

# Talan<sup>★</sup>

## **DATA AS A PRODUCT (DAAP) : UN MODÈLE POUR PROFITER DE LA VALEUR QUE LA DATA PEUT APPORTER**

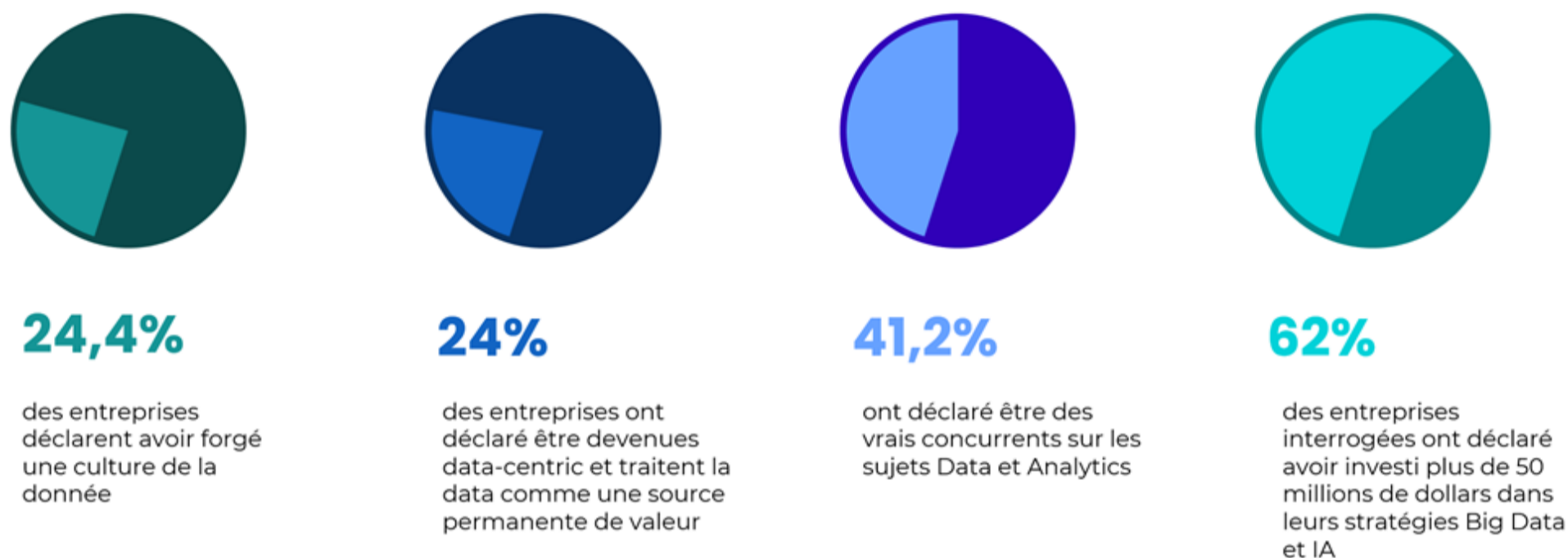
**B**ien que la grosse majorité des entreprises reconnaissent le pouvoir des données, la plupart ont du mal à libérer tous leur potentiel. Les investissements dans la data doivent le plus souvent fournir simultanément une valeur à court terme et en même temps jeter les bases d'utilisations futures, alors même que les technologies de données évoluent de manière imprévisible, que de nouveaux types de données émergent et que le volume de données ne cesse d'augmenter.

En règle générale, les entreprises lancent des programmes massifs pour répondre aux besoins des utilisateurs finaux à travers des pipelines de données qui deviennent souvent fragiles et problématiques, en créant leurs propres systèmes de liaison point à point. De plus, cela génère des difficultés organisationnelles plus larges, telles que des désaccords entre différents départements d'une entreprise sur les faits essentiels du business.



En effet, dans les rapports annuels de [NewVantage Partners](#), une enquête annuelle menée auprès des cadres supérieurs des entreprises sur les thèmes de l'adoption des données et de l'IA par les entreprises, a montré une tendance à mettre beaucoup d'efforts et d'investissements dans la création de plateformes Data et Analytics, qui connaissent pourtant de faibles taux de réussite (**figure 1**).

