



DATA MESH

LE B.A.-BA D'UNE STRATÉGIE RÉUSSIE

 INVENTIV·IT
www.inventiv-it.fr

Inventiv IT est un accélérateur de votre transformation digitale, et vous propulse vers l'excellence opérationnelle en décuplant votre capacité à exploiter la data et les solutions technologiques.

EDITO

Comment accéder facilement à des données se trouvant dans différents départements ? Nous nous sommes posé cette question depuis des décennies jusqu'à convenir qu'il était impératif de repenser notre approche de la gestion et de l'exploitation de nos ressources numériques. C'est dans ce contexte que le concept de Data Mesh émerge comme une réponse innovante et révolutionnaire.

Le Data Mesh ne se contente pas de suivre les chemins déjà tracés par les architectures traditionnelles de gestion des données. Au contraire, il remet en question les fondements même de notre façon de penser et d'agir en matière de données.

En reconnaissant l'existence des silos de données et en les intégrant dans une approche globale et cohérente, le Data Mesh ouvre de nouvelles perspectives pour les entreprises désireuses de devenir véritablement data driven.

Dans ce livre blanc, nous vous invitons à naviguer dans l'univers du maillage des données. De ses origines à ses applications concrètes, en passant par ses principes fondamentaux et ses défis, nous explorerons ensemble les multiples facettes de ce cadre de travail novateur.

Que vous soyez un expert de la donnée ou simplement curieux de découvrir les dernières tendances en matière de gestion des données, ce livre blanc vous fournira les connaissances et les perspectives nécessaires pour comprendre et apprécier pleinement l'importance du data mesh dans le business d'aujourd'hui et de demain.

Cette exploration du maillage des données vous inspirera et vous incitera à repenser votre approche de la gestion des données au sein de votre organisation. Car, comme le dit si bien l'adage, dans l'ère du Big Data, celui qui maîtrise ses données maîtrise son destin.

Bonne lecture
L'équipe de rédaction - Data Fabric



*Du latin **obrir** : «faire que ce qui était fermé ne le soit plus», «déplacer ce qui empêche le libre passage», «donner accès à»; ouvert; open. Un mot tellement usuel et pourtant, à lire son étymologie et son histoire (sa lignée ? ;-), il semble presque à lui seul définir l'approche "Data Mesh".*

En effet, comment espérer innover sans être ouvert sur l'avenir, collaborer sans être ouvert aux autres, grandir sans être ouvert au changement, bref prospérer. Encore faut-il pouvoir mesurer, mesurer pour comparer afin d'améliorer, améliorer l'utilisation de ses ressources, améliorer ses performances, bref améliorer sa productivité.

Pour mesurer il faut de la Donnée, de la Donnée de qualité donc gouvernée, que seuls les métiers connaissent, que seule une plateforme polyglotte, multcloud, sécurisée et ouverte à tous les formats peut intégrer et servir; bref une bonne alchimie d'Hommes, de méthodes et de technologies; afin de réaliser ce maillage efficient de l'information, cher à Zhamak Dehghani, pour permettre aux grandes entreprises de se réinventer et aux moins grandes de réussir leur passage à l'échelle, pour construire ce deuxième quart de siècle qui commence.

Paris, le 5 juin 2024

Hugues Journeau
Solutions Architect,
Databricks France



SOMMAIRE

Introduction	Pages 03
1. Evolution vers le data mesh : origines et trajectoire	Pages 04-08
Origines du Data Mesh.....	05
Principes fondamentaux du Data Mesh.....	06
Mesure de la réussite et retour sur investissement du Data Mesh.....	07
Comparaison avec les approches traditionnelles : Data Lake, Data Warehouse et Data Mart.....	08
2. Démarche de mise en oeuvre - Use Case d'un leader du transport en IDF	Pages 09-14
Use case d'un leader du transport en IDF	10
Challenges et objectifs du projet Data Mesh	11
Etapes pour adopter le Data Mesh.....	12
3. Les piliers du data mesh	Page 14
Propriété décentralisée de domaine des données	
La Data en tant que produit	
Infrastructure de données en libre-service en tant que plateforme	
Gouvernance des données informatique fédérée	
4. Technologies et outils associés au data mesh	Pages 15-18
Solutions selon les étapes de construction du Data Mesh	16
Focus sur les avantages d'utiliser Databricks.....	17 - 18
5. Evolutions du Data Mesh	Page 19
Conclusion	Page 20
Pour aller plus loin	Page 21

INTRODUCTION AU DATA MESH

Le Data Mesh ou maillage de données en français, représente bien plus qu'une simple approche technologique. C'est un cadre de travail complet qui vise à transformer les organisations en entités véritablement orientées données, capables de tirer parti de la richesse de leurs ressources informationnelles pour prendre des décisions éclairées et rester compétitives sur le marché.

Depuis des décennies, les entreprises luttent contre ce que l'on appelle la « tragédie des silos », où les données sont cloisonnées dans différents départements et systèmes, rendant difficile leur accès et leur utilisation à des fins stratégiques. Des solutions telles que les Data Lakes, Data Warehouses et Data Hubs ont été développées pour centraliser ces données dispersées, mais elles ont souvent échoué à résoudre complètement le problème.

Contrairement à ces architectures centralisées, le Data Mesh adopte une approche radicalement différente. Plutôt que de chercher à éliminer les silos de données, il reconnaît leur existence et cherche à les exploiter de manière efficace pour créer une organisation agile et centrée sur les données. Le Data Mesh ne se limite pas à une technologie spécifique, mais combine des aspects technologiques, des processus, des compétences et une gouvernance appropriée pour transformer une entreprise en une entité réellement « data driven » et capable de valoriser son patrimoine data.

1. EVOLUTION VERS LE DATA MESH : ORIGINES ET TRAJECTOIRE

Dans cette première partie, nous allons plonger au cœur du Data Mesh en définissant clairement ce qu'il est et en explorant ses concepts fondamentaux.

Nous discuterons également de la mesure de la réussite et ROI de l'approche et des points distinctifs avec leurs architectures traditionnelles.

Cette introduction théorique servira de base pour comprendre les sections suivantes du livre blanc, où nous aborderons des aspects plus pratiques et techniques de la mise en œuvre du maillage des données.

