

Une architecture de données moderne sous des contraintes de sécurités élevées

Par Eric Gagnaux, Directeur DATA chez keyIT

4 avril 2024

Agenda

- 🚩 keyIT en 4 slides
- 🚩 Introduction à la donnée
- 🚩 Les fondamentaux
- 🚩 Déclenchement d'un projet Data
- 🚩 Collecte, préparation et stockage
- 🚩 Plateforme moderne de données
- 🚩 La partie sécurité

Présentation de keyIT

keyIT

Des bases solides

Management constitutionnel

Notre organisation au sein de l'entreprise

Marly - Vernier

Deux points de rencontre situés à Fribourg et Genève
Prochainement: Valais

Les données La sécurité

Notre raison d'être

01.01.2021

Création de la
Société Anonyme

CHF 250k

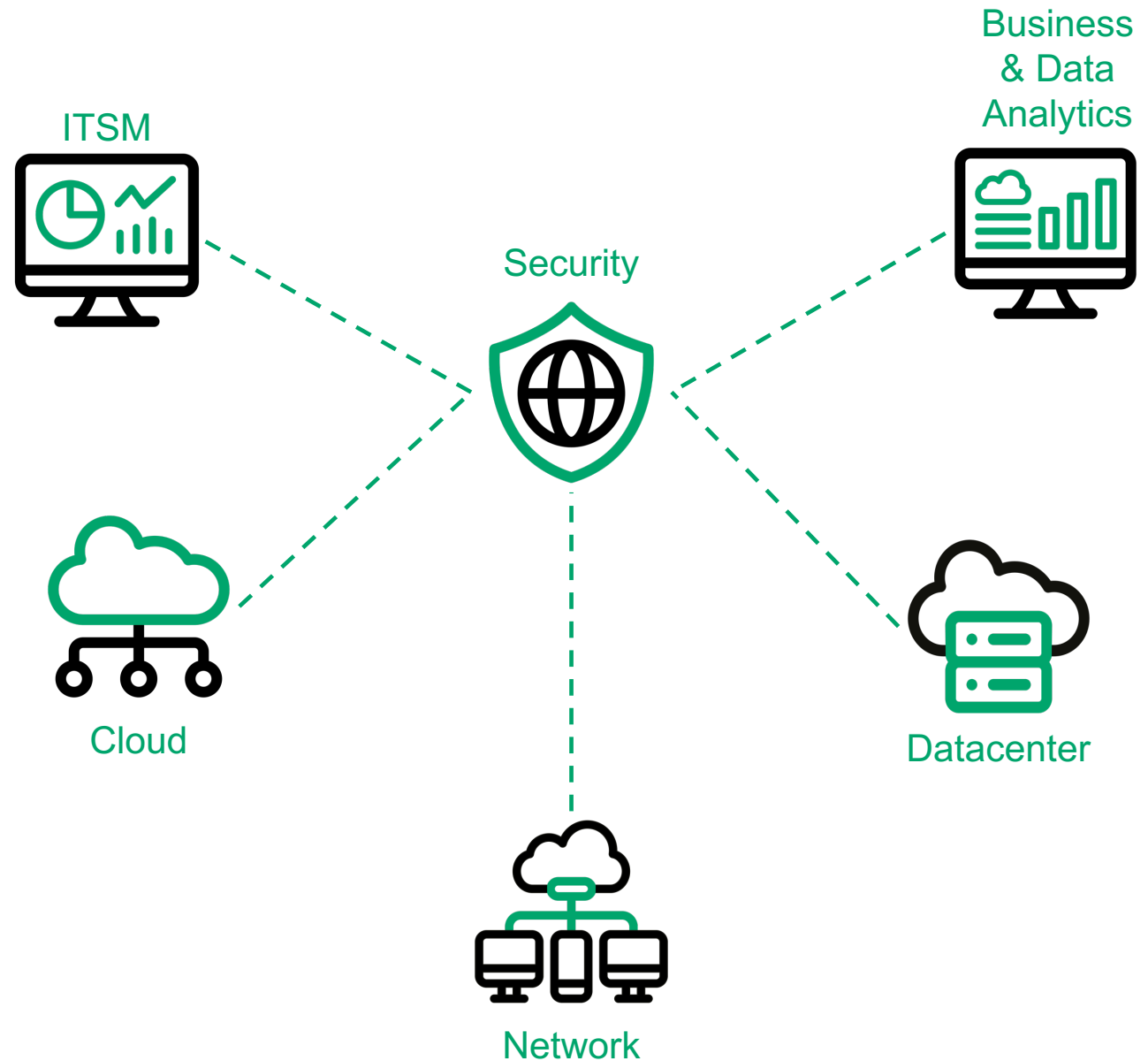
de capital action

**Nbr de
pers.**

23 au 1^{er} avril 2024

En 1 coup d'oeil

- 🚩 Sécurité
- 🚩 Réseau
- 🚩 Cloud
- 🚩 DataCenter
- 🚩 Business & Data Analytics
- 🚩 ITSM