#### Rapport NewVantage Partner 2021



**Figure 1 : Les résultats de l'enquête reflètent les challenges auxquels les entreprises sont confrontées**

Inversement, les entreprises qui réussissent à tirer de la valeur des données le font en traitant les données comme un produit commercial. Lorsqu'une entreprise développe un produit, elle essaie de maximiser les ventes en répondant aux besoins d'autant de types de clients que possible, souvent en créant une offre standard qui peut être adaptée à différents utilisateurs.

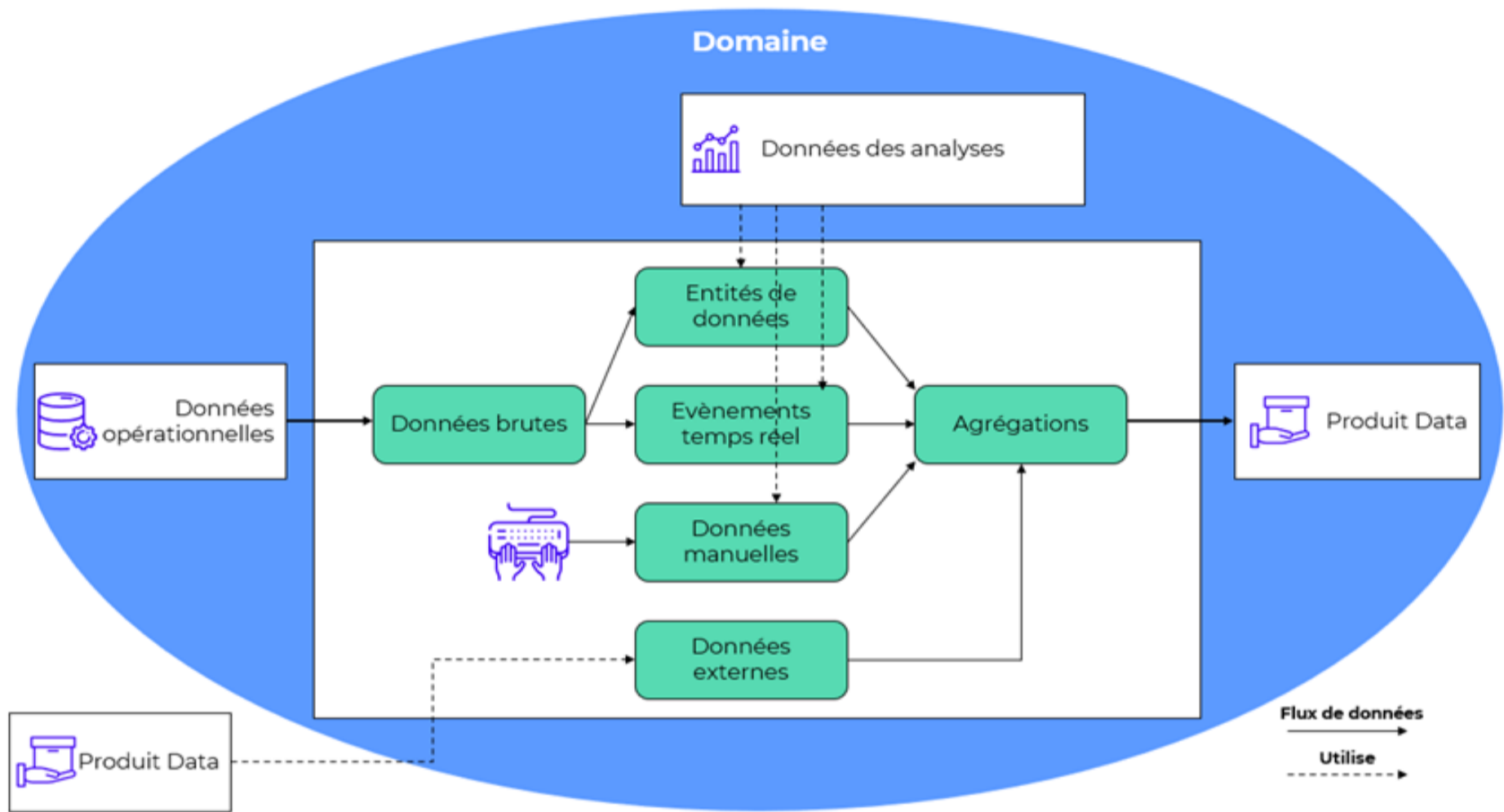
Les produits de données fonctionnent de la même manière, leur rôle est de fournir les moyens pour développer l'activité et améliorer les ROI, que ce soit à travers l'aide à la décision, une vision client 360, la personnalisation de produits et services, etc.