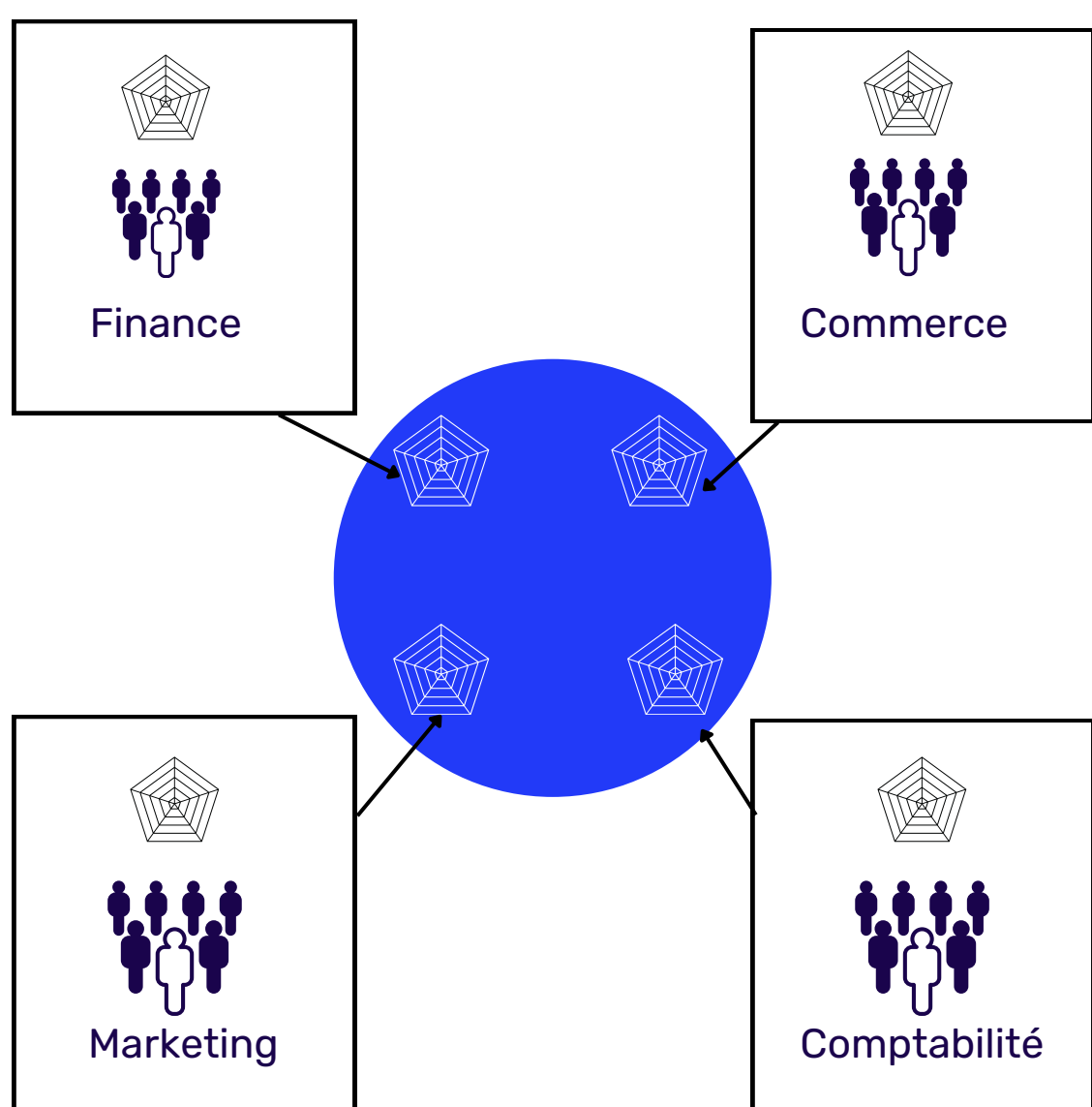
En vous familiarisant avec ces concepts, vous serez mieux préparé à envisager l'adoption de cette approche dans votre organisation et à saisir les opportunités qu'elle offre pour améliorer la gestion et l'exploitation de vos données.

ORIGINES ET PRINCIPES FONDAMENTAUX DU DATA MESH



01 Origines du Data mesh

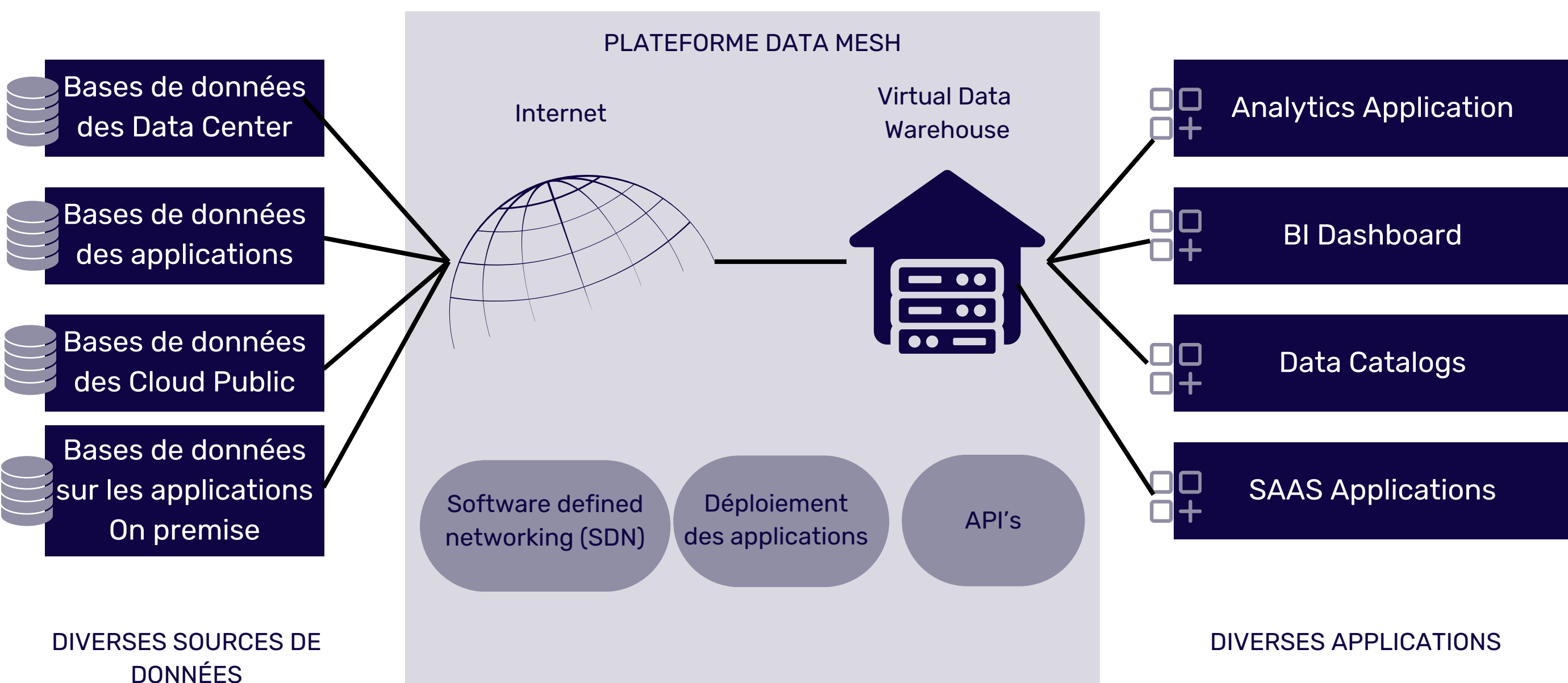
Schema 1 : Gestion autonome des données par département



Le Data Mesh représente une approche révolutionnaire dans la gestion et l'exploitation des données au sein des organisations. Contrairement aux approches traditionnelles centralisées, où les données sont souvent considérées comme un actif appartenant à un seul département ou service, le Data Mesh reconnaît la nature distribuée et complexe des données dans les entreprises modernes.

