des données avant leur analyse approfondie. Les données pouvaient être stockées et provisionnées avant d'en déterminer la valeur ajoutée réelle pour l'entreprise. Il était ainsi possible d'indexer l'ensemble des données de l'entreprise et de les préparer pour des questions métier inconnues.
- **Prise en charge de nombreux modèles d'ingestion et de consommation.** Un référentiel unique pour l'ingestion et la consommation des flux de données, avec accès structuré aux documents et jusqu'aux colonnes.

Au final, de nombreux data lakes ne servent plus qu'à accumuler la data. Ils ne servent plus à organiser, à générer des insights ni à valoriser les données pour l'entreprise.

Pourtant, malgré ces avantages, la vision du data lake n'a pas encore été pleinement concrétisée. Les data lakes offrent des capacités de stockage immenses et la possibilité, pour les experts, de créer de nouveaux pipelines pour certains types d'analytics et de transformations de data. Toutefois, les entreprises consacrent des ressources au déploiement de produits open source destinés aux développeurs afin de tirer une valeur ajoutée, même limitée, de leurs data lakes.

### Un référentiel unique peut-il tout régir ?

Pour les entreprises, les data lakes représentaient la promesse d'enfin disposer d'un référentiel de Big Data pour stocker l'ensemble de leurs données et à partir duquel elles allaient pouvoir lancer des analyses. Ils allaient offrir « un référentiel unique pour tout stocker ».

Mais, pour toute une série de raisons, le data lake ne tient pas ses promesses. Et la désillusion des entreprises ne fait que grandir.<sup>1</sup>

Au final, de nombreux data lakes ne servent plus qu'à accumuler la data. Ils ne servent plus à organiser, à générer des insights ni à valoriser les données pour l'entreprise. Ils ne répondent pas à ces trois besoins essentiels :

- 1 Devenir le référentiel unique pour l'ensemble des données
- 2 Fournir des données exploitables à la demande
- 3 Garantir une gestion de la data adaptée à l'entreprise

Si les entreprises sont aussi frustrées vis-à-vis des data lakes, c'est précisément parce que ceux-ci ne répondent pas à ces besoins.

## DU DATA LAKE À LA MARKETPLACE DES DONNÉES

Afin de pouvoir gérer les flux chaotiques du Big Data, établir leur gouvernance et permettre leur exploitation générale, les entreprises doivent pouvoir créer et gérer facilement des flux pour l'ensemble de leurs données. Ceci doit inclure les aspects suivants :

**Incorporation constante de nouvelles données.** Au fur et à mesure de l'arrivée de nouvelles sources, celles-ci doivent être évaluées et incorporées dans la data de travail utilisée par l'entreprise. Les données ne peuvent pas simplement être insérées, un processus automatique et évolutif d'intégration doit être mis en place.

1. [Early Adopter Research](#) décrit toute une série de raisons pour lesquelles les data lakes ont échoué, dans sa mission de recherche « [Saving Your Data Lake](#) » (Sauver votre data lake).

**Création d'un cycle vertueux dans l'utilisation de la data.** Une fois que de nouvelles données sont collectées, les utilisateurs doivent être en mesure d'y accéder immédiatement et de les améliorer en y ajoutant des métadonnées métier. L'analyse et le profilage automatiques enrichissent les données avec encore plus de métadonnées. Les utilisateurs peuvent encourager d'autres utilisateurs à accéder aux données et à les exploiter. Avec le temps, les données sont plus nombreuses, de nouvelles métadonnées sont ajoutées et le nombre de personnes qui les utilisent augmente. Ce cycle finit par se perpétuer automatiquement, au bénéfice de l'entreprise.

**Lancement de produits data.** Au fur et à mesure que la compréhension des cas d'usage relatifs aux données intégrées évolue, la finalité d'une supply chain de la donnée est un produit data qui peut être exploité au sein des applications, des tableaux de bord et de tout autre moyen dans le but de favoriser la sensibilisation et l'efficacité des actions. Au fil du temps, l'ajout incrémentiel de jeux de données permet à l'entreprise de développer une compréhension plus solide et plus complète de l'ensemble de ses données.

**Augmentation du nombre de pratiquants de la data.** Étant donné le volume exponentiel des données, de nombreuses personnes doivent être en mesure de prendre part au processus d'ingestion, d'évaluation et de « productisation » de la data. L'augmentation du nombre d'utilisateurs représente également un avantage pour l'entreprise. En effet, chacun apporte son propre contexte, ce qui donne de la valeur ajoutée à plus de données. Ce partage permet d'élargir la compréhension des données de l'entreprise, et donc de donner un caractère plus pertinent à l'analytics.

Les marketplaces des données qui offrent ces fonctionnalités représentent donc la prochaine évolution en matière de gestion de la data. Elles constitueront, à l'avenir, une solution complète et évolutive. Elles apportent une solution aux problèmes non résolus du data lake.

La marketplace des données affine et étend l'intention initiale du data lake. Elle permet aux entreprises de connaître l'emplacement de toutes leurs données et d'assurer le suivi de leur utilisation. Elles savent ainsi quelles données sont utilisées et peuvent donc déterminer, de façon avisée, à quelles données consacrer leurs ressources afin de les améliorer. Elles savent également comment organiser la data dans l'ensemble de l'entreprise. La marketplace des données permet aux entreprises de répondre à des questions plus nombreuses et plus précises sur leur data et de rester compétitives en termes d'analytics.



L'évolution des référentiels de data

“ 451 Research estime que la nécessité de filtrer, de traiter, de transformer et de gérer les données, dans le but de les adapter aux multiples cas d'usage d'analytics, est cruciale pour produire de la valeur à partir du data lake. La gouvernance des données et la préparation de la data en libre-service constituent des éléments clés d'un data lake fonctionnel et des marketplaces des données associés. Par ailleurs, les insights et recommandations basés sur le machine learning représentent un aspect de plus en plus important dans l'accélération de la génération de valeur à partir des données de l'entreprise. »

MATT ASLETT  
451 RESEARCH

### Démocratisation des données : plus d'autonomie pour les utilisateurs

Les marketplaces des données favorisent la démocratisation des données, une nécessité absolue pour l'avenir de la data. La démocratisation des données signifie que les utilisateurs décident quelles données sont importantes et avec quelles données ils souhaitent travailler, sans devoir faire appel à un intermédiaire du service informatique. La marketplace des données porte donc bien son nom : il s'agit d'une place de marché dynamique, où les utilisateurs décident ce qui présente le plus de valeur. Les entreprises peuvent évaluer les données les plus populaires en fonction de leur utilisation, puis décider où elles souhaitent investir. Les utilisateurs peuvent acheter des produits data au sein même du marketplace, puis associer ces produits à d'autres produits qu'ils ont créés afin de lancer des projets data-driven réellement puissants.

## FONCTIONNALITÉS DE LA MARKETPLACE DES DONNÉES

La marketplace des données réunit toutes les fonctionnalités positives du data lake et du data warehouse en un seul produit qui répond à tous les besoins de l'entreprise en termes de data. En d'autres mots, la marketplace des données reprend, étend et améliore les fonctionnalités du data lake et du data warehouse en offrant :



**Toutes les données en une vue unique et simplifiée**



**Des données exploitables à la demande**



**Une gestion de la data adaptée à l'entreprise**



**Toutes les données regroupées dans une vue simplifiée** qui contient :

- Toutes les sources de données, validées et profilées
- Une évaluation de la qualité de chaque jeu de données, y compris de chaque colonne
- Un catalogue intelligent et intégré des métadonnées techniques et commerciales, qui organise, documente et décrit toutes les données collectées
- Une gestion automatisée de la data pour simplifier et ajuster la distribution des données prêtes pour l'analytics



**Des données exploitables à la demande.** Au lieu de recourir à une liste isolée et statique, les utilisateurs recherchent et récupèrent les données au moyen d'une interface navigable qui couvre tous les niveaux de compétences techniques. Les nouvelles données étant rapidement incorporées, les utilisateurs disposent toujours de données actualisées. Les fonctionnalités spécifiques doivent inclure, entre autres :

- La désignation des données « prêtes à l'emploi », où la qualité est garantie et les champs sensibles sont protégés
- Une expérience d'achat assistée, qui permet à l'utilisateur de trouver la meilleure data et de la comprendre
- La possibilité de réutiliser et de collaborer, afin que les analystes et les utilisateurs puissent facilement partager la data au sein de l'organisation
- Un suivi de traçabilité qui enregistre l'origine, l'évolution et la signification de la data