### Qu'est-ce qu'un produit data ?

Un produit data fournit un ensemble de données de haute qualité, prêt à l'emploi, auquel les personnes d'une organisation peuvent facilement accéder et les appliquer à différents challenges métiers. En effet, tout ce qui est poussé vers un «environnement data de production» auquel l'entreprise peut accéder est un produit. Cela peut être une vision client 360 avec toutes les informations sur les interactions avec le service client, les transactions, etc. De même, un terminal des données financières et des analyses est également un produit data.

Les produits data devraient être créés autour des domaines métiers - dans l'e-commerce, on trouve par exemple: commandes, livraisons, clients, etc. - et adressés aux consommateurs de données et donc traités comme des API publiques, destinés aux data scientists et aux data analysts, qui devraient être traités comme des clients, pour lesquels

il faudrait garantir une totale satisfaction. Une caractéristique principale des produits data est leur interopérabilité avec d'autres produits data dans l'écosystème, ce qui permet d'avoir une meilleure connectivité entre eux et de les combiner pour créer de nouveaux produits.



**Figure 2 : Flux de données de création d'un produit data**

Dans la **figure 1**, nous pouvons voir les flux de données qui mènent à la création des produits data. Les données opérationnelles sont souvent ingérées sous forme de données brutes et non structurées.

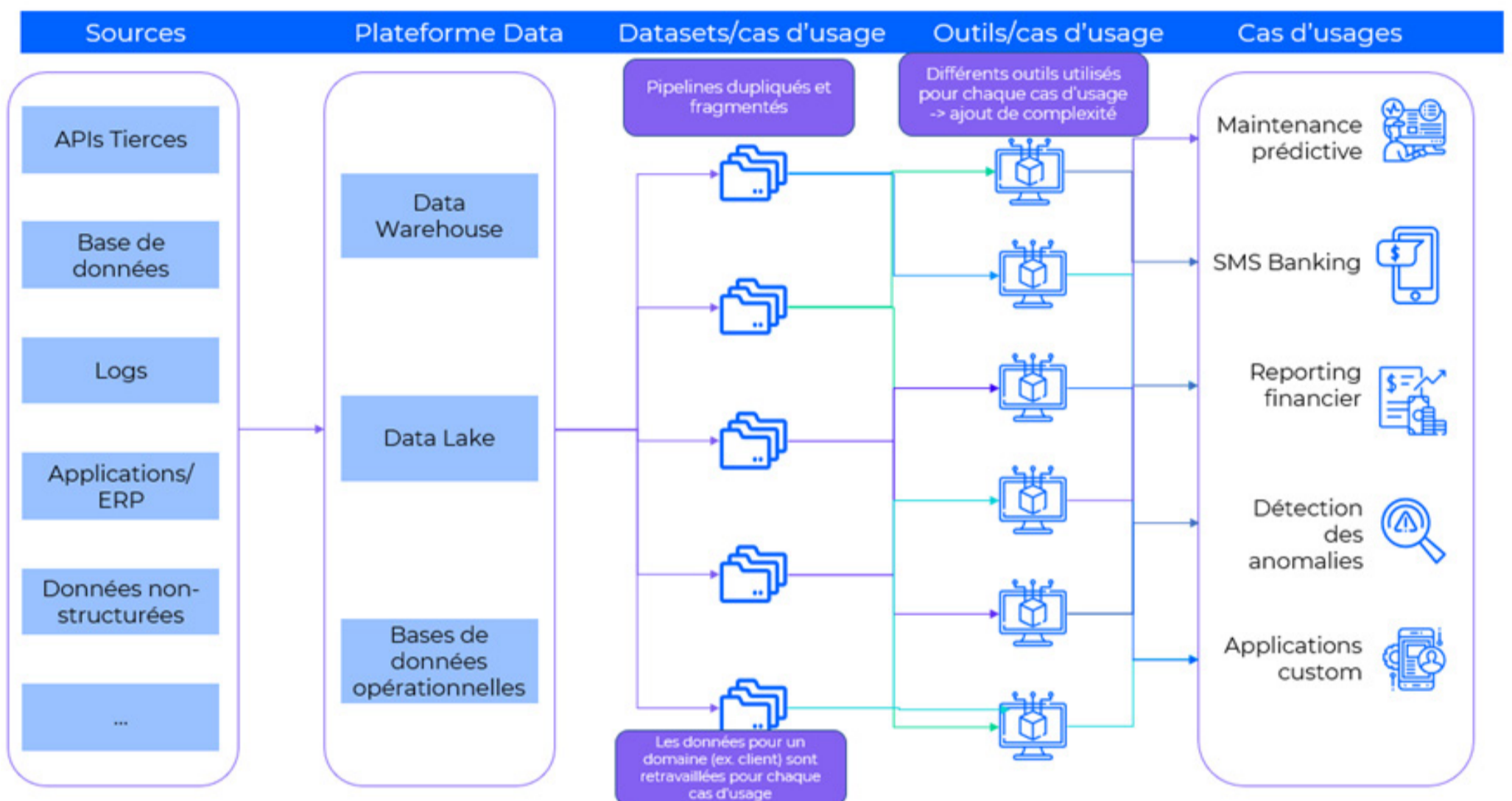
Dans une étape de prétraitement, les données brutes sont nettoyées et structurées en événements et entités. Les événements, par exemple la validation d'une commande ou la livraison d'un colis, sont souvent ingérés en temps réel. Les entités de données représentent des objets métiers tels que des expéditions ou des articles dont l'état change au fil du temps.

En pratique, nous voyons souvent des données saisies manuellement ou importées. Par exemple, des saisies manuelles dans les tableurs, ou des données textuelles telles que les emails. Les produits data peuvent également être créés à partir de d'autres produits de données bien gouvernés. Ils sont dérivés de l'agrégation des événements, des entités, des données manuelles et externes.



## Comparaison du modèle data traditionnel et du modèle Data as a Product

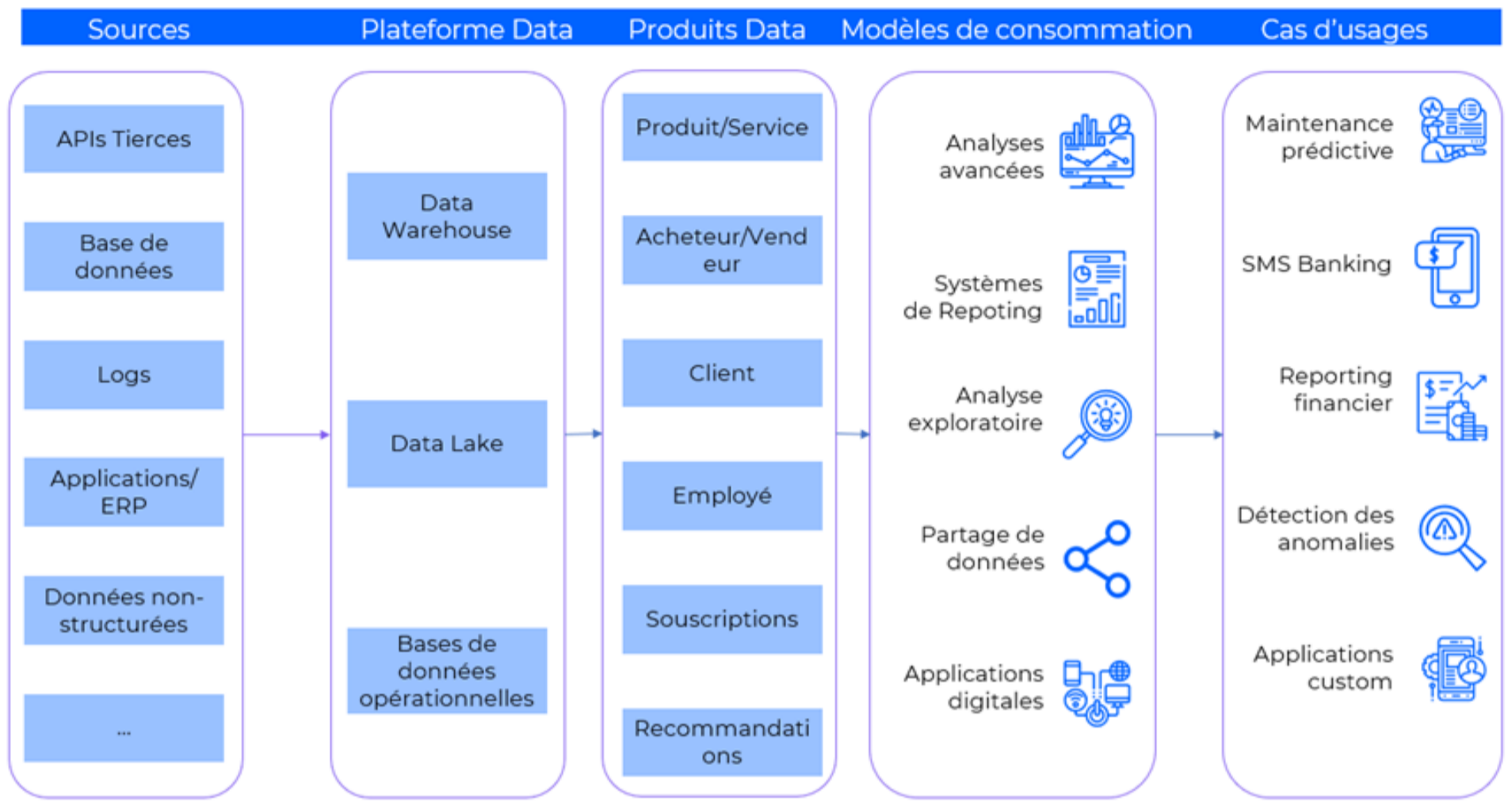
Dans le cas de l'approche traditionnelle, les équipes créent des pipelines de données à partir des sources et adaptent ces données pour leurs propres cas d'usage. Ceci implique des étapes redondantes et dupliquées et nous nous retrouvons face à un nombre impactant de pipelines non gouvernés, incapables de se communiquer entre eux.



<sup>1</sup>Figure 3 : Modèle data traditionnel

Le Data Lake est un espace de stockage qui reçoit tous les types de données : structurées (Données de bases relationnelles, feuilles de calcul...) ou non structurées (texte, audio, vidéo...) sans qu'il ait besoin de respecter un schéma spécifique pour les flux entrants. Au contraire, un Data Warehouse ne conserve que des données structurées et est destiné à la BI (Business Intelligence) et à la prise de décision<sup>1</sup>

Tandis qu'avec le modèle Data as a Product, les équipes créent des solutions pour leurs propre cas d'usages en exploitant les produits data déjà créés d'une manière standardisée et qui peuvent être consommés de plusieurs façons (Analyses avancées, Reporting, etc.). Ceci permet de simplifier l'architecture et de réduire le temps pour réaliser de la valeur à partir des données.