Ce concept remonte à Zhamak Dehghani, qui a présenté pour la première fois cette idée lors d'une conférence en 2018. Depuis lors, le concept a gagné en popularité et a suscité un grand intérêt dans la communauté des données. Il s'agit essentiellement d'une réponse aux défis posés par les énormes volumes de données, leur diversité croissante et la nécessité de les exploiter de manière efficace et évolutive.

Schéma 2 : Exemple de Plateforme Data Mesh



ORIGINES ET PRINCIPES FONDAMENTAUX DU DATA MESH



02 Principes fondamentaux du Data Mesh

Le Data Mesh repose sur quatre principes fondamentaux qui révolutionnent la manière dont les organisations gèrent et exploitent leurs données.

Distribution des responsabilités

Le Data Mesh attribue la responsabilité de la gestion des données à chaque équipe au sein de l'organisation. Chaque équipe est responsable de la collecte, du traitement et de la qualité des données qu'elle produit. Cela favorise une meilleure compréhension des besoins métier et améliore la qualité des données



Automatisation des pipelines des données

Le Data Mesh préconise l'automatisation des processus d'ingestion, de traitement et de préparation des données. Cette automatisation réduit les erreurs manuelles, améliore l'efficacité opérationnelle et permet de déployer rapidement de nouvelles fonctionnalités.



Une entreprise Data Driven favorise une culture de la donnée et une collaboration interpersonnelle

Décentralisation de l'architecture technique

Le Data Mesh adopte une architecture distribuée au lieu de le centraliser. Les données sont gérées et traitées là où elles sont produites. Il améliore la scalabilité, la résilience et les performances du système.



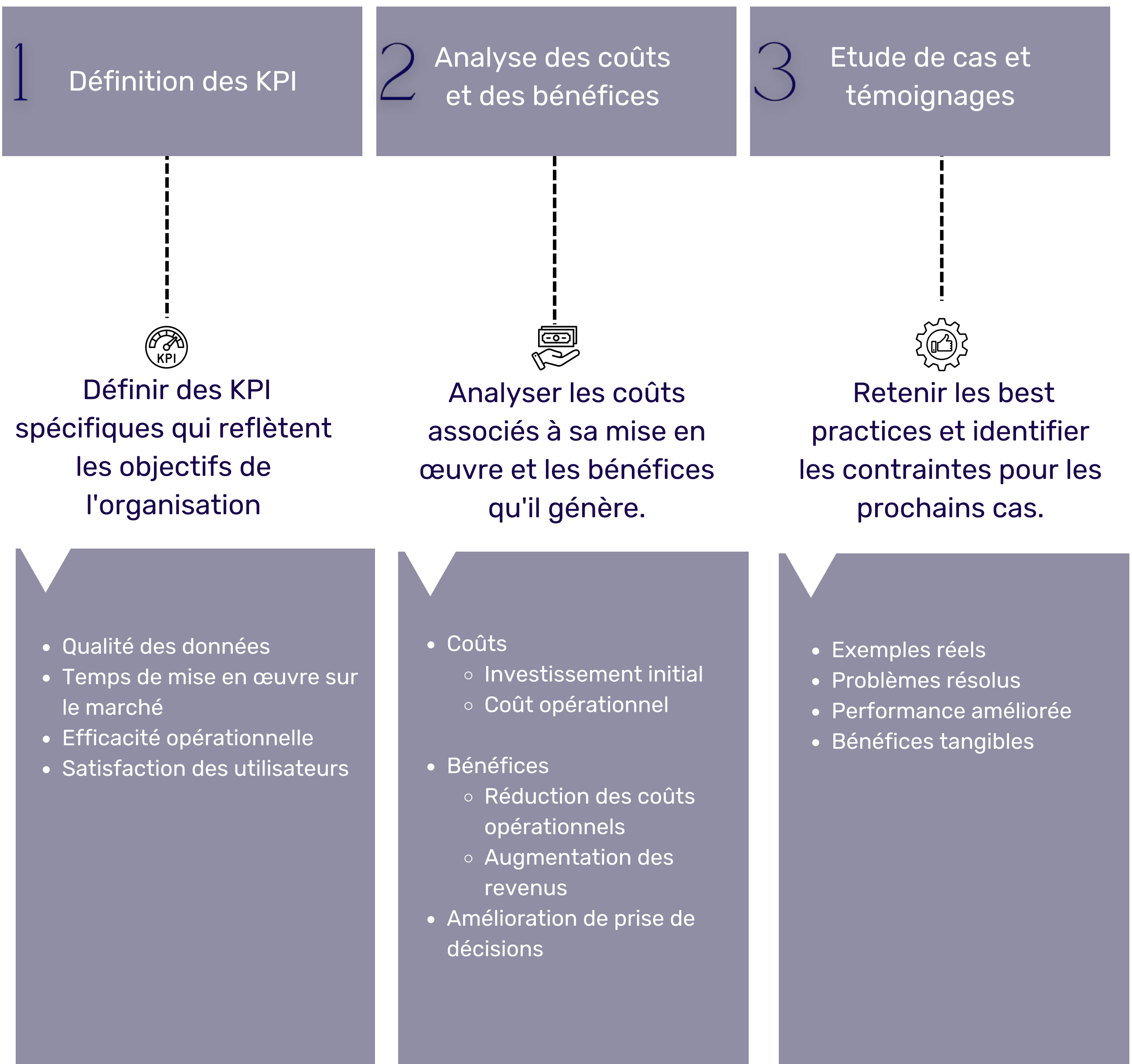
Culture de la donnée

Le Data Mesh encourage une culture d'entreprise centrée sur la donnée. Les data owners doivent collaborer étroitement avec d'autres parties prenantes, y compris les data stewards, les ingénieurs de données et les analystes, pour garantir une gestion cohérente et efficace des données.

ORIGINES ET PRINCIPES FONDAMENTAUX DU DATA MESH



03 Mesure de la réussite et ROI du data mesh



ORIGINES ET PRINCIPES FONDAMENTAUX DU DATA MESH

Tableau comparatif des architectures de données

Architecture	Approche	Model
Data Mesh	Approche décentralisée de la gestion des données, où chaque domaine gère ses propres données comme des produits	
Data Lake	Grand dépôt de données brutes, non transformées.	
Data Warehouse	Dépôt centralisé de données structurées, optimisées pour l'analyse	
Data Mart	Sous-ensemble d'un data warehouse pour des besoins spécifiques	

Bien que les architectures traditionnelles comme les Data Lakes et les Data Warehouses aient leurs mérites et continuent d'être utilisées efficacement dans de nombreux contextes, le Data Mesh représente une évolution vers une gestion des données plus agile, scalable et centrée sur les besoins métier.

2. DEMARCHE DE MISE EN OEUVRE DU DATA MESH

Après avoir exploré les fondements théoriques et les principes clés du Data Mesh, nous allons nous pencher sur les aspects pratiques de sa mise en œuvre. La transition vers un modèle de Data Mesh représente une transformation significative pour toute organisation, impliquant non seulement des changements technologiques mais aussi une réorganisation culturelle et structurelle.

Dans cette section, nous aborderons les étapes clés nécessaires pour adopter efficacement le Data Mesh, illustrées par un cas d'étude concret d'entreprise ayant réussi cette transition.

Nous examinerons également les défis courants auxquels votre organisation peut être confrontée.

