



Reprenez le contrôle de vos données face aux géants américains

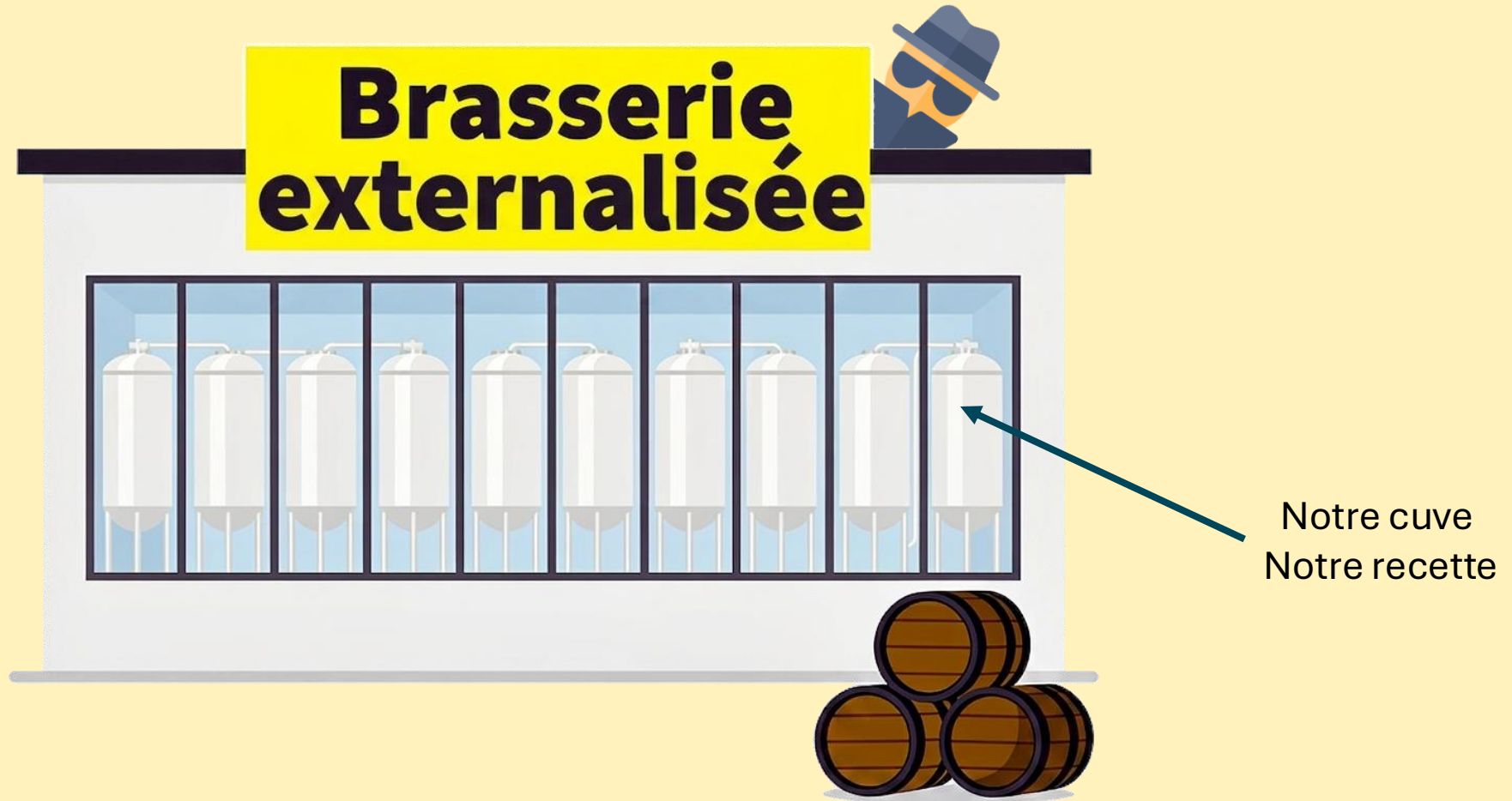
Nathan Leclercq / Jonathan Fritsch

Notre brasserie



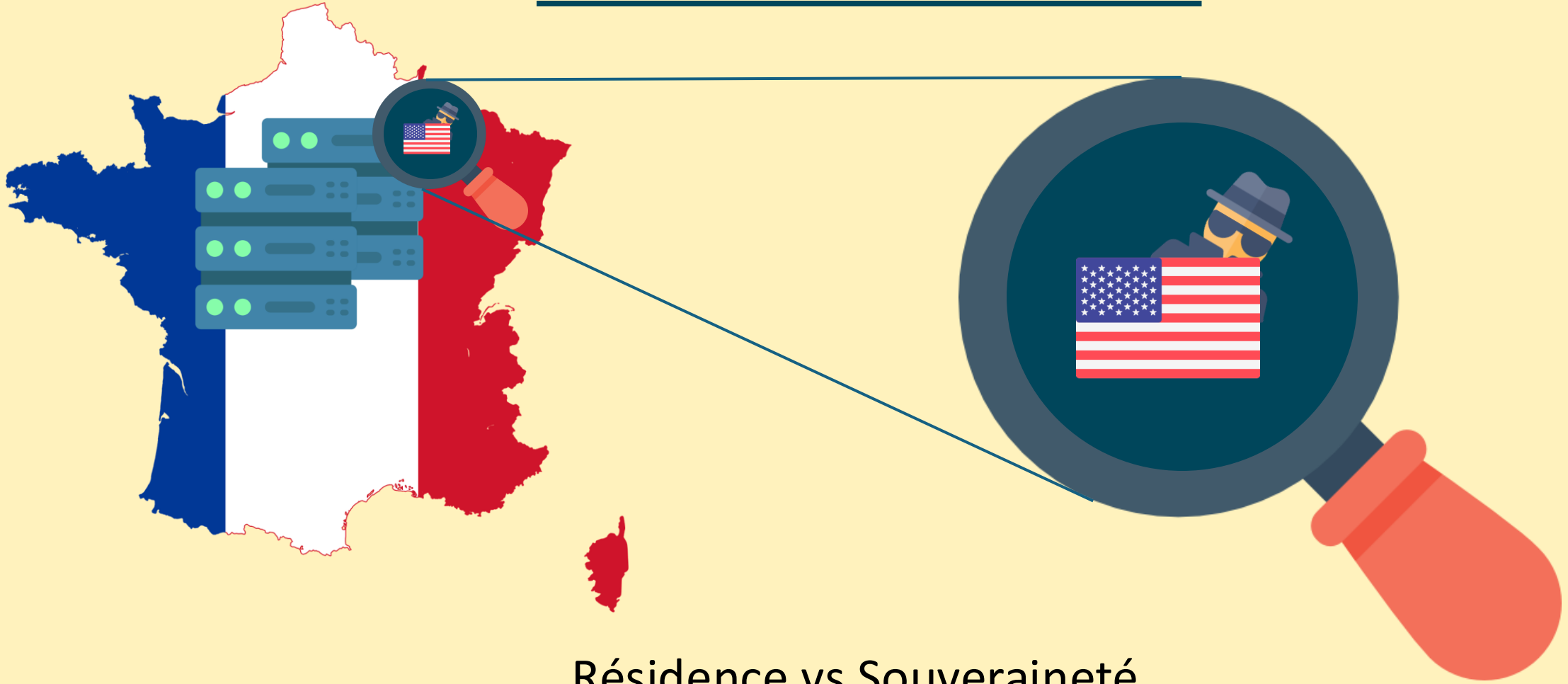
Notre cuve
Notre recette

Leur brasserie



Cloud Act

2018



Résidence vs Souveraineté

Intro

Nyukom



Réseau



Téléphonie

70 entités cibles

Intro

Contexte
métier

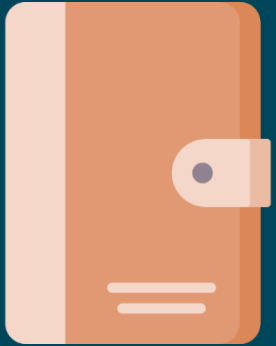
La brasserie ferme



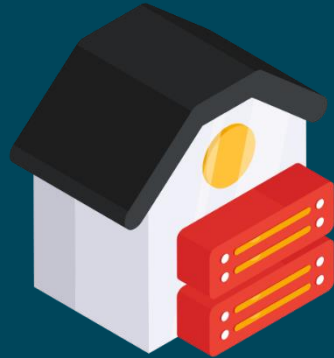
Intro

Contexte
métier

Les contraintes



Recettes
secrètes



Local
équipé



Budget
fixe



Équipe
technique

Intro

Contexte
métier

La Brasserie



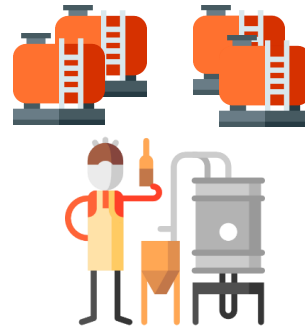
Champs

Données sources



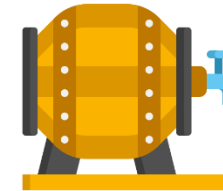
Entrepôt

Data Lake
Données brutes



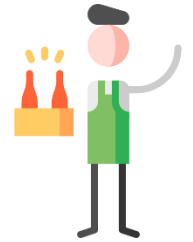
Brassage

Transformation



Fûts

Data Warehouse
Données préparées



Distributeur

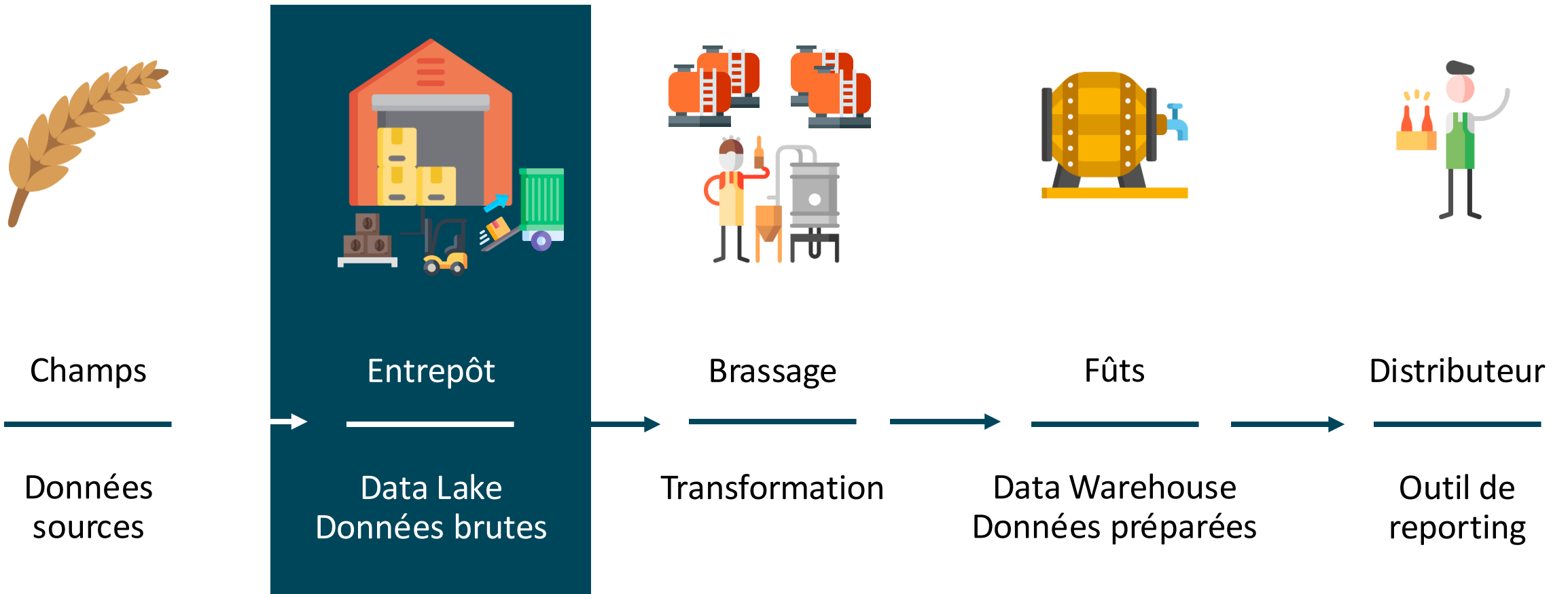
Outil de reporting

Intro

Contexte
métier

Architecture

La Brasserie

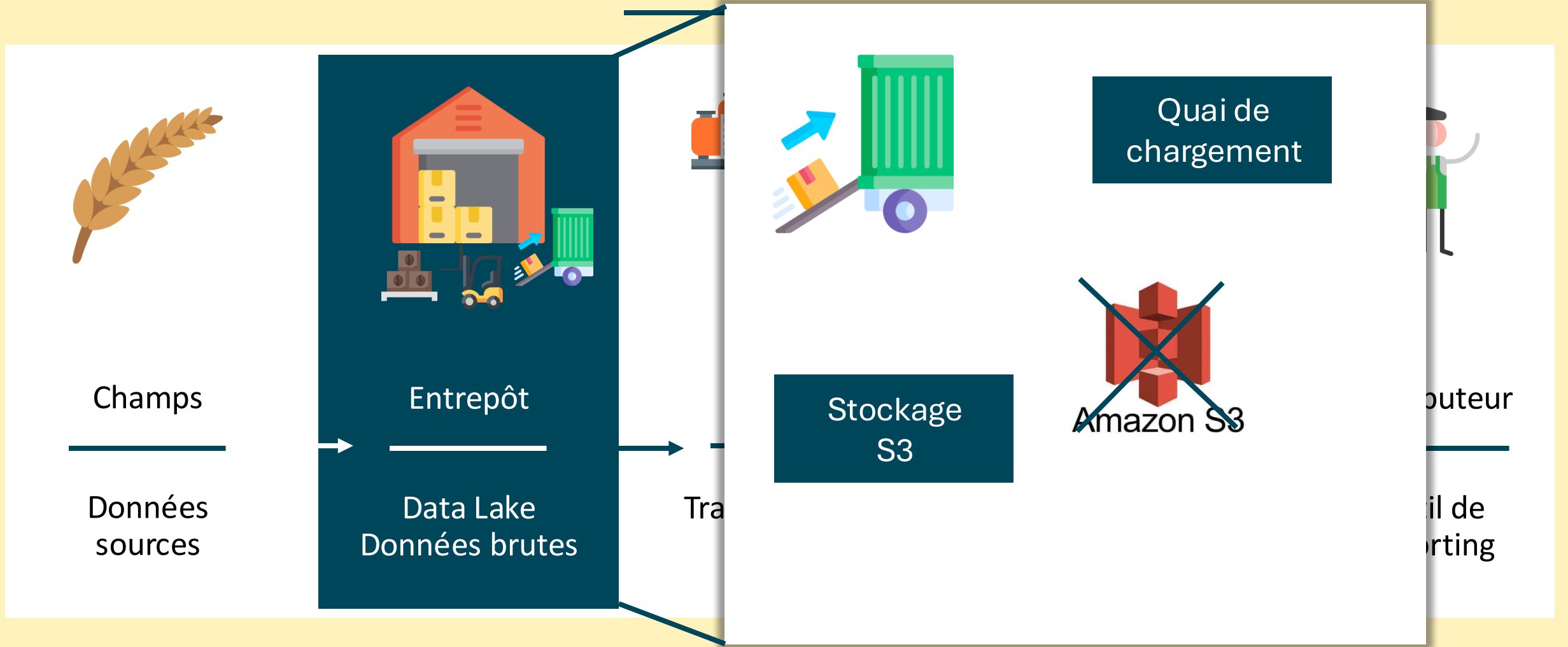


Intro

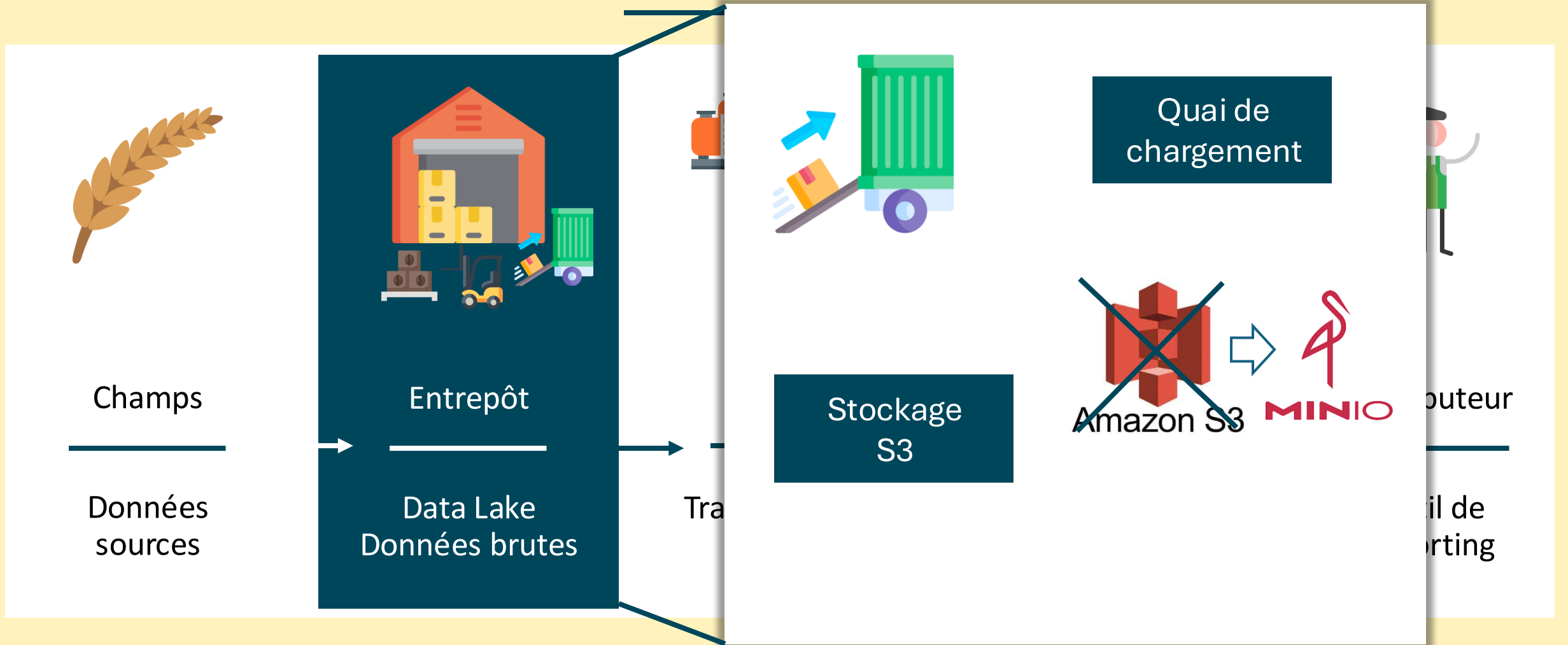
Contexte
métier

Architecture

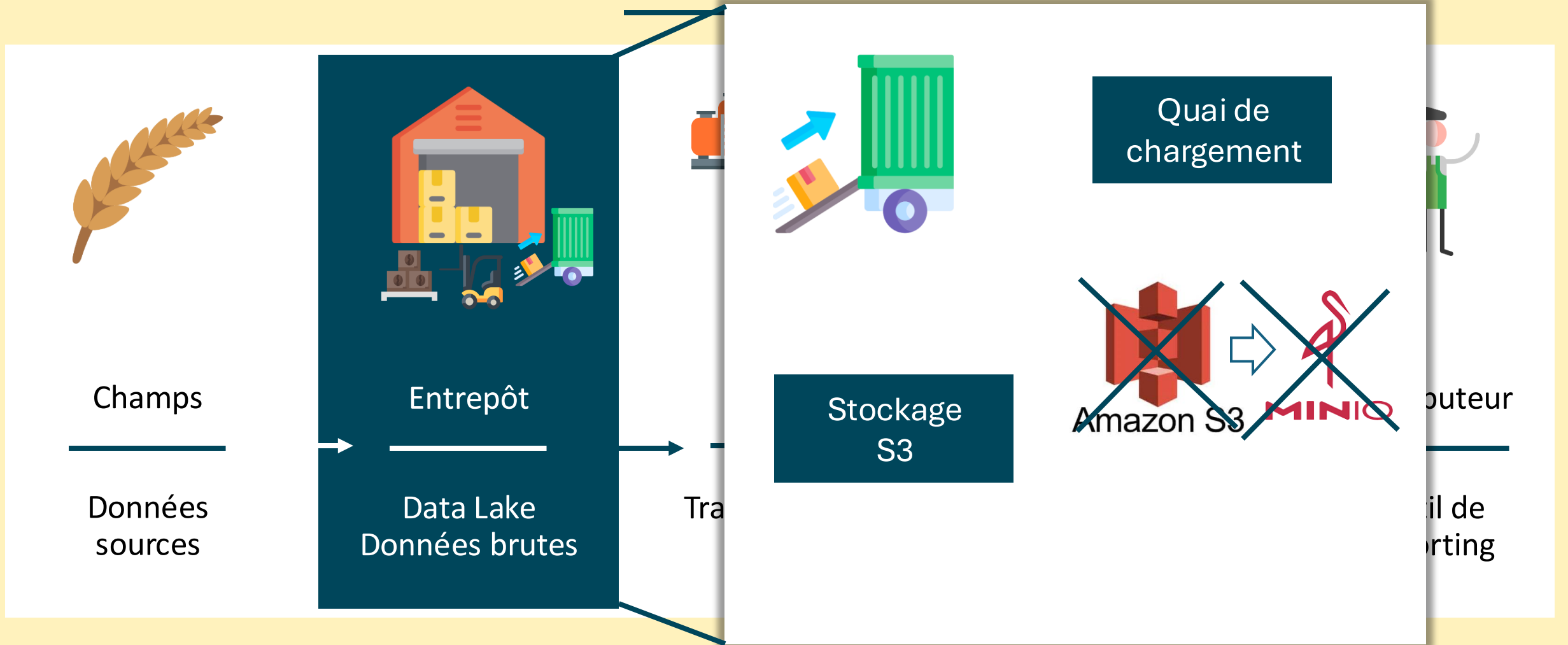
La Brasserie



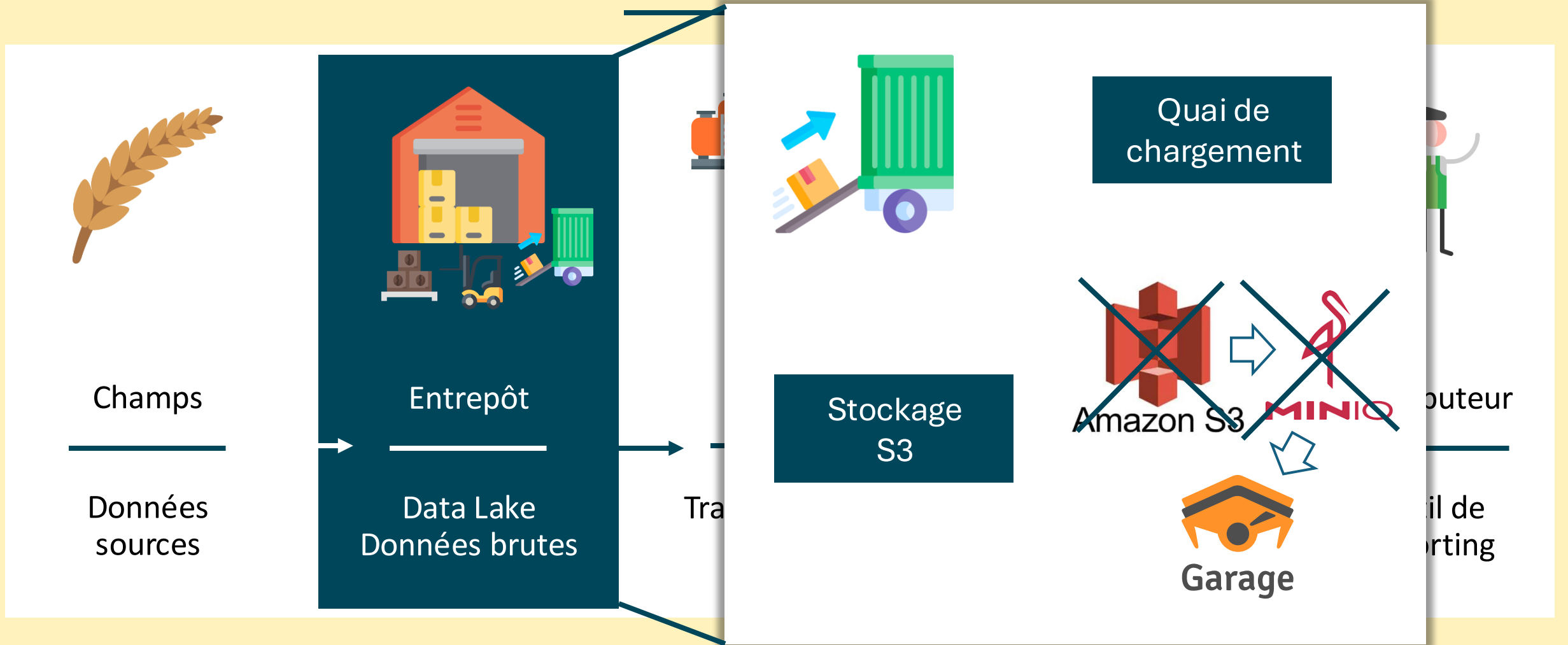
La Brasserie



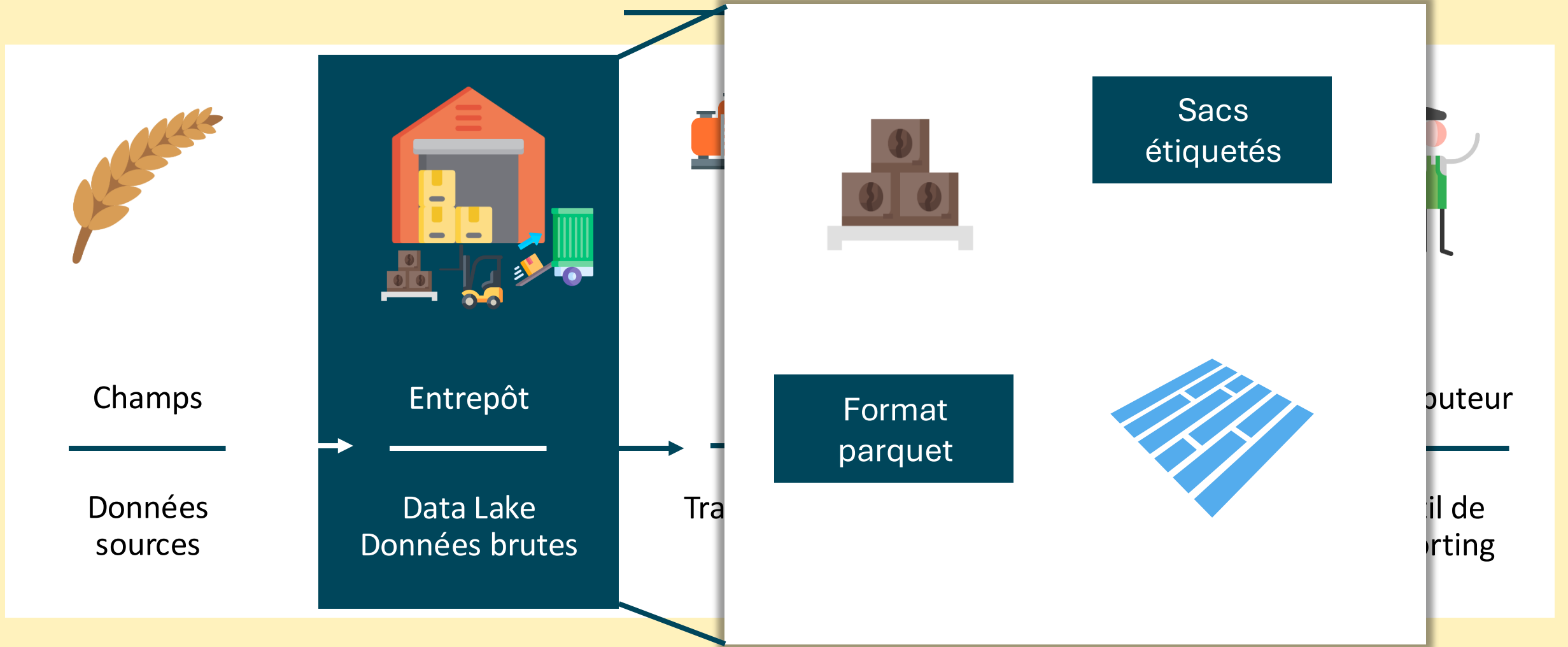
La Brasserie