**Une gestion de la data adaptée à l'entreprise.** Contrairement au data lake, la marketplace des données est totalement mature au niveau architectural, ce qui signifie qu'elle intègre la sécurité, la gouvernance, les métadonnées et la protection des données. La marketplace des données est hautement évolutive et agile, elle offre :

- Une protection et un contrôle de bout en bout. Intégration automatique avec les structures de sécurité et de gouvernance de l'entreprise
- Une évolutivité intégrée. S'exécute en natif sur les plateformes de Big Data pour offrir des performances élevées et une évolutivité linéaire entre différentes configurations, notamment sur site, dans le Cloud, hybride et multi-cloud
- Une architecture ouverte. Assure un retour sur investissement rapide et s'intègre en toute transparence avec les systèmes déjà en place
- Une connectivité avec les applications d'analytics et les sources de données courantes, ainsi que la prise en charge des formats de data complexes et hérités

Non seulement la marketplace des données réunit les meilleures fonctionnalités du data warehouse et du data lake en une seule plateforme, mais elle offre également d'autres fonctionnalités. Surtout et avant tout, elle donne aux entreprises une agilité accrue sur leur data. Les données deviennent plus faciles à trouver et à comprendre puisqu'elles sont prêtes à l'emploi. Les données ne sont pas compartimentées et les utilisateurs ne dépendent pas du service informatique pour y accéder.

	DATA WAREHOUSES	DATA LAKES	MARKETPLACES DES DONNÉES
<p><b>TOUTES LES DONNÉES EN UNE VUE UNIQUE ET SIMPLIFIÉE</b></p> <hr/> <p>Vue intégrée de toutes les données. Documentée, précise et complète</p>	<p><b>F</b></p> <ul style="list-style-type: none"> <li>• Conçu pour répondre à des questions commerciales spécifiques</li> <li>• Limité aux données de BI et de reporting</li> <li>• Inaccessibilité de la plupart des données d'entreprise</li> <li>• Pas de Big Data accessible</li> </ul>	<p><b>C</b></p> <ul style="list-style-type: none"> <li>• N'inclut pas facilement toutes les données d'entreprise</li> <li>• Nouvelles données pas automatiquement vérifiées, contrôlées en termes de qualité ou documentées lors de l'ingestion</li> <li>• Résultat : la data du lake est opaque, contaminée et difficile à explorer</li> </ul>	<p><b>A</b></p> <ul style="list-style-type: none"> <li>• Inclut toutes les données de l'entreprise, des données brutes aux données prêtes à l'emploi</li> <li>• Smart Data Catalog incluant toutes les métadonnées techniques, opérationnelles et commerciales</li> </ul>
<p><b>DES DONNÉES EXPLOITABLES À LA DEMANDE</b></p> <hr/> <p>Libre-service. Demandes de nouvelles données satisfaites en quelques heures</p>	<p><b>C</b></p> <ul style="list-style-type: none"> <li>• Accès uniquement à la data dans le warehouse</li> <li>• Les nouvelles données prennent des mois</li> <li>• Réutilisation de data limitée sur plateforme unique</li> <li>• Réutilisation générale et collaboration impossibles</li> </ul>	<p><b>C</b></p> <ul style="list-style-type: none"> <li>• Accès impossible à la data pour la plupart des utilisateurs</li> <li>• Nécessite des compétences techniques et de programmation spécialisées.</li> <li>• Les utilisateurs techniques passent plus de 70 % du temps à trouver et préparer les jeux de données</li> <li>• Réutilisation générale et collaboration impossibles</li> </ul>	<p><b>A</b></p> <ul style="list-style-type: none"> <li>• Expérience d'achat de type Amazon</li> <li>• Les utilisateurs trouvent, comprennent et préparent eux-mêmes les jeux de données</li> <li>• Nouvelles données ajoutées en quelques heures</li> <li>• Permet une collaboration et une réutilisation générale</li> </ul>
<p><b>UNE GESTION DE LA DATA ADAPTÉE À L'ENTREPRISE</b></p> <hr/> <p>Sécurité, gouvernance, métadonnées et protection des données sensibles. Prise en charge des plateformes informatique actuelles et futures</p>	<p><b>A</b></p> <ul style="list-style-type: none"> <li>• Fonctionnalités complètes et matures de gestion de la data adaptée à l'entreprise</li> </ul>	<p><b>F</b></p> <ul style="list-style-type: none"> <li>• Manque de fonctionnalités de gestion de la data adaptée à l'entreprise</li> <li>• Tente de combler les lacunes en intégrant des produits open source ; travail supplémentaire requis</li> </ul>	<p><b>A</b></p> <ul style="list-style-type: none"> <li>• La marketplace des données offre des fonctionnalités complètes et matures de gestion de la data adaptée à l'entreprise</li> </ul>

## Enregistrement de données externes

La marketplace des données possède toute la « tuyauterie » nécessaire pour permettre l'évaluation, le catalogage, l'exploration et l'accès à l'ensemble des données, y compris la data hébergée hors du data lake. Ceci s'appelle l'enregistrement de la data. Il s'agit d'inclure toutes les données dans le catalogue, indépendamment de leur localisation dans l'entreprise, de traiter l'arrivée des nouveaux jeux de données de façon transparente, puis d'être capable de réutiliser ces jeux de données de façon économique.

Avec la marketplace des données, les entreprises bénéficient d'une vue unifiée des données dans le data lake, ainsi que des données enregistrées dans d'autres systèmes. Peu importe si les données se trouvent à l'intérieur ou à l'extérieur de l'entreprise, sur site ou dans le Cloud. Cela permet aux entreprises d'adopter une approche agile en matière de gestion de la data. Plutôt que de charger toutes les données dans le data lake avant de les utiliser, les analystes et autres intervenants clés peuvent effectuer une recherche dans le catalogue et exploiter les données issues de n'importe quelle source. Les demandes au sein de l'entreprise favorisent la maturation et la croissance de la marketplace.

Ce concept s'étend naturellement au-delà d'un simple data lake et des systèmes d'entreprise pour inclure une multitude de data lakes et de sources de data dans le Cloud. Avec la marketplace des données, les entreprises peuvent utiliser ce catalogue étendu pour voir les données qu'elles possèdent, et celles qu'elles n'ont pas. Elles peuvent analyser les métadonnées avant d'appeler leur service informatique pour qu'il provisionne la data entre les systèmes. Enfin, les entreprises peuvent, au fil du temps, surveiller les activités à forte valeur ajoutée associées à leur data, puis utiliser ces informations pour optimiser leur façon de gérer, de stocker, d'améliorer et « d'opérationnaliser » leurs données.

## Extraction automatique des métadonnées et des informations de profilage

La marketplace des données permet aux entreprises de procéder au profilage et à l'extraction automatique des métadonnées. Grâce aux métadonnées collectées automatiquement par la marketplace, vous pouvez constituer un référentiel qui vous permettra de « productiser » votre data de façon innovante.

L'analyse automatisée peut améliorer le catalogue de données en identifiant des modèles dans la data. Par exemple, la signature statistique d'une colonne ou d'un tableau peut être utilisée pour reconnaître et baliser des données sensibles, telles que les informations personnelles identifiables. La génération de métadonnées par l'analyse de métadonnées permet au système de gagner en intelligence au fil du temps. L'automatisation est une approche évolutive, adaptée à l'entreprise, pour la gestion de volumes de données en croissance constante et pour tirer une valeur ajoutée maximale des données de l'entreprise. Ces règles varient d'un secteur à l'autre, en fonction des besoins.

## Analyser les données d'utilisation

Quelles données sont adaptées à l'entreprise ? Il existe un signe absolument clair démontrant que les données sont prêtes : elles sont activement utilisées. Une marketplace des données collecte également les données d'utilisation. Vous êtes ainsi en mesure de gérer votre data par niveaux de valeur, en fonction de leur degré d'utilisation.

Vous pouvez alors mieux identifier vos priorités en termes de data, délivrer des résultats professionnels pertinents le plus tôt possible et concevoir rapidement des produits data.

Toutes ces métadonnées contribuent à une plus grande agilité du processus de développement. Avec un catalogue de data basé sur la marketplace, vous pouvez identifier les opportunités et les défis, vous pouvez hiérarchiser les cas d'usage en fonction de leur valeur afin d'obtenir une valeur ajoutée immédiate, tout en améliorant chaque fois un peu plus la facilité d'utilisation du data lake.

## La marketplace des données et l'écosystème du Big Data

Il est important de reconnaître que la marketplace des données ne prétend pas résoudre complètement tous les problèmes. Certaines technologies émergentes fournissent déjà un support pour le data lake, allant des tissus de données aux catalogues de data, en passant par les outils de préparation des données et de développement d'applications. Malheureusement, toutes ces bonnes idées ne forment pas un produit. La marketplace des données est le produit qui unit ces sources de données et outils afin de proposer aux entreprises la solution intégrée qu'elles recherchent. Mais une marketplace des données offre également une grande flexibilité : l'exploitation à la volée de cette donnée avec la technologie de votre choix.

## FAIRE VIVRE UNE MARKETPLACE DES DONNÉES

Pour qu'une marketplace des données fonctionne dans la pratique, les entreprises doivent mettre en place un processus de bout en bout qui s'appliquera systématiquement à toutes leurs données. Toutes les données doivent être cataloguées dans le data lake. Ainsi, le profilage et les autres formes d'analyses pourront générer un jeu complet de métadonnées décrivant ces données. Le catalogage exhaustif de la data permet également de réaliser l'analyse et l'exploitation des données au niveau de détail requis, quel qu'il soit.

L'ensemble complet de ces métadonnées permet la création d'un catalogue exhaustif de métadonnées, lequel facilite un accès global en libre-service. Les utilisateurs, quel que soit leur niveau technique, peuvent voir comment les autres utilisent les mêmes données et mobiliser celles-ci à bon escient, sans devoir faire appel au service informatique.

Avec une marketplace des données, vous pouvez conditionner et distribuer la data de façon plus efficace. Celle-ci sera alors mieux utilisée. Et surtout, en permettant leur accès en libre-service, la marketplace des données place la donnée dans les mains des utilisateurs métier. Ceci est crucial pour favoriser le changement et accélérer l'innovation en :

- 1 Réduisant les coûts de gestion des données
- 2 Accélérant les délais de réponse
- 3 Augmentant massivement la productivité des équipes d'analytics et des consommateurs de données

Étendre le nombre de personnes capables d'utiliser la data permet de collecter davantage de données d'utilisation. La marketplace gagne ainsi en efficacité.

La marketplace des données intègre les retours des « clients » (les consommateurs de données) afin d'identifier les jeux de données les plus utiles. Une fois l'ensemble de leurs données reliées au sein d'un catalogue, les entreprises peuvent déterminer, à partir des retours du marché, les données les plus utiles, mais aussi comment elles sont utilisées, nettoyées et améliorées.

Le suivi, tout au long de leur cycle de vie, des données brutes converties en données prêtes à l'emploi, mais aussi l'utilisation des retours, par le crowdsourcing, provenant des consommateurs de données, permettent aux entreprises d'identifier et de classer leurs jeux de données en fonction de leur valeur. La classification même de la data de cette façon permet de s'assurer que toutes les données présentent quelque utilité, quel que soit leur rang. En effet, ce qui n'était encore qu'un « marécage » de données devient à présent un ensemble organisé.

À partir de là, les entreprises peuvent investir dans la préservation de leurs données les plus précieuses. Les données qui se révèlent les plus utiles peuvent alors recevoir l'attention qu'elles méritent. Elles peuvent être nettoyées, façonnées, intégrées et converties en un produit facile à utiliser.

Étendre le nombre de personnes capables d'utiliser la data permet de collecter davantage de données d'utilisation. La marketplace gagne ainsi en efficacité.

Au final, la marketplace des données permet de distribuer à la fois des produits data et les formes brutes de la data qui ont mené à leur création. L'analyse automatisée peut améliorer le catalogue de données en explorant les modèles de data. Ceci permet de rationaliser l'organisation des données, en particulier des données sensibles. Le système tout entier gagne en intelligence au fil du temps.

Enfin, la marketplace des données ne limite ni l'usage ni l'expansion des outils existants. Elle unifie plutôt les outils de sécurité, de Big Data et de gouvernance au sein d'une supply chain commune et transparente.

Une fois qu'elles commencent à utiliser les marketplaces des données, les entreprises constatent rapidement l'efficacité supérieure de la marketplace par rapport à celle du data warehouse et du data lake, lesquels ressemblent davantage à des « marécages » de data. Les utilisateurs peuvent y entrer et y exploiter la data. Ils peuvent trouver les données dont ils ont besoin en quelques minutes, par eux-mêmes, et non plus soumettre des demandes au service informatique, sachant qu'une réponse pouvait prendre plusieurs jours ou plusieurs semaines, voire plusieurs mois. Les utilisateurs peuvent partager la data au sein de l'organisation, donnant ainsi de plus en plus de valeur aux données. L'organisation toute entière gagne en efficacité et en agilité avec sa data. Ceci est essentiel étant donné que de nouvelles sources de données continuent d'affluer.

Voici comment différents services de l'organisation tirent parti de la marketplace des données.

Les marketplaces des données encouragent les utilisateurs à partager la data au sein de l'organisation, donnant ainsi de plus en plus de valeur aux données.

## Gestionnaires data

Les rôles du personnel informatique, des data stewards et des ingénieurs data intègrent désormais de nouvelles dimensions en termes de visibilité, ainsi que de nouveaux processus pour les activités qui ont toujours été « ad hoc ». Les gestionnaires data peuvent désormais :

- Prendre de meilleures décisions quant à la façon d'allouer les ressources et agir lorsque la marketplace des données identifie une opportunité
- Faire en sorte que les jeux de données les plus utilisés deviennent plus utiles, plus propres et plus affinés
- Exécuter un processus agile de création et d'amélioration des jeux de données standardisés (produits data)
- Mieux allouer et cibler les ressources de formation et de conseil en fonction des signaux de la marketplace identifiant les données utilisées par les différents groupes
- Déterminer quels outils tiers sont les plus souvent utilisés par différents segments et lesquels ont le plus d'impact
- Accepter facilement l'ajout de nouvelles données dans la marketplace, étant donné qu'un processus défini permet désormais d'intégrer les jeux de données avec des informations descriptives complètes et possibilité de recherche. Le catalogue peut signaler lorsqu'une nouvelle source de données existe déjà dans la marketplace, afin d'éviter la prolifération inutile de la data. Les nouvelles données créent de la valeur, pas un « marécage »
- Enregistrer la data externe dans la marketplace pour étendre considérablement la portée du catalogue, encourager et assurer le suivi de l'utilisation des données qui ne sont pas stockées dans la marketplace

## Utilisateurs des données

Une marketplace des données change également l'expérience utilisateur de façon significative. Les utilisateurs peuvent désormais :

- Effectuer des recherches dans un catalogue complet, riche en métadonnées sur chaque jeu de données, y compris des informations de profilage
- Trouver des données externes qui sont enregistrées dans la marketplace
- Compiler leurs propres jeux de données
- Tirer parti des jeux de données créés par d'autres utilisateurs et collaborer avec eux
- Voir quelles données sont les plus populaires à partir des signaux de la marketplace
- Utiliser un nombre croissant de produits data organisés et proposés par les gestionnaires data.

## CONCLUSION

De façon générale, la marketplace des données est simplement une dénomination pour un nouveau paradigme de gestion de la data, nécessaire pour faire face au volume croissant de données aujourd'hui disponibles. Les entreprises peuvent choisir d'ignorer ces nouvelles données ou de les gérer avec d'anciens processus devenus de véritables goulets d'étranglement. Ce qui signifie qu'elles renoncent à en tirer profit.

Celles qui choisissent de trouver un moyen de prendre le contrôle sur cette problématique doivent faire face à certaines réalités. Pour réussir, elles devront trouver la façon de :

- Permettre la collecte, le profilage et la description des nouvelles données et permettre aux utilisateurs de les trouver
- Façonner la data dans les formats les plus conviviaux
- Faire en sorte que les données les plus populaires puissent être trouvées par, ou recommandées à, ceux qui pourraient en avoir besoin
- Impliquer davantage de personnes dans ce processus global

La vision de la marketplace des données peut faire de tout cela une réalité. C'est un excellent point de départ pour résoudre un problème d'une complexité effrayante.

La marketplace des données transforme un problème (gérer l'avalanche de nouvelles données) en une opportunité (délivrer davantage de valeur à l'entreprise). Si vous pensez que vous n'avez pas les moyens de gérer cette transformation, il est peut-être temps d'envisager la mise en place d'une marketplace des données.

Ce livre blanc a été rédigé par Early Adopter Research en partenariat avec Qlik

Contactez-nous



© 2019 Early Adopter Research