DE LA THÉORIE À LA PRATIQUE DU DATA MESH

01

CAS D'USAGE : MISE EN OEUVRE DU DATA MESH CHEZ UN LEADER DU TRANSPORT EN ILE-DE-FRANCE

LE CLIENT EN QUELQUES CHIFFRES

PLUS DE 70 000 COLLABORATEURS
PLUS DE 270 MÉTIERS DIFFÉRENTS

+ 3 MILLIARDS DE VOYAGEURS
CHAQUE ANNÉE

LEADER DU TRANSPORT EN IDF
AVEC UN CA DE PLUS DE 6MDS €

PROBLÉMATIQUE

Inventiv IT a accompagné un des leaders de transport en Ile-de-France qui faisait face à une multitude de projets de gestion des données (8 projets au total) avec des données éparpillées et mal structurées. Cette dispersion entraînait des difficultés de coordination, de définition des processus et de structuration efficace des données. La complexité de l'organisation avec ses nombreux départements et vastes ensembles de données compliquait davantage la tâche de donner du sens à ces données dispersées.

SOLUTIONS

Pour répondre à ces défis, l'entreprise décide de mettre en place une plateforme basée sur le concept du Data Mesh. Cette approche permet de structurer et de coordonner les différents cas d'usage en silotant les ensembles de données par domaines spécifiques tels que les ressources humaines (RH), la propreté, la fraude, la gestion des stocks et des actifs.

Outils utilisés : Data Bricks, AWS

RÉSULTATS



Capacité d'analyse
des données
améliorée



Augmentation de
l'efficacité
opérationnelle



Réduction de la pression
sur les équipes centrales
de la gestion des données



Accessibilité
accrue des
données

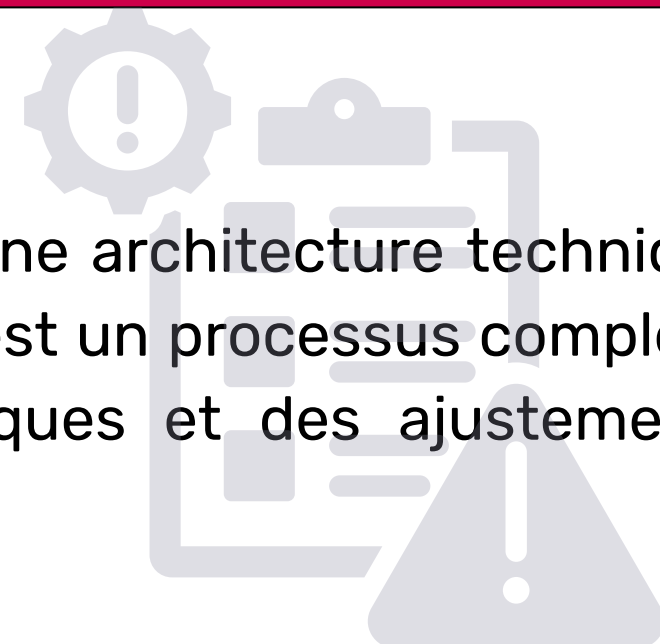
DE LA THÉORIE À LA PRATIQUE DU DATA MESH

02 CHALLENGES DE LA MISE EN ŒUVRE DU PROJET DATA MESH

Lors de la mise en place de la plateforme de Data Mesh, nous avons fait face à plusieurs défis. Nous vous avons identifié quelques challenges que vous pourriez rencontrer :

GESTION DES IMPRÉVUS

La mise en place d'une plateforme de Data Mesh nécessite une architecture technique robuste et bien pensée. La construction de cette plateforme est un processus complexe et chronophage, souvent ralenti par des imprévus techniques et des ajustements nécessaires



CONCEPT DE DATA MESH ENCORE NOUVEAU

Le Data Mesh étant une approche relativement nouvelle, l'adoption de nouvelles méthodologies requiert souvent un changement culturel. Certains collaborateurs peuvent être réticents ou manquer de connaissance sur le Data Mesh.

ABSENCE D'OWNERSHIP DE LA DONNÉE

Une des pierres angulaires du Data Mesh est la responsabilisation des équipes sur la gestion des données (data ownership). Dans beaucoup d'organisations, cette culture n'est pas encore instaurée, ce qui peut engendrer des difficultés pour déterminer qui est responsable de quoi.

GOVERNANCE DÉCENTRALISÉE

Assurer une gouvernance décentralisée des données tout en maintenant la sécurité et la conformité aux réglementations peut s'avérer délicat, en particulier dans des secteurs réglementés comme la finance ou la santé

DE LA THÉORIE À LA PRATIQUE DU DATA MESH

03 ÉTAPES DE MISE EN OEUVRE D'UN DATA MESH

La mise en place d'un data mesh est un processus complexe qui nécessite une planification minutieuse et une exécution coordonnée. Elle peut se faire en 8 étapes clés :



01

Compréhension et adoption des principes du data mesh

- Propriété décentralisée de domaine des données
- Data en tant que produit
- Infrastructure de données en libre-service en tant que plateforme
- Gouvernance des données informatique fédérée

Définir des politiques de gouvernance qui assurent la qualité, la sécurité et la conformité des données tout en permettant une certaine flexibilité.

Mettre en place des outils pour surveiller l'utilisation des données et appliquer les politiques de gouvernance



06

Gouvernance fédérée des données

- Etablissement des règles de gouvernance
- Outils de surveillance et de gestion



02

Évaluation et préparation de l'Infrastructure

- Evaluation des capacités actuelles
- Choix des outils et technologies

Analyser l'infrastructure existante pour identifier les lacunes et les besoins.

Sélectionner les technologies adaptées pour soutenir le data mesh (plateformes de gestion de données, les outils de transformation des données, et les solutions de sécurité)



04

Définition des domaines de données

- Identification des domaines
- Assignation des responsabilités

Définir clairement les domaines de données en fonction des besoins opérationnels et fonctionnels

Assigner des équipes responsables pour chaque domaine de données.

Mise en place de l'Infrastructure en libre-service

- Développement de plateformes de données
- Automatisation et standardisation



Construire ou intégrer des plateformes qui permettent aux équipes de gérer et de partager leurs données facilement

Mettre en œuvre des outils pour automatiser la gestion des données et standardiser les pratiques de traitement et de partage des données.



03

Formation et changement culturel

- Formation des équipes
- Promouvoir la culture du produit de données

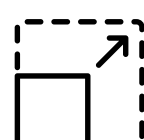
Former les équipes aux nouveaux outils, processus et responsabilités liés au data mesh

Encourager une mentalité où chaque équipe considère ses données comme un produit avec des utilisateurs finaux spécifiques

Échelle et élargissement

- Élargissement progressif
- Évaluation et Adaptation

08



Étendre le data mesh à d'autres domaines de données de manière progressive.

Continuer à évaluer l'efficacité du data mesh et apporter des améliorations continues.

Déploiement et itération

- Lancement pilote
- Feedback et amélioration Continue

Commencer par un déploiement pilote dans un ou plusieurs domaines pour tester et affiner l'approche.

Recueillir les retours d'expérience et ajuster les processus et les outils en conséquence.

La mise en place d'un data mesh nécessite une transformation organisationnelle et technologique significative. Elle implique la collaboration étroite entre les équipes de données et les domaines fonctionnels pour assurer un alignement stratégique et opérationnel.

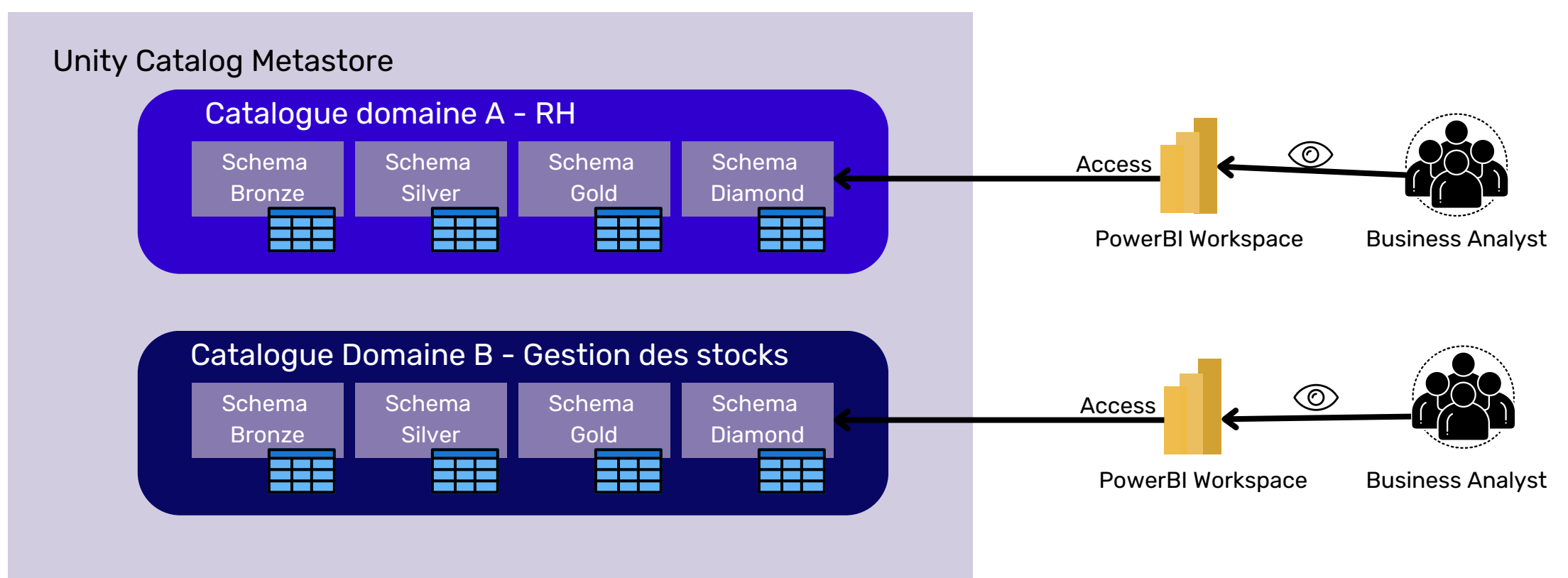
DE LA THÉORIE À LA PRATIQUE DU DATA MESH



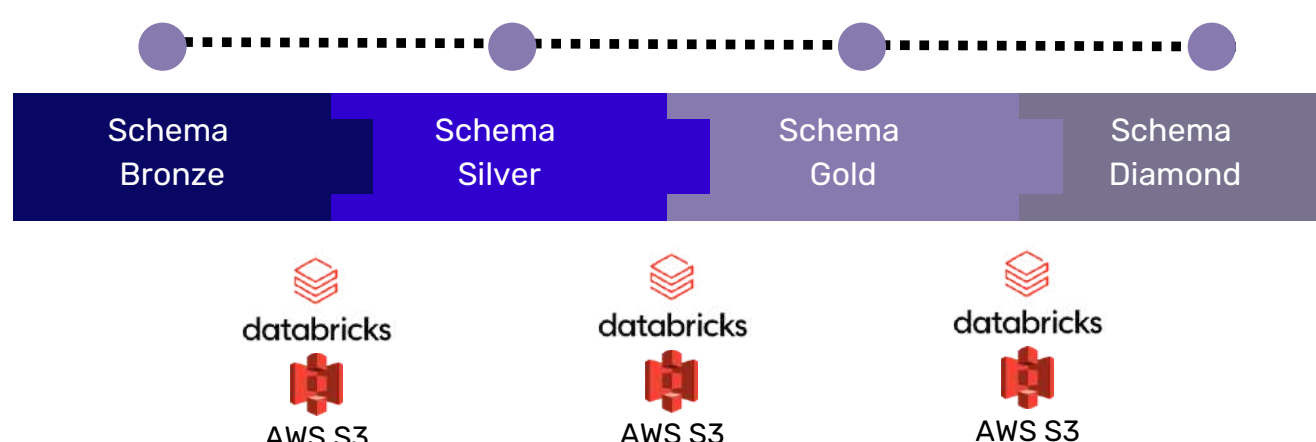
04

EXEMPLE D'ARCHITECTURE DATA MESH

Une architecture Data Mesh garantit la décentralisation des domaines de données dans un premier temps. Par la suite, il est question d'assurer l'interopérabilité entre les différentes équipes. L'architecture Data Mesh permet également l'accès en libre service à ces données via une plateforme. Et surtout, il se repose sur une gouvernance centralisée et qui s'assure de la sécurité des données.



Les différents niveaux de schemas sont enrichis successivement via des jobs databricks et les tables stockés sur S3 AWS. Les données peuvent être utilisées d'un domaine à l'autre suivant les règles RGPD



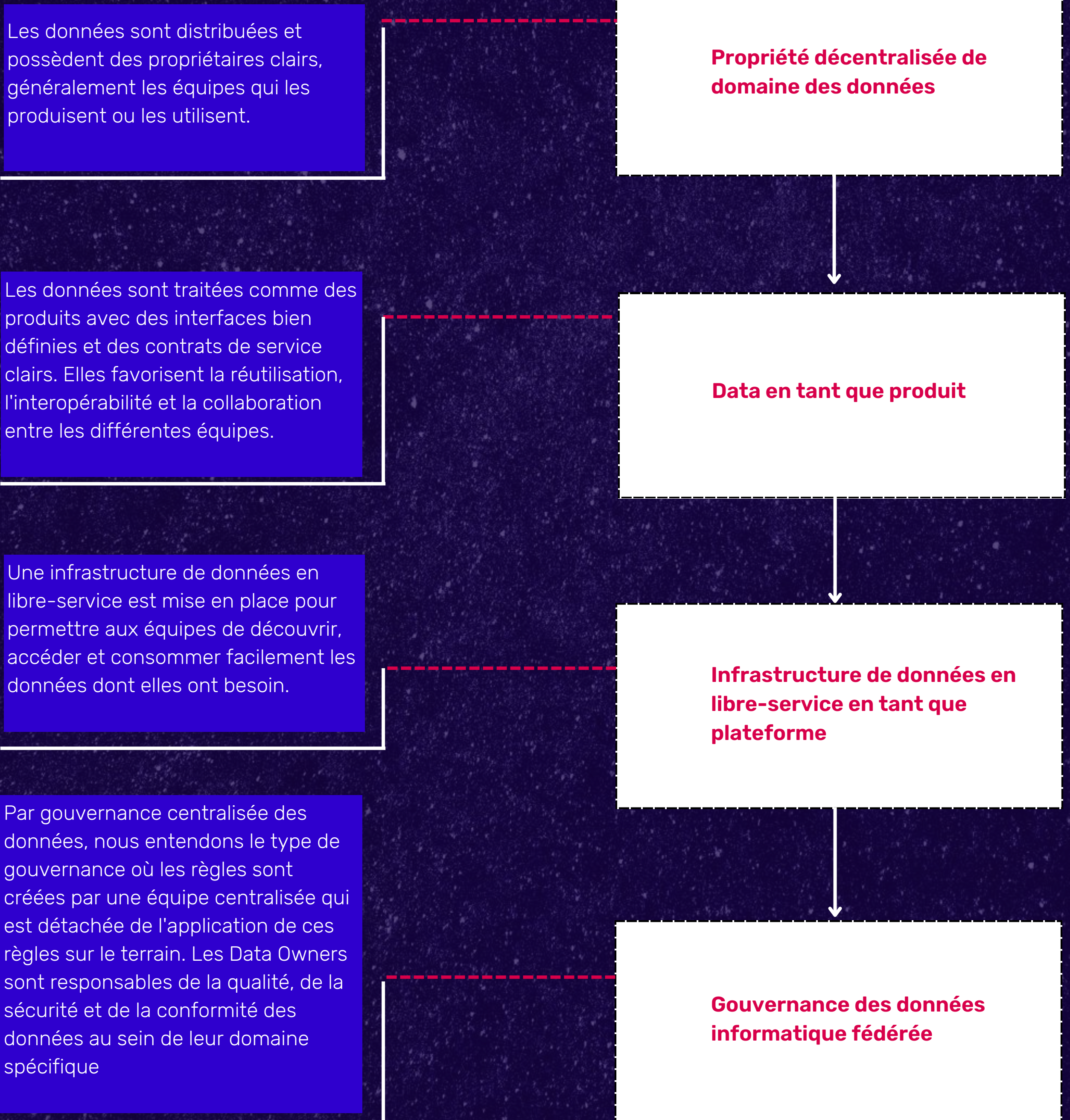
3. LES PILIERS DU DATA MESH

Dans la seconde partie de ce livre blanc, nous avons mis en pratique le data mesh en partant d'un cas d'usage aux étapes de mise en oeuvre.

La troisième partie de ce livre blanc se concentrera sur les quatre piliers fondamentaux qui sous-tendent le Data Mesh. Ces piliers sont les éléments sine qua non qui permettent aux organisations de mettre en oeuvre cette nouvelle architecture de manière efficace et de maximiser les avantages potentiels.

Ils constituent le cœur de la transformation vers une organisation véritablement data-driven.

Le Data mesh repose sur 4 piliers fondamentaux



4. LES TECHNOLOGIES ET OUTILS ASSOCIÉS DATA MESH

Nous avons vu comment le Data Mesh distribue les responsabilités, décentralise l'architecture technique, automatise les pipelines de données et favorise une culture de collaboration et de données.

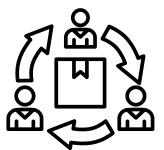
Cependant, pour que cette transformation soit efficace et durable, il faudrait disposer des bonnes technologies et des outils appropriés. C'est là que la quatrième partie de ce livre blanc intervient.

Dans cette section, nous allons détailler les technologies et les outils qui permettent de mettre en œuvre et de tirer pleinement parti du Data Mesh.

SOLUTIONS SELON LES ÉTAPES DE CONSTRUCTION DU DATA MESH

01 Technologies clés

Le Data Mesh n'est pas lié à des technologies ou outils spécifiques, mais plutôt à une approche organisationnelle et architecturale pour la gestion décentralisée des données. Cependant, certaines technologies et outils peuvent faciliter la mise en œuvre du maillage de données.



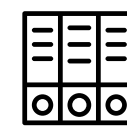
Plateformes de données distribuées

Comme Hadoop, Spark, Kafka, etc. Elles permettent de stocker et traiter des données de manière décentralisée par domaine.



APIs et services de données

Pour exposer les produits de données de manière découvrable, adressable et interopérable



Catalogage des données

Pour découvrir et comprendre les produits de données disponibles. Certains outils de catalogage mentionnés sont : Collibra, Apache Atlas, etc.



Outils d'ingestion et de transformation de données

Pour créer les pipelines de données par domaine, comme Apache NiFi, Talend, etc.



Outils de gouvernance et de sécurité des données

Pour appliquer les politiques de gouvernance de manière décentralisée, comme Apache Ranger, Collibra, Delta Sharing, etc.



Outils complémentaires

- Outils de développement et d'intégration continue: Pour créer et déployer les produits de données de manière agile.
- Outils d'observabilité et de monitoring: Pour surveiller la santé et la qualité des produits de données.
- Outils d'analyse de données: Pour exploiter les produits de données, comme Tableau, PowerBI, etc.

SOLUTIONS SELON LES ÉTAPES DE CONSTRUCTION DU DATA MESH

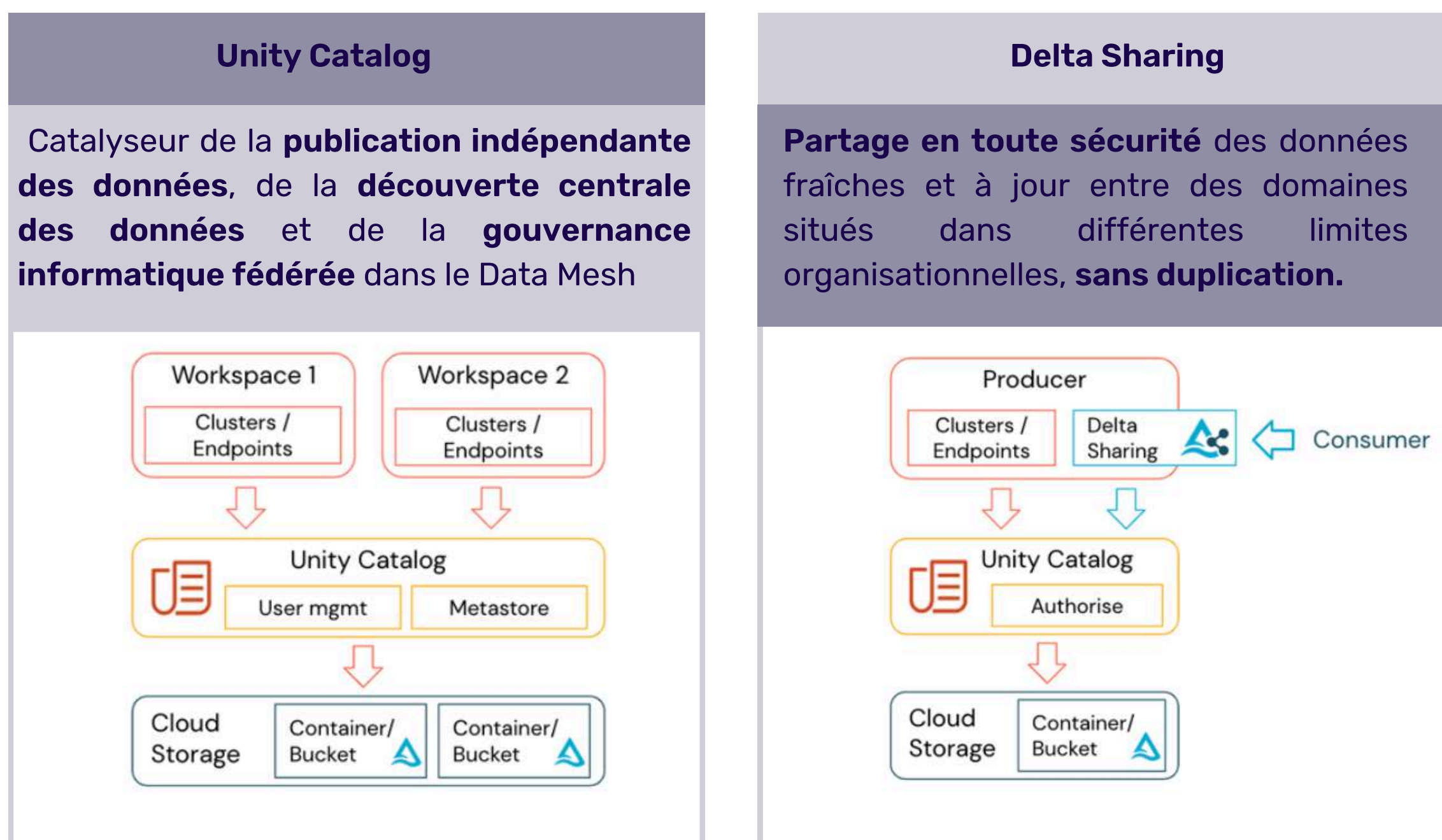
02 Databricks, une solution facilitatrice du Data Mesh