**Figure 4 : Modèle Data as a Product**

## Comment appliquer l'approche Data as a Product ?

### Avoir l'alignement des parties prenantes sur les produits à développer

Comme la data est aussi un produit, les clients internes sont également les parties prenantes. Il est donc indispensable de faire du partenariat avec les principaux consommateurs de données lors de la création de la feuille de route de produits data, développer des SLAs (Service-Level Agreements), standardiser le format des produits data pour améliorer leur interopérabilité et gérer la data comme un produit.

1

### Appliquer les méthodes de gestion de produits

Comprendre les problèmes et les besoins du client (consommateurs des données). Faire un cadrage du périmètre des projets data et les décomposer en sprints itératifs, et développer des produits data qui non seulement résolvent ces problèmes mais vont au-delà des attentes et peuvent servir comme base d'entrée pour d'autres produits data.

2

## Mettre en place une infrastructure libre-service

Mettre à disposition des consommateurs de données des interfaces standardisées pour découvrir et accéder à tous les produits data en se basant sur une infrastructure data libre-service, et rendre faisable pour les équipes de chaque domaine de gérer le cycle de vie de leurs produits de données en toute autonomie.

# 3

## Prioriser une gouvernance de données fédérée

Mettre en place des normes et politiques et de gouvernance de données définies au niveau de toute l'entreprise, ainsi que des processus de gestion de la qualité et la disponibilité des données et du cycle de vie data. En parallèle, il faut permettre aux équipes responsables des produits data d'avoir un large degré d'autonomie quant à la façon dont ces normes et politiques sont mises en œuvre pour leurs produits, et définir leurs propres SLAs, SLOs (Service-Level Objectives) et SLIs (Service-Level Indicators).

# 4

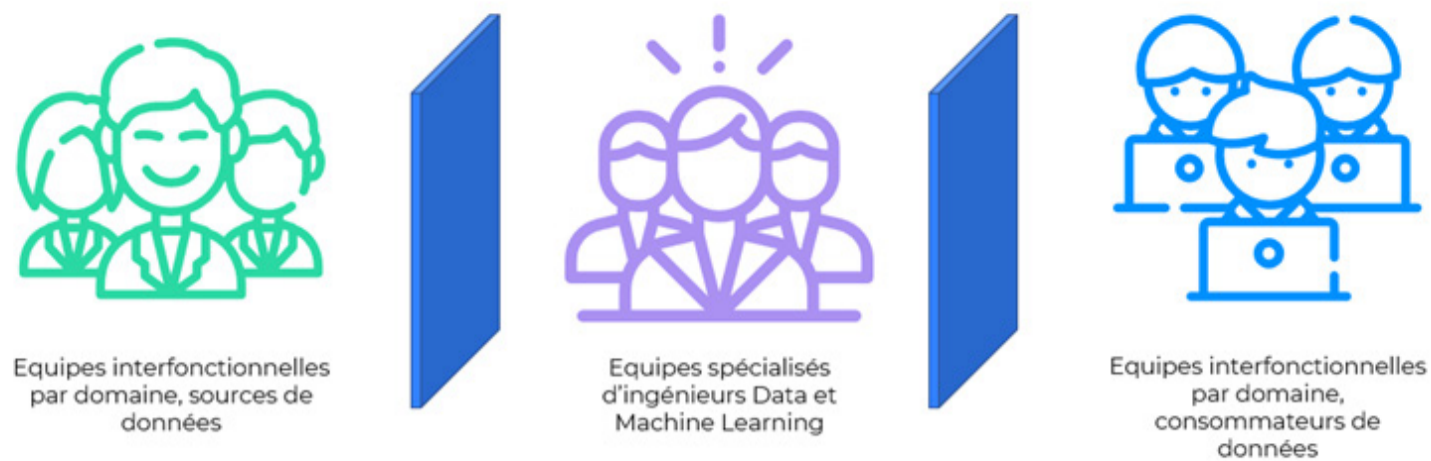
## Choisir la bonne structure d'équipe

Plusieurs styles d'organisations peuvent être envisagés en fonction de la taille de l'entreprise et les besoins métiers. En général, nous préconisons une organisation hybride avec une équipe data centrale qui gère l'infrastructure data et fournit les données sources et des équipes par domaines qui se focalisent sur la création de produits data liés à leurs domaines métiers, qui s'approprient leurs données et en sont responsables.

# 5

## Comment transformer votre organisation pour implémenter ce modèle ?

Lorsque nous nous rapprochons pour observer les organisations typiques autour des plateformes data, ce que nous trouvons est une équipe data ingénieurs hyper-spécialisés et cloisonnés des unités opérationnelles qui génèrent ou utilisent les données. De plus, ces équipes d'ingénieurs sont souvent formées en fonctions des expertises techniques de chacun et manquent des connaissances métiers. **(Figure 5)**



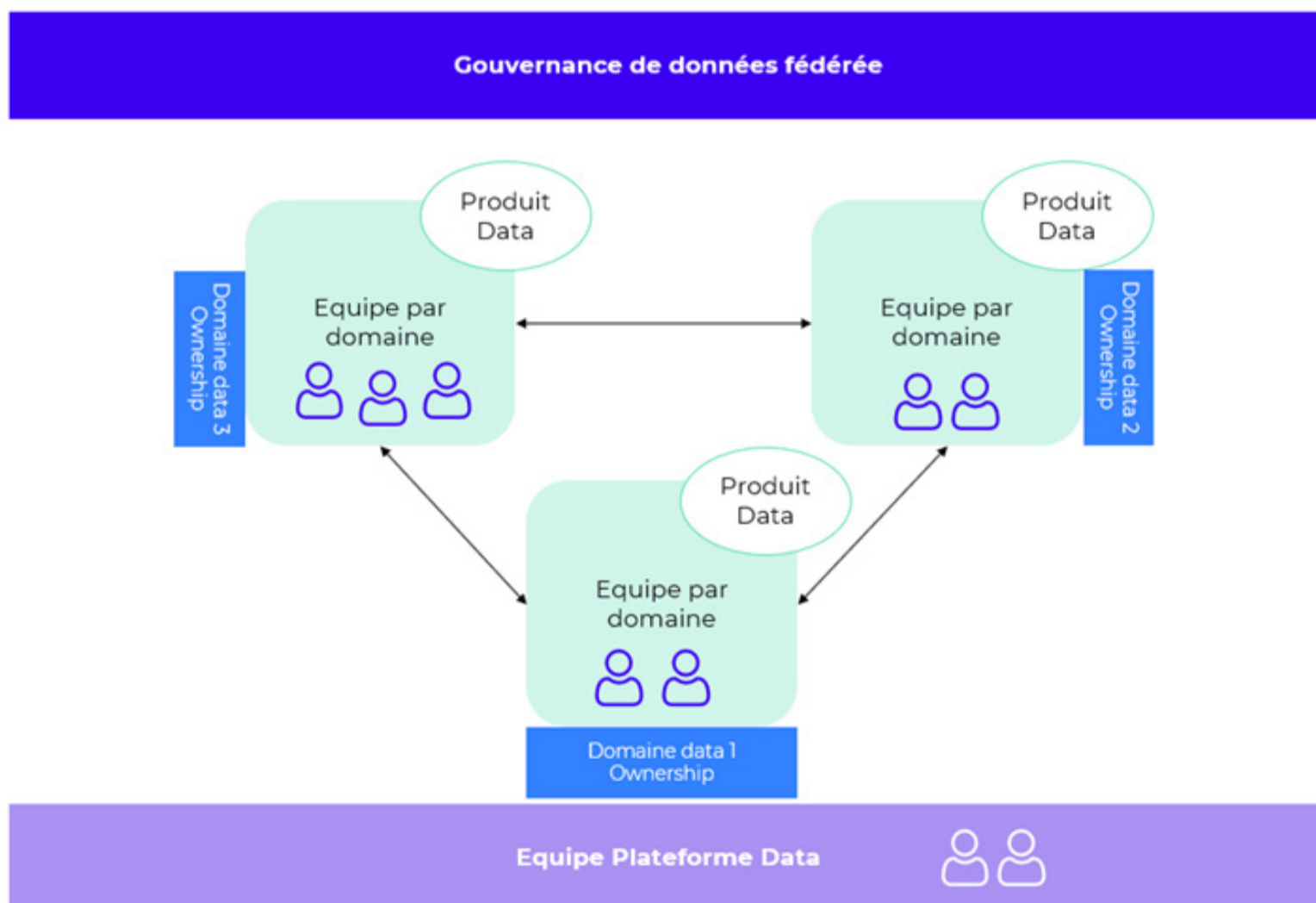
**Figure 5 : Organisation traditionnelle avec des équipes hyper-spécialisées et cloisonnées**

En réalité, nous nous retrouvons face à des équipes productrices de données déconnectées les unes des autres et des consommateurs de données frustrés qui se battent pour une place prioritaire dans le backlog des équipes de data ingénieurs déjà surchargés.

Nous préconisons une organisation plus adaptée dans la **figure 6**. Cette approche décentralisée permet aux équipes de chaque domaine d'effectuer

elles-mêmes leurs propres analyses de données.

Chaque domaine à son équipe et ses données opérationnelles et analytiques. L'équipe du domaine ingère des données opérationnelles et construit des modèles de données pour effectuer ses propres analyses et créer des produits data en fonction des besoins d'autres domaines.



**Figure 6 : Modèle opérationnel efficace pour organiser les responsabilités autour de la data**



Les représentants des équipes des domaines se réunissent pour définir les normes et les politiques globales de gouvernance de données. Ces règles servent comme base pour contrôler la création des produits data.

Ils couvrent des aspects tels que l'interopérabilité pour standardiser les méthodes de consommation des produits data ainsi que leurs formats, et les préparer également à être exposé à l'extérieur à des fins de partage et de monétisation à la consommation. Les normes de documentation des produits data pour les rendre plus faciles à découvrir et à comprendre. Les normes et responsabilités pour sécuriser l'accès aux données en définissant des processus IAM (Identity Access Management).

Ils établissent également des règles de conformité et de protection des données personnelles.

Pour unifier la stack technologique utilisée pour la création des produits data, une plateforme data libre-service et gérée par les équipes Data et Infrastructure centrales peut être mise en place. Cette plateforme est équipée des outils et capacités analytiques permettant aux équipes des domaines de construire les modèles d'analyse et de prise de décision. Des capacités de création, découverte, monitoring et accessibilité aux produits data, ainsi que l'automatisation de l'application des politiques globales de data gouvernance. L'équipe Infrastructure a pour rôle de créer les outils nécessaires pour rendre les équipes de chaque domaine, autonomes pour consommer et créer les produits data.

Une architecture type permettant de mettre en avant ce modèle de données est celle apportée par le Data Mesh : une nouvelle approche pour concevoir des architectures de données modernes en adoptant des changements organisationnels ainsi qu'une approche data-centric. L'idée est que les données doivent être facilement accessibles et interconnectées dans l'ensemble de l'entreprise et au-delà. Et plus important, elles doivent apporter une valeur métier !

## Références :

- <https://developer.confluent.io/learn-kafka/data-mesh/data-ownership/>
- [https://www.montecarlodata.com/blog-how-to-treat-your-data-like-a-product/#:~:text=%E2%80%9CData%20as%20a%20Product%20\(DaaP,personalized%20products%2C%20or%20detecting%20fraud](https://www.montecarlodata.com/blog-how-to-treat-your-data-like-a-product/#:~:text=%E2%80%9CData%20as%20a%20Product%20(DaaP,personalized%20products%2C%20or%20detecting%20fraud)
- <https://hbr.org/2022/07/a-better-way-to-put-your-data-to-work>
- <https://apnews.com/press-release/business-wire/technology-business-stock-markets-north-america-lung-disease-6c256b3af23f4f04978befb6e671e56d>
- <https://www.datamesh-architecture.com/#why>
- Zhamak, D. (2022). Data Mesh: Delivering Data-Driven Value at Scale