Business & Data Analytics chez keyIT

à l'avant-garde d'un avenir
fondé sur les données



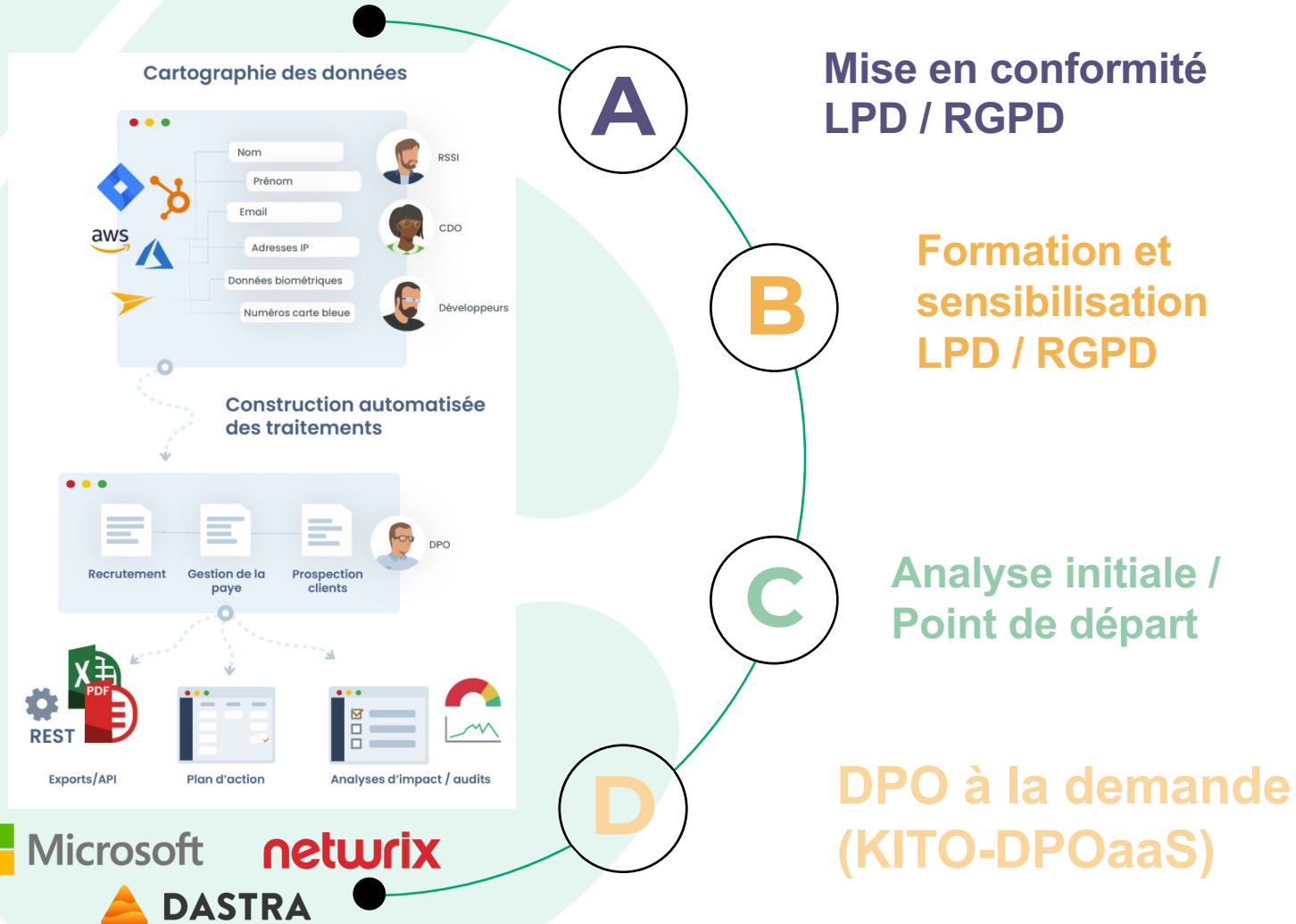
Chez keyIT, **nous utilisons des analyses de données avancées** pour vous aider à rester en tête dans votre domaine.

L'analyse des données est essentielle à l'innovation dans toutes les industries, aidant les entreprises à ne pas se contenter de suivre, mais à prendre les devants.

Nous nous consacrons à l'avancement des marchés divers, considérant les défis comme des opportunités de croissance.



DPO as a Service

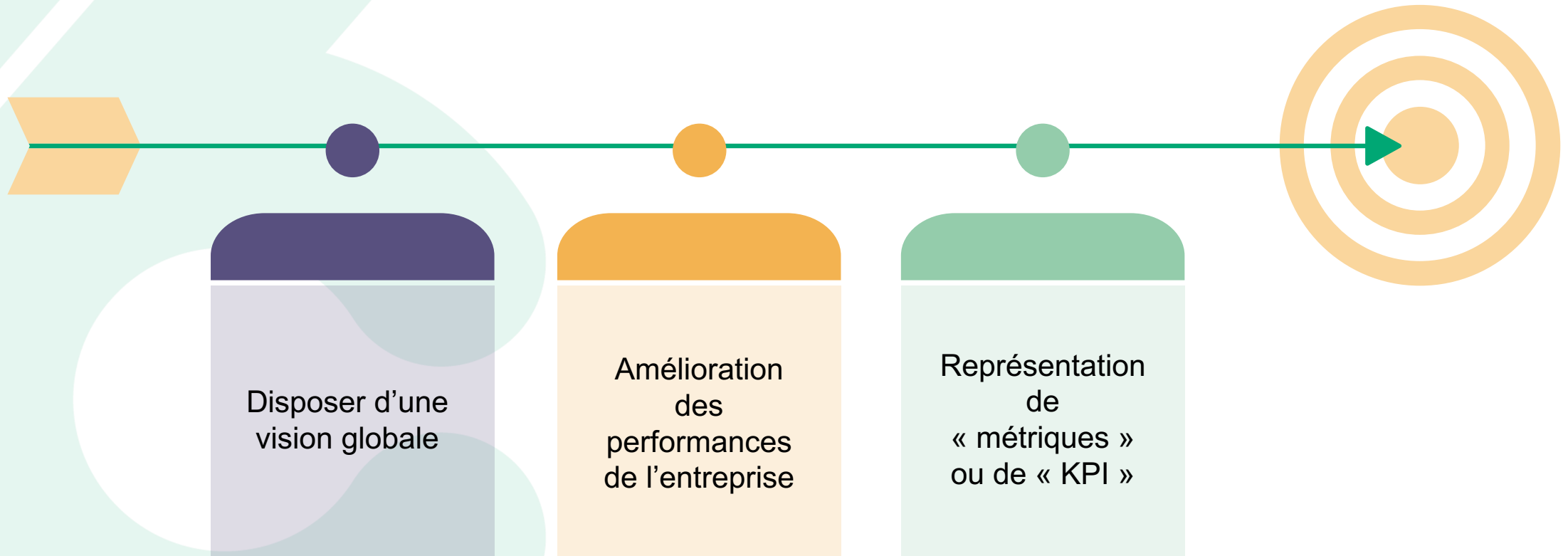


- ✓ Cartographie des données
- ✓ Transferts à l'étranger
- ✓ Registre des traitements
- ✓ Sous-traitants
- ✓ Gestion des incidents et des demandes

Introduction à la donnée

keyIT

Pourquoi utiliser ses données?



Cas d'usage

EDUCATION:

Augmentation des revenus et meilleures prédictions des effectifs pour une école de renommée internationale

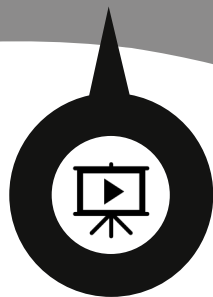


INDUSTRIE:

Prévention de pannes pour une entreprise de location d'engins de chantiers

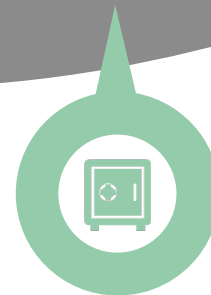


ASSURANCES: Aide à la prise de décision stratégique pour l'évolution des technologies pour les 5 prochaines années



MEDIA: Optimisation des données clients pour une meilleure orientation du marketing et des achats sur l'acquisition de nouveaux produits

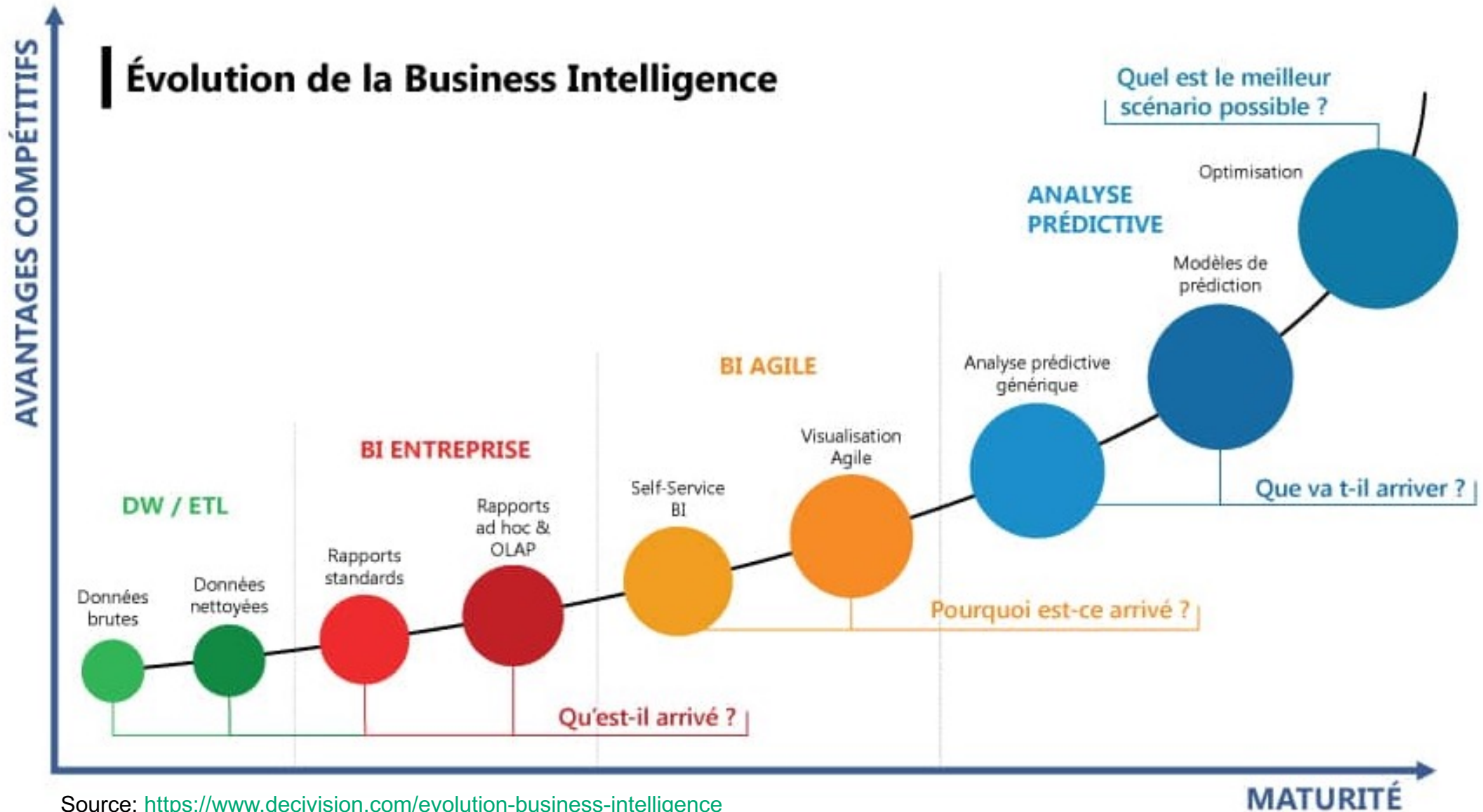
BANQUE: Détection de fraudes sur des outils financiers



SANTÉ: Assistance proactive aux soignantes dans une clinique spécialisée pour enfants sous respirateur artificiel



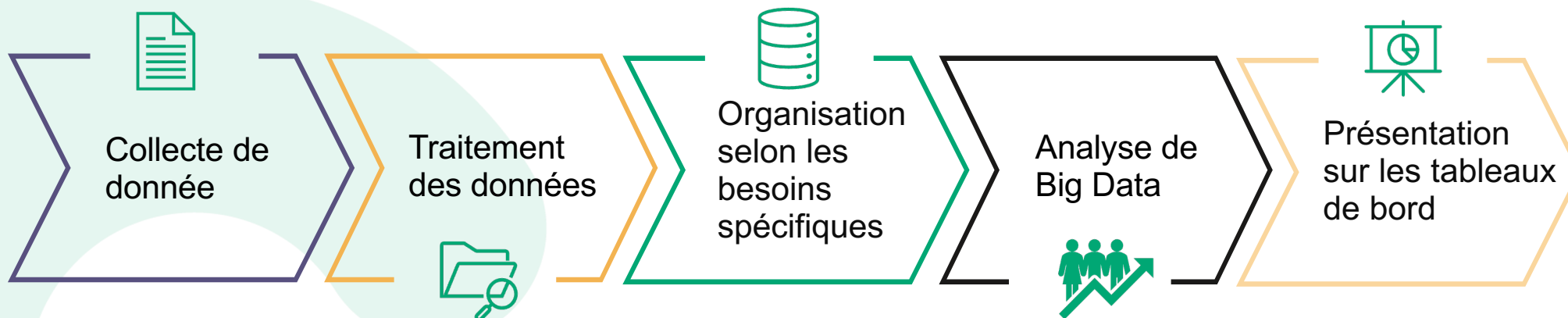
Evolution de la donnée



Les fondamentaux

keyIT

Le processus du début jusqu'au tableau de bord



Tout ce que l'entreprise connaît et collecte est versé dans le **datalake**, brut et non filtré.

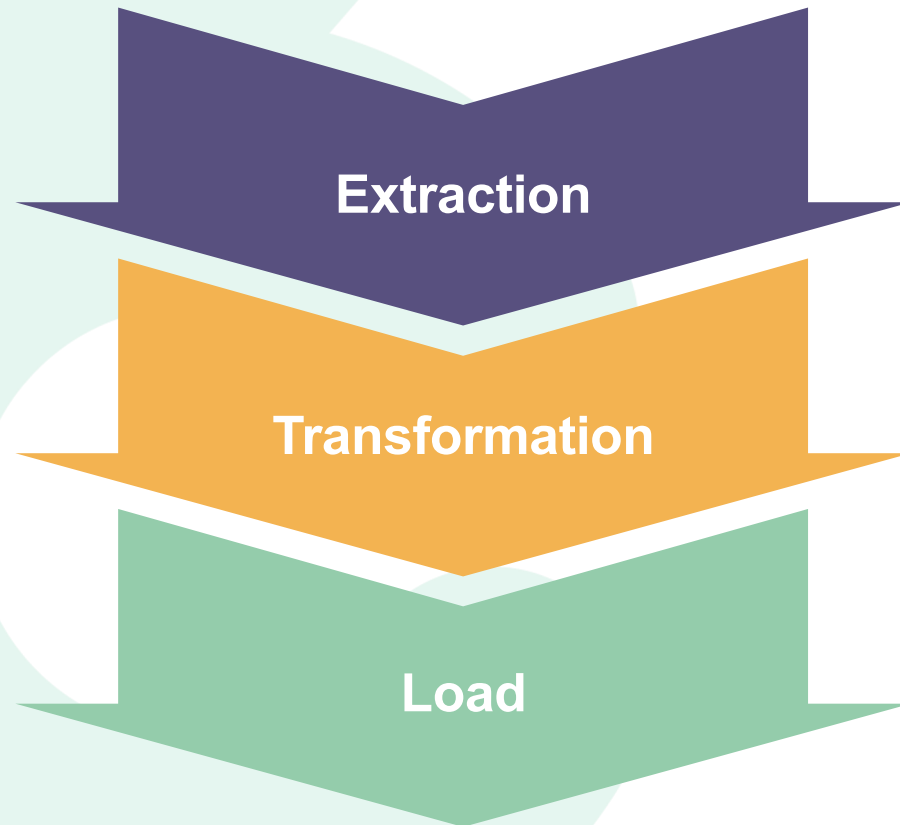
À l'aide de **l'ETL**, les données sont **extraites, transformées et chargées** dans l'entrepôt de données où elles sont organisées et préparées pour l'analyse.

Des parties de ces données sont ensuite organisées en **datamarts** pour des équipes ou départements spécifiques.

Des outils et techniques spéciaux sont utilisés pour filtrer **de grands volumes de données** afin de trouver des informations utiles.

Enfin, les informations et les données sont présentées sur **des tableaux de bord** de manière facile à comprendre

ETL / ELT (Extract-Transform-Load)



Les données sont extraites de différentes sources de données telles que des fichiers, des bases de données ou des API

Les données sont transformées pour correspondre aux exigences du modèle de données de l'entrepôt de données

Les données sont chargées dans l'entrepôt de données à l'aide d'un processus automatisé

Entrepôt de données Data Warehouse (DWH)



🌿 Base de données centralisée :

Stockage et organisation de données provenant de diverses sources

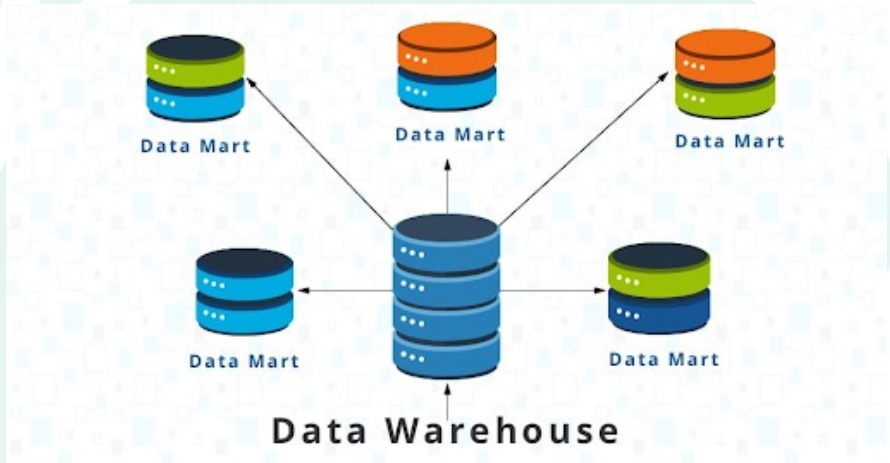
🌿 **Objectif :**

- Accès facile et rapide aux données pour l'analyse et la prise de décision
- Optimisé pour des requêtes complexes
- Données sécurisées

🌿 **Composition :**

- Datamart

Datamart



🌿 C'est un sous ensemble du DWH, généralement organisé autour d'un sujet métier (RH, Comptabilité etc.)

🌿 **Objectif :**

- Accès sécurisé
- Périmètre réduit donc plus facile à appréhender
- Organisé et dédié à un besoin

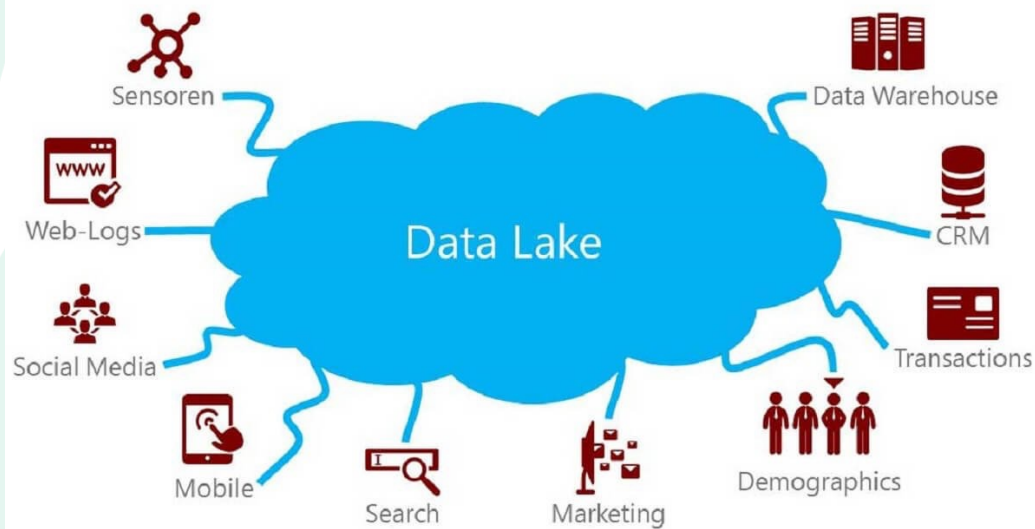
🌿 **Composition :**

- Dimensions
- Faits

BigData

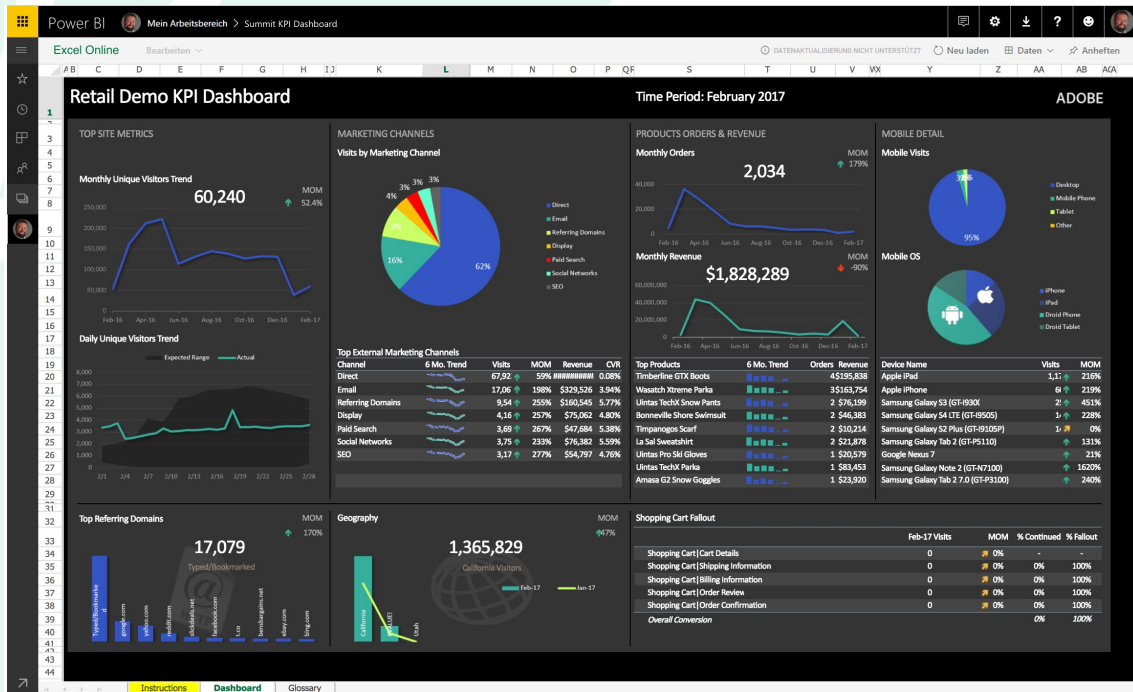
- 🌿 **Un volume de données important :**
 - On parle de Pétaoctet de données (1 000 To)
- 🌿 **Une multitude de source de données :**
 - Réseaux sociaux
 - IoT
 - log
- 🌿 **Une croissance continue et importante :**
 - Ajout de données à la milliseconde
- 🌿 **Nécessite des technologies spécifiques :**
 - Hadoop, Spark

DataLake



- 🌿 Lieu de stockage centralisé et flexible dans le Cloud
- 🌿 Tout type de données structurées, semi structurées et non structurées
- 🌿 Capacité de stockage illimitée
- 🌿 Sécurisé
- 🌿 Accessible

Reporting & Tableau de bord



Met en forme les données, généralement sous forme graphique afin de :

- Aider les prises de décision
- Piloter une activité via des indicateurs visuels
- Alerter
- Partager des analyses

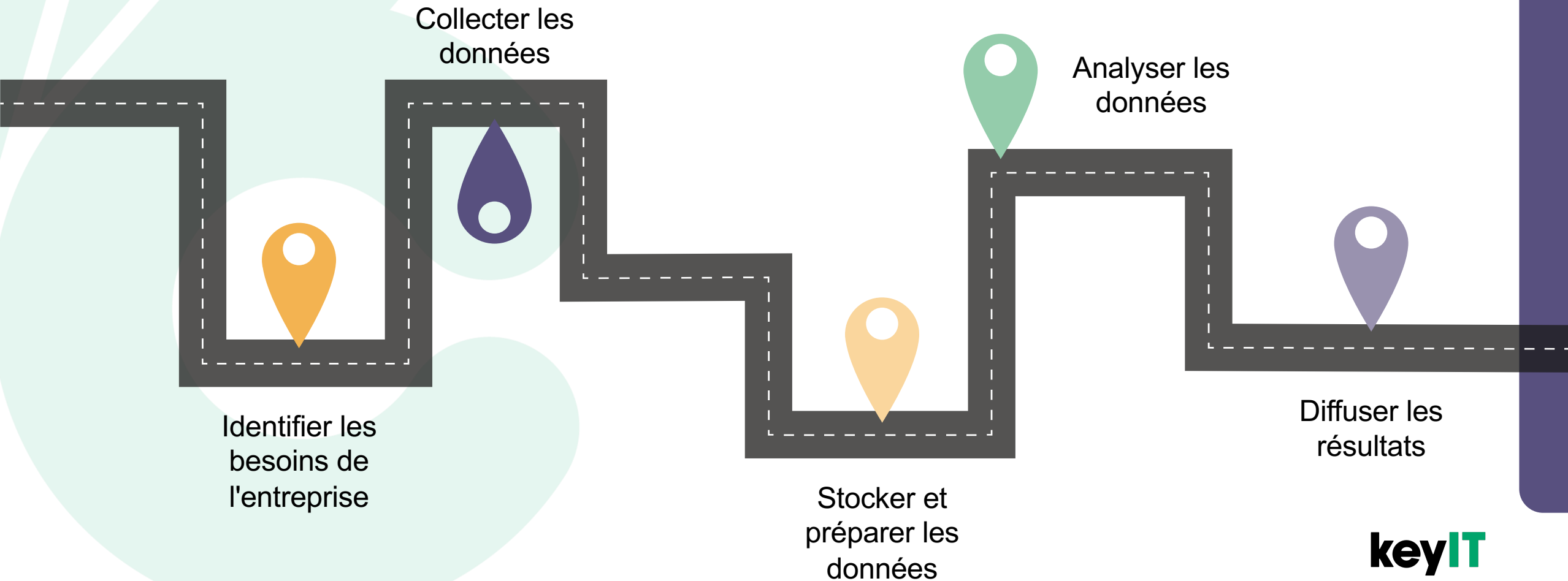
Déclenchement d'un projet Data

keyIT

Pourquoi créer un projet BI

- 🌿 Améliorer la prise de décision
- 🌿 Gagner en efficacité
- 🌿 Optimiser la performance
- 🌿 Anticiper les tendances

Comment créer un projet BI



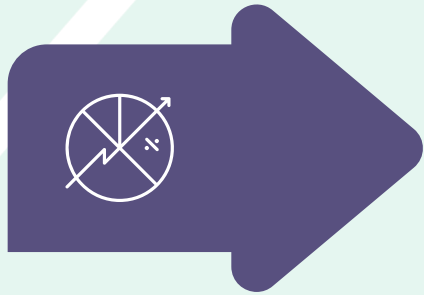
Les profils des différents intervenants

Rôles	Appartenance	Définition	Périmètre d'action
Data Engineer	Prestataire /Client	Responsable de la conception, de la construction et de la maintenance des infrastructures de données	ETL/ELT, Ingestion des données, Spark,...
Data Scientist	Prestataire /Client	Responsable de l'analyse des données pour en extraire de la valeur et faire des prédictions	Machine Learning, Intelligence Artificielle
Business Analyst	Prestataire /Client	Personne du métier qui analyse les données pour identifier les problèmes et les opportunités d'amélioration	Qualification des besoins et documentation
Data Protection Officer	Client	Responsable de la protection des données et la conformité avec les lois, et de la sensibilisation des employés	Suivi et validation de la conformité des données
Data Steward	Client	Responsable de la gestion des données au sein d'un département. Surveillant aussi la qualité des données	Suivi et responsabilité des données

**Collecte,
préparation et
stockage**

keyIT

La collecte des données



Identifier les sources de données

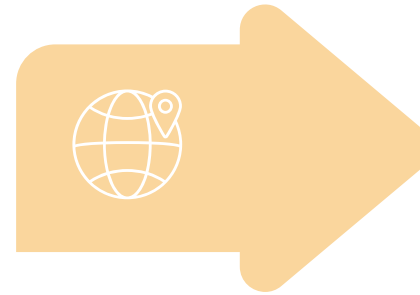


Planifier les méthodes de collecte de leurs fréquences

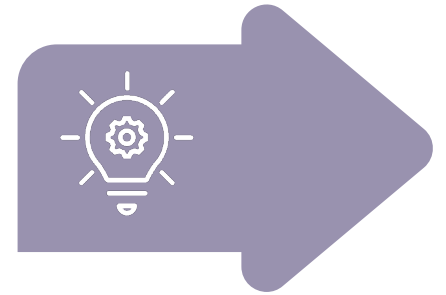


Collecter les données:

- Mettre en place une infrastructure
- Assurer la qualité des données, nettoyer et normaliser les données



Stocker et protéger les données



Analyser les données

Sources de données



Relational Databases

(structured)



Unstructured

images, video, audio, free text
(no structure)



Semi-Structured

csv, logs, json, xml
(loosely-typed)



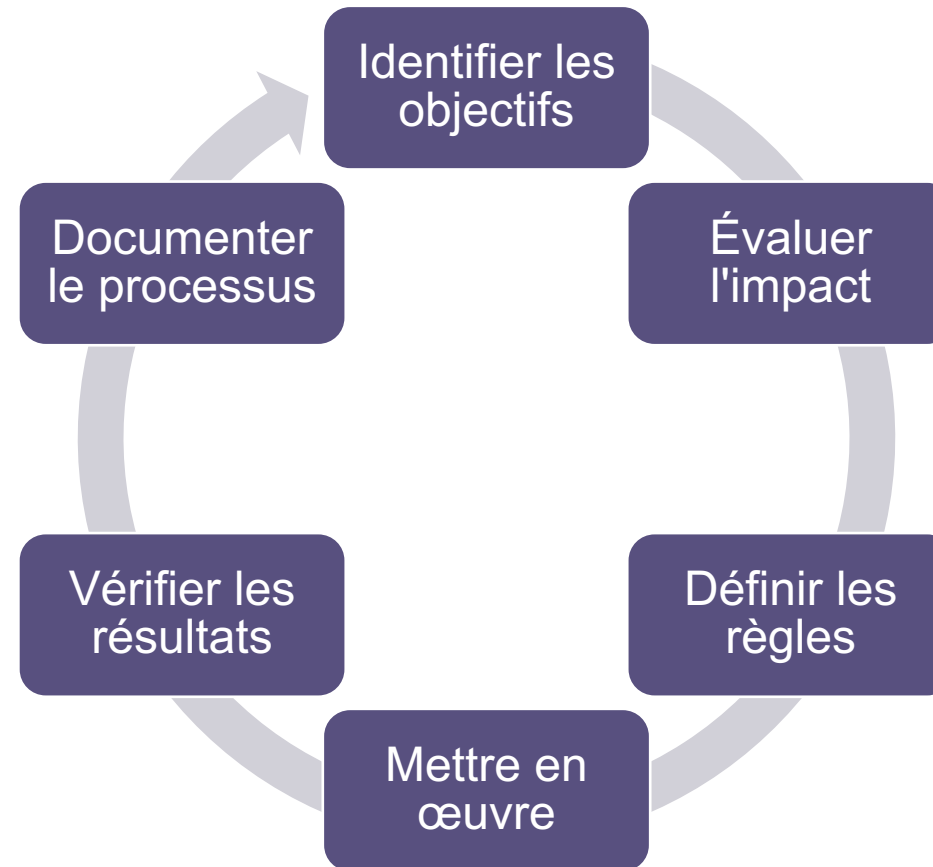
Big Data Streams



IoT Devices

Streaming

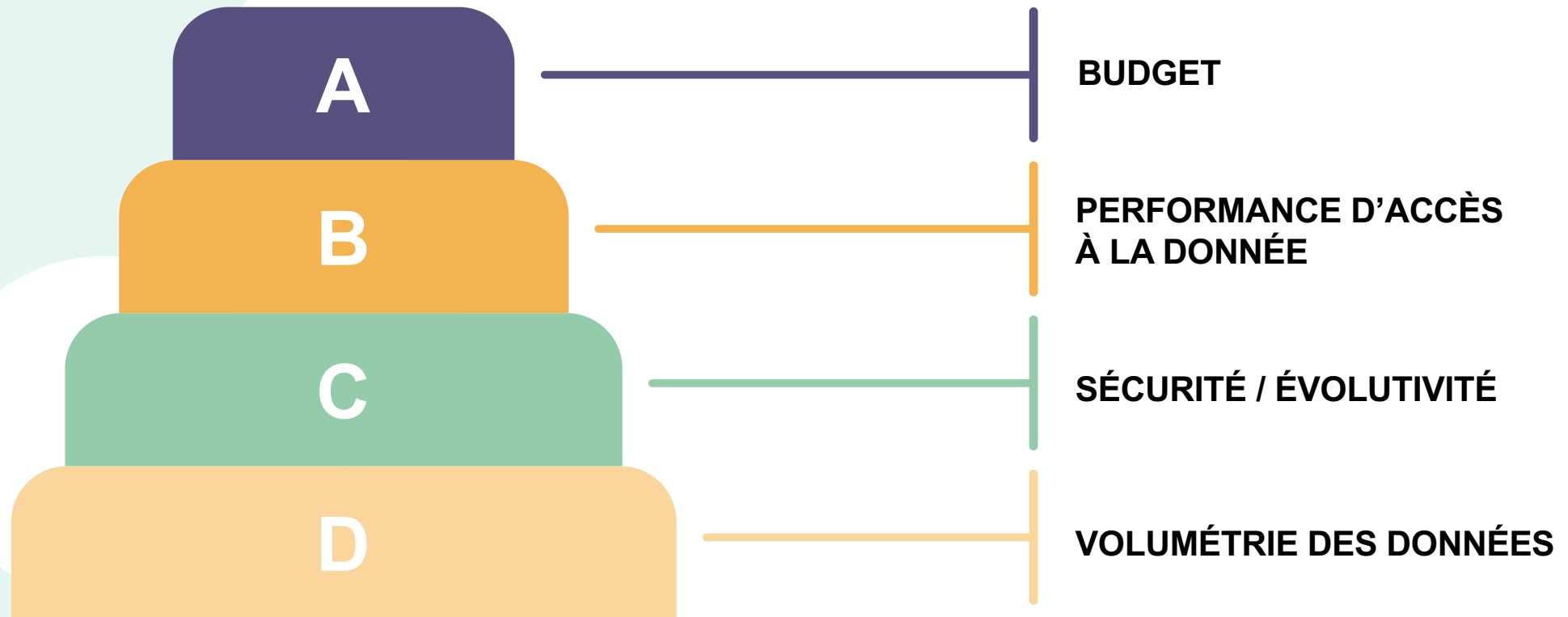
Nettoyage et transformation



Cloud vs On Premise

	Cloud	On Premise
Évolutivité	+++	---
Flexibilité	+++	--
Coût	++	--
Maintenance	+++	---
Performance	+++	+/-
Contrôle	---	+++
Sécurité	+++	+

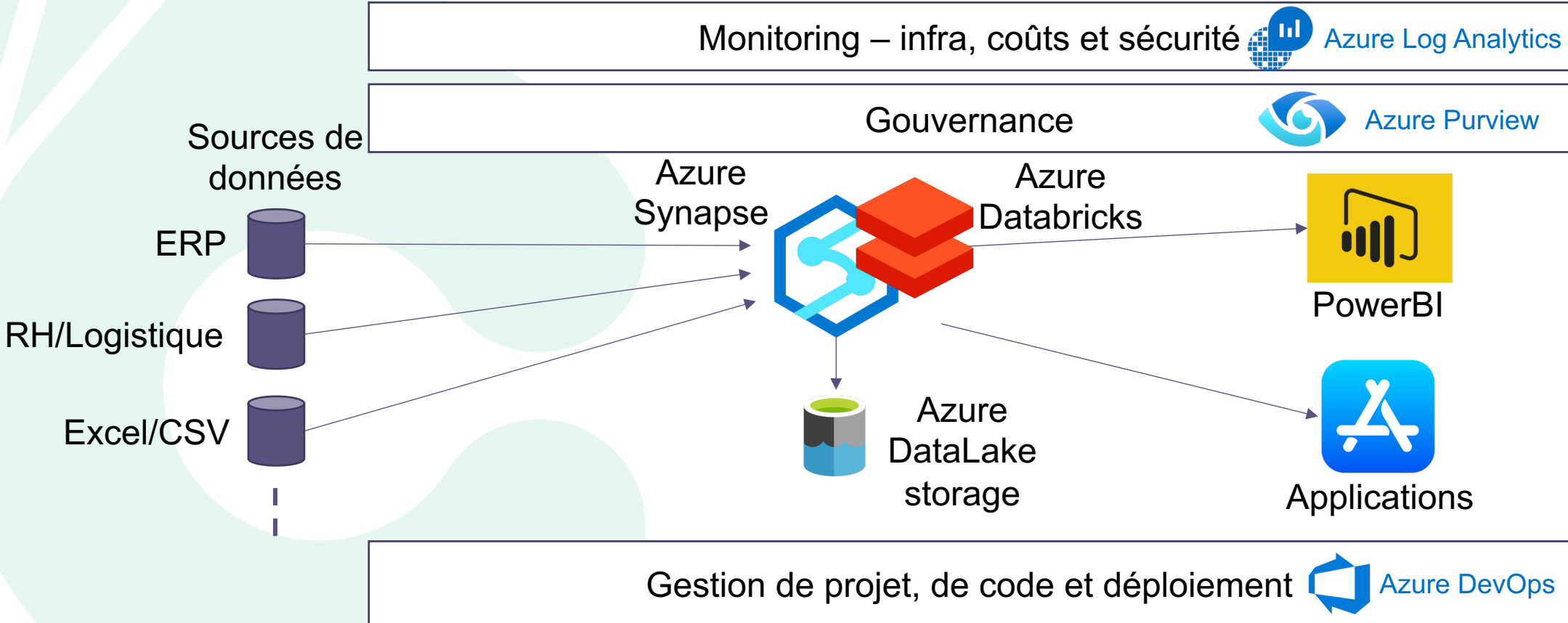
Comment choisir mon stockage ?



**Plateforme moderne
de données**

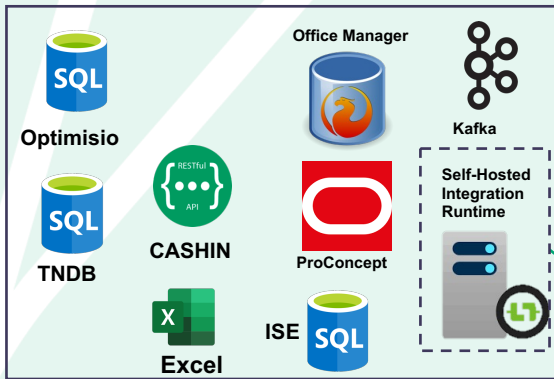
keyIT

Schéma bloc d'une plateforme moderne BI

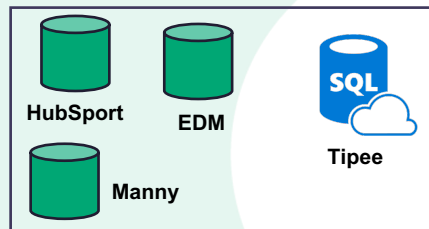


LAKEHOUSE- Architecture avec ADF + Databricks

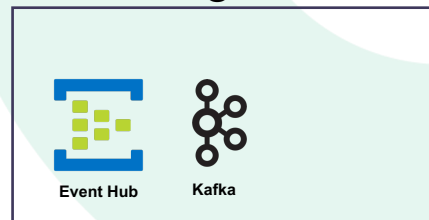
On-prem
Source Systems



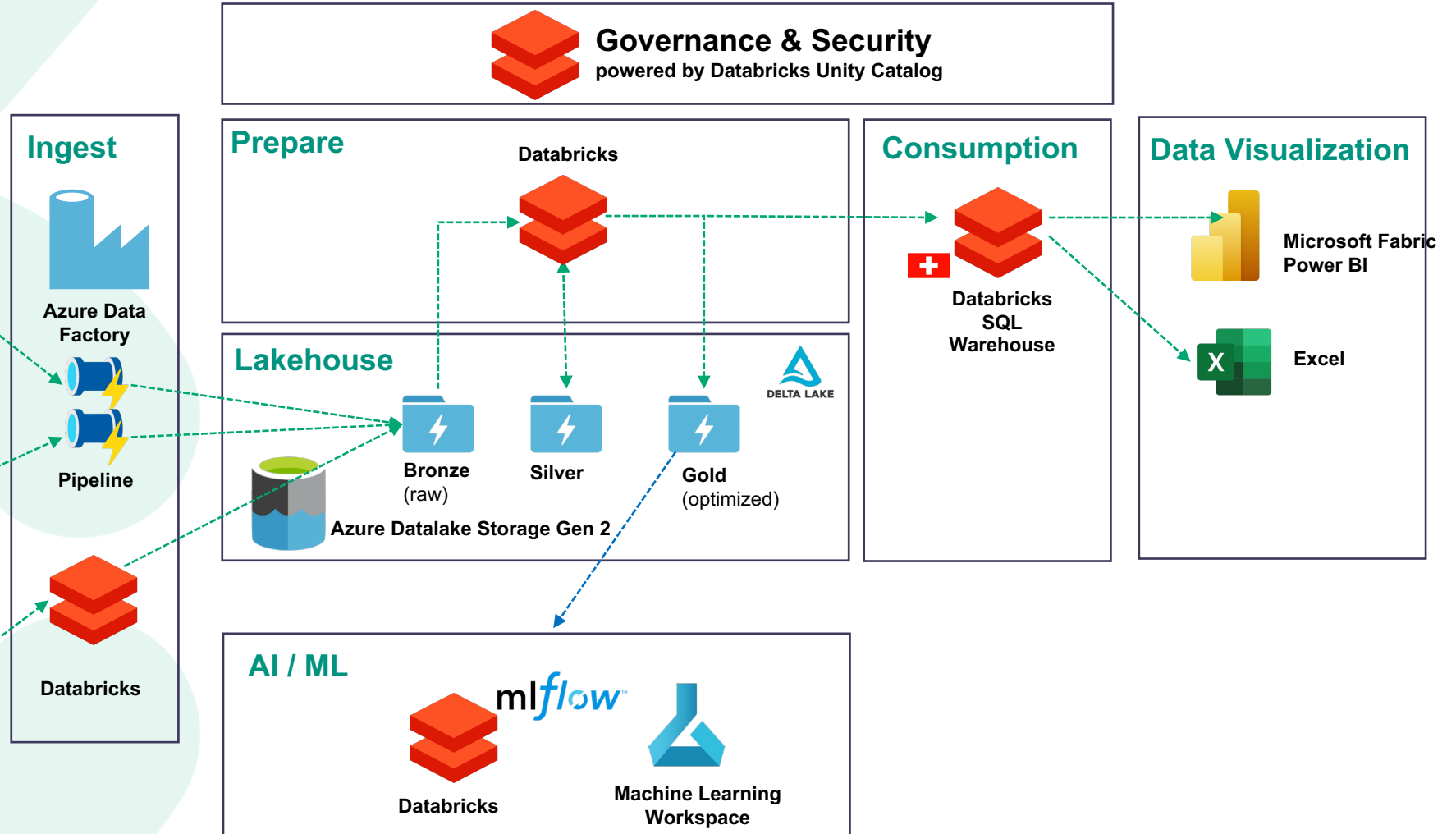
Cloud Data



Streaming Data



Real Time Ingestion



La sécurité

keyIT

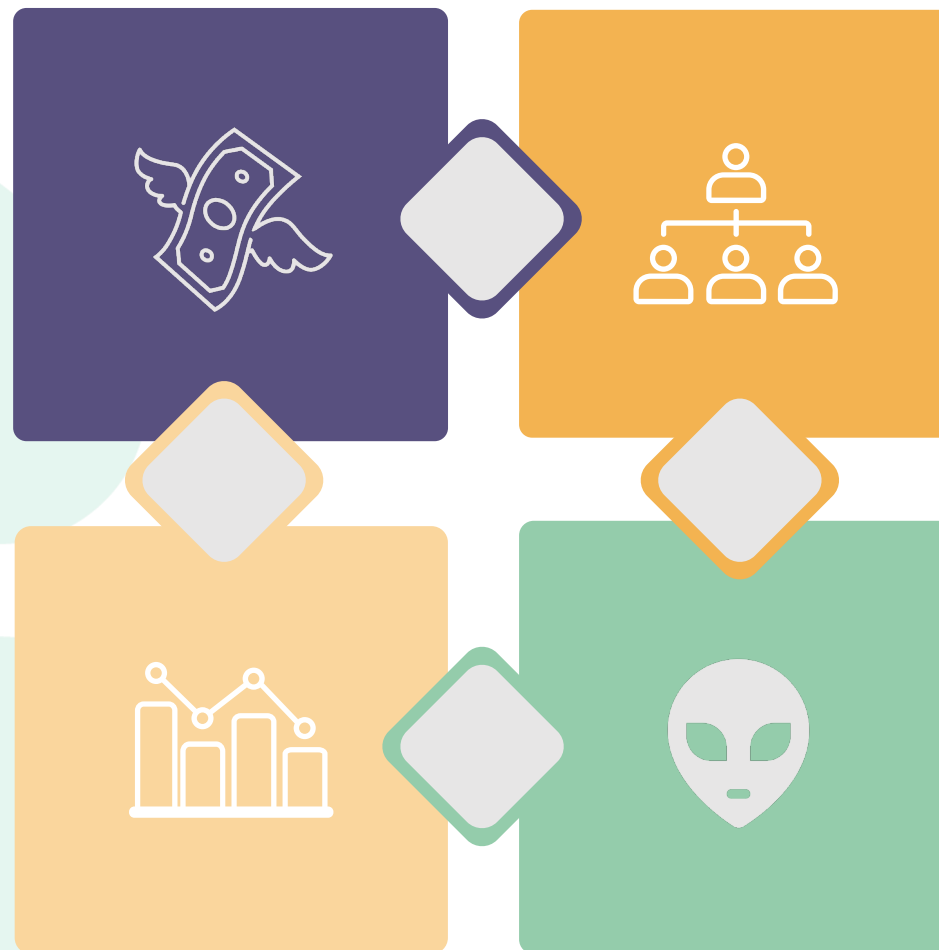
Naviguer à travers les cybermenaces

4,45 millions de dollars

Le coût moyen mondial d'une violation de données en 2023, soit une augmentation de 15 % en trois ans.

1 sur 3

Nombre de violations identifiées sur les outils développés dans l'organisation



Les données personnelles des clients

sont les données les plus coûteuses et les plus souvent compromises.

Le phishing et les informations d'identification volées ou compromises

Ils ont été à l'origine de respectivement 16 % et 15 % des violations

Source: Cost of a Data Breach Report 2023, IBM

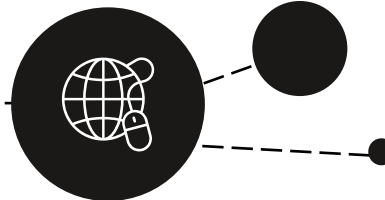
keyIT

Les éléments de sécurité à prendre en compte

Hébergement des Données et Infrastructure



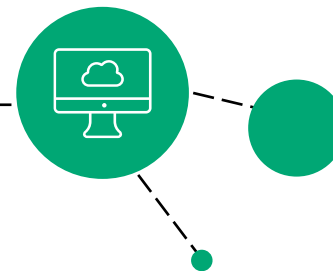
Configuration du Réseau et du Stockage



Gestion des Accès et Gouvernance



Contrôle d'Accès et Protection des Données



Conformité Réglementaire



Les éléments de sécurité qui peuvent (devraient) être appliqués



Hébergement de l'infrastructure entièrement **en Suisse**.



Ingestion sécurisée des données **onPremise** via une Gateway (avec ou sans VPN)

Gestion des accès RBAC basée sur **Microsoft Entra ID**

Gouvernance des données assurée par Unity Catalog



Possibilité d'infrastructure réseau **100% privée** (VPN + Private EndPoints)



Redondance **de stockage des données** Zone-redundant storage (ZRS)



Contrôle d'accès au niveau Catalogue, Schéma, Table, Views ou Dossier/Fichier

Conformité RGPD, ISO27001, 27017, 27018, SOC1,2,3, CCPA, HIPAA, PCI-DSS, HITRUST, IRAP

Questions ?

Associez-vous à keyIT

Scannez le code



Mes coordonnées: Eric Gagnaux, ega@keyit.ch, 079 350 57 59

Spécialistes en conversion des données en insights actionnables, nous dynamisons l'efficacité, l'innovation et la durabilité au sein des divers industries.

Appel à l'action :

Connectez-vous avec nous pour explorer comment nos solutions d'analyse de données sur mesure peuvent propulser votre entreprise vers l'avant.

- **Suivez-nous sur LinkedIn @keyIT sa**
- **Envie d'en savoir plus ? :**

Envoyez-nous un message direct sur LinkedIn ou contactez-nous à sales@keyit.ch