Databricks facilite la construction d'un Data Mesh grâce à l'exploitation d'une plateforme Databricks Lakehouse. Databricks Lakehouse est une architecture ouverte qui permet une grande souplesse dans l'organisation et la structuration des données.

Cette plateforme Cloud native combine le protocole **Unity Catalog** pour unifier la gestion des données et assurer leur scalabilité, **Delta Sharing** pour le partage en toute sécurité, et également un **Workspace Management** pour gérer les produits de données à l'aide d'une infrastructure commune indépendante du domaine, en libre-service.

Pour mettre en œuvre un Data Mesh avec Databricks, on utilise des espaces de travail dédiés et des workflows automatisés via **Databricks Workflows et Databricks Terraform**

Schéma 9 : Fonctionnalités de Databricks Lakehouse pour activer un maillage de données



Source : [Databricks.com](https://databricks.com)

Un Data Mesh clarifie les objectifs et les besoins de l'entreprise pour maximiser la productivité et la valeur des données. Databricks Lakehouse propose une plateforme ouverte et évolutive, conçue pour répondre à ces exigences avec une interopérabilité, une efficacité et une simplicité accrues.



databricks

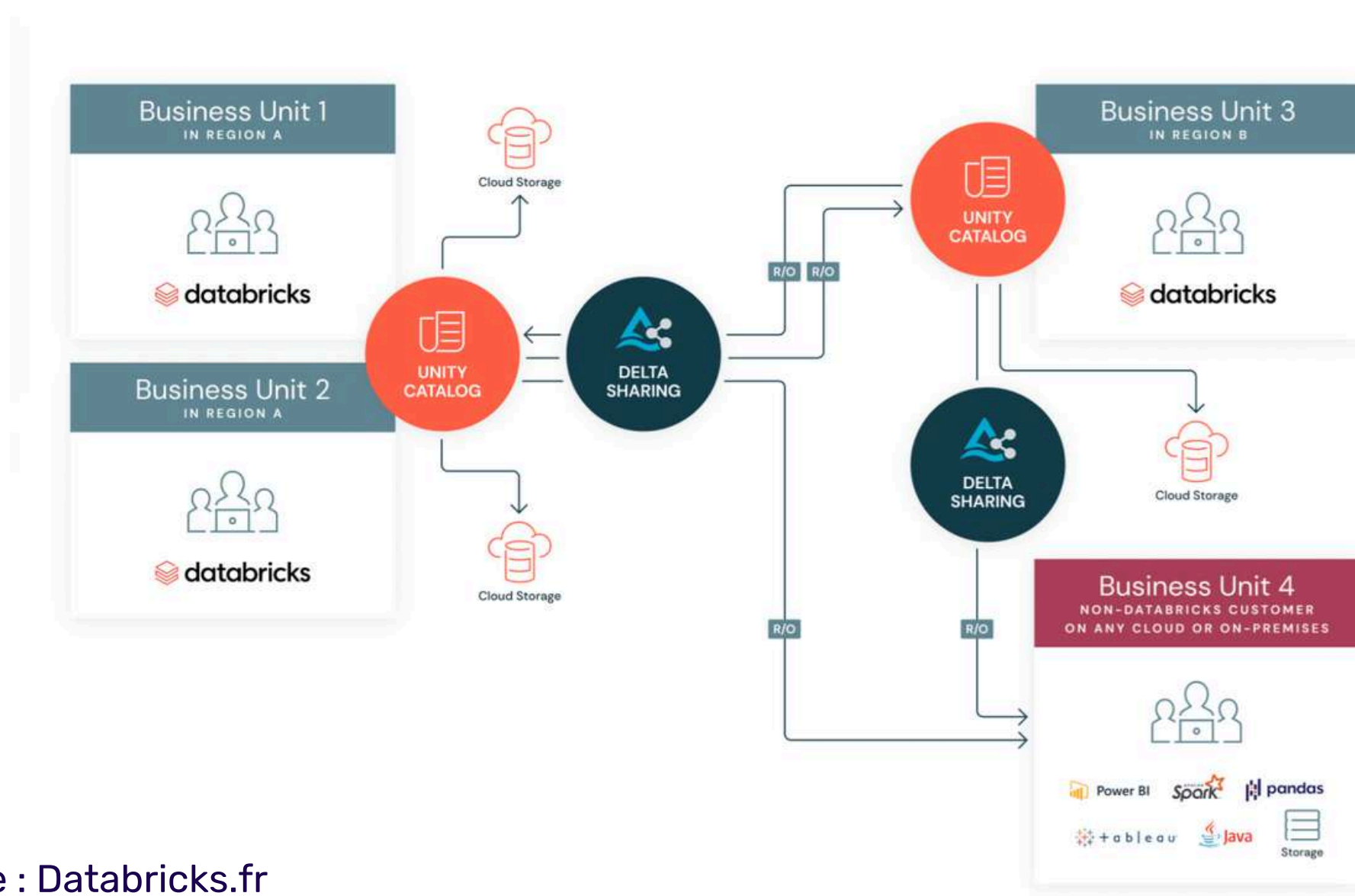
SOLUTIONS SELON LES ÉTAPES DE CONSTRUCTION DU DATA MESH

Databricks, une solution facilitatrice du Data Mesh

La combinaison de Unity Catalog, Delta Sharing et Databricks Marketplace permet de répondre aux exigences de gouvernance fédérée des données dans un Data Mesh.



Exemple de maillage de données avec Delta Sharing pour partager en toute sécurité des données avec des unités commerciales et des filiales à travers des cloud ou des régions sans copier ou répliquer les données.



Source : Databricks.fr

Les technologies Databricks sont des facilitateurs clés pour mettre en œuvre avec succès une architecture Data Mesh décentralisée et fédérée pour la gestion des données analytiques



databricks

5. EVOLUTION DU DATA MESH

Dans cette section, nous explorerons les tendances clés liées au Data Mesh.

Nous examinerons également comment l'évolution des pratiques de collaboration et de culture au sein des organisations renforce l'adoption et l'efficacité du maillage de données.

Les entreprises devront adapter leurs structures et leurs mentalités pour tirer pleinement parti de cette approche innovante.

Tendances émergentes dans le Data Mesh

Le data mesh est une tendance émergente dans la gestion des données d'entreprise qui gagne en popularité. Nous avons identifié quelques principales tendances liées à cette approche :

● Adoption croissante du data mesh

De plus en plus d'entreprises adoptent le data mesh pour faire face à la complexité et à la diversité croissantes des données. Le Data Mesh décentralise, démocratise et traite les données comme un produit dans un modèle fédéré, permettant un accès facilité aux données.

● Puissante découverte de données à grande échelle (data discoverability)

Avec l'adoption du Data Mesh, les entreprises se tournent vers des catalogues de produits de données organisés et facilement consommables plutôt que des catalogues de données techniques traditionnels. Cela permet aux utilisateurs d'accéder directement aux données de manière sécurisée.

● Conception axée sur les domaines métiers

Le data mesh applique les principes de la conception axée sur les domaines, alignant les données sur les besoins spécifiques des différents domaines métier. Il améliore l'agilité, la pertinence et la précision des données.

● Collaboration et responsabilisation organisationnelles

Le data mesh encourage la collaboration entre les domaines métier et la responsabilisation des équipes pour la qualité des données. Cela favorise une culture de responsabilité partagée.

CONCLUSION

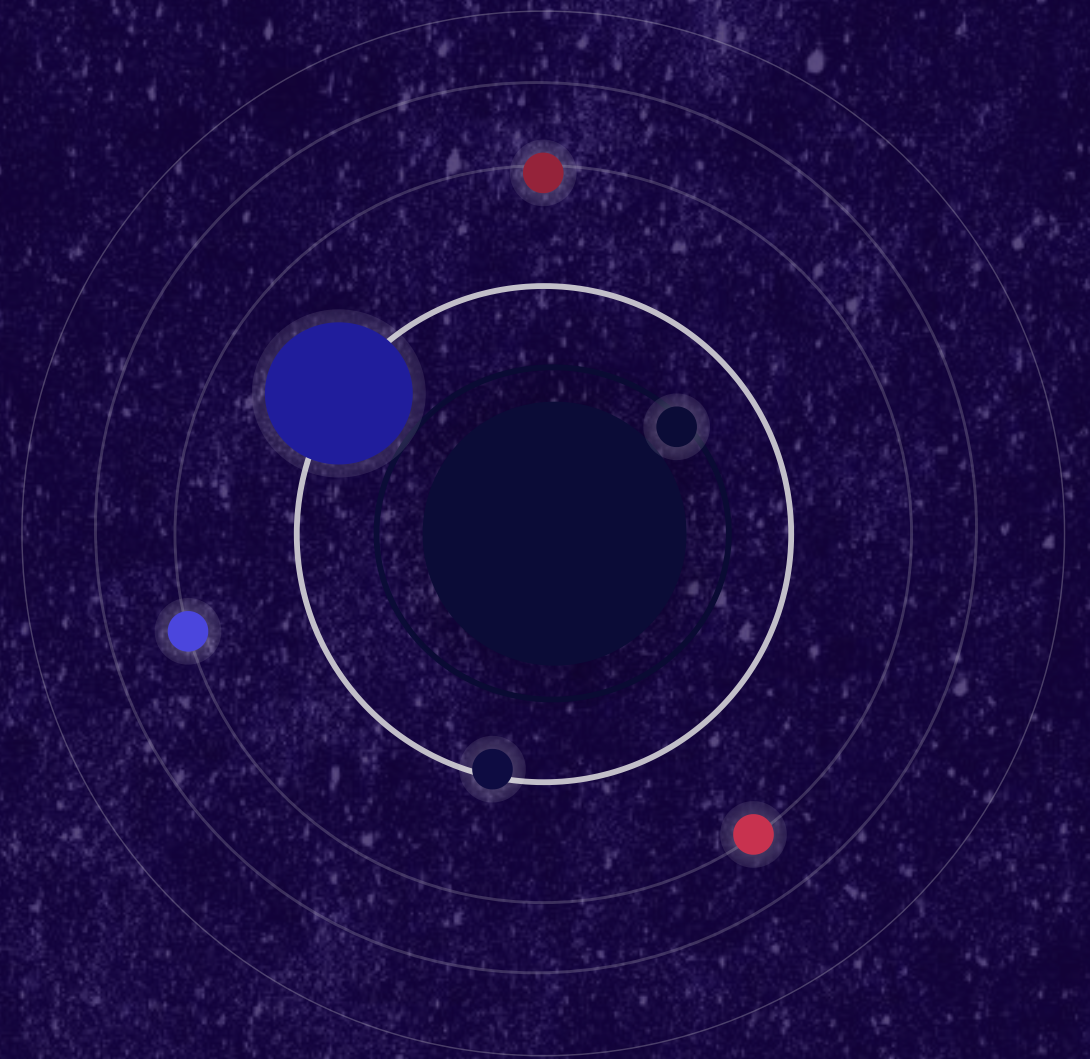
Le Data Mesh n'est pas seulement une innovation technologique, mais aussi un changement de paradigme dans la manière dont nous concevons et utilisons les données.

En adoptant cette approche, les organisations peuvent non seulement surmonter les limitations des modèles traditionnels, mais aussi ouvrir la voie à une exploitation plus efficace et stratégique de leurs données. Nous espérons que ce livre blanc vous a fourni des insights précieux et des directives pratiques pour envisager et réussir l'adoption du maillage de données dans votre propre contexte organisationnel.

4. POUR ALLER PLUS LOIN

Discutons de votre vision de devenir Data driven et de mettre en place une plateforme data intégrant une architecture Data Mesh.

Inventiv IT vous accompagne dans votre programme de mise en oeuvre :



Audit et Conseil

Nous réalisons pour vous un audit de maturité de votre gouvernance actuelle.

Stratégie et Roadmap

Nous vous proposons une stratégie accompagnée d'une roadmap détaillée.

Solutions /outils selon vos usages

Nous vous aidons à définir les solutions adéquates, et nous vous accompagnons dans le déploiement

Inventiv IT est partenaire certifié de



EN SAVOIR PLUS

RÉDACTEURS ET CONTRIBUTEURS



Sami Ouassaidi
Architecte Data



Hugues Journeau
Solutions Architect
Databricks



Sandra Randria
Chief Marketing Officer



Christel Bitemo
CP - Marketing Digital



<https://inventiv-it.fr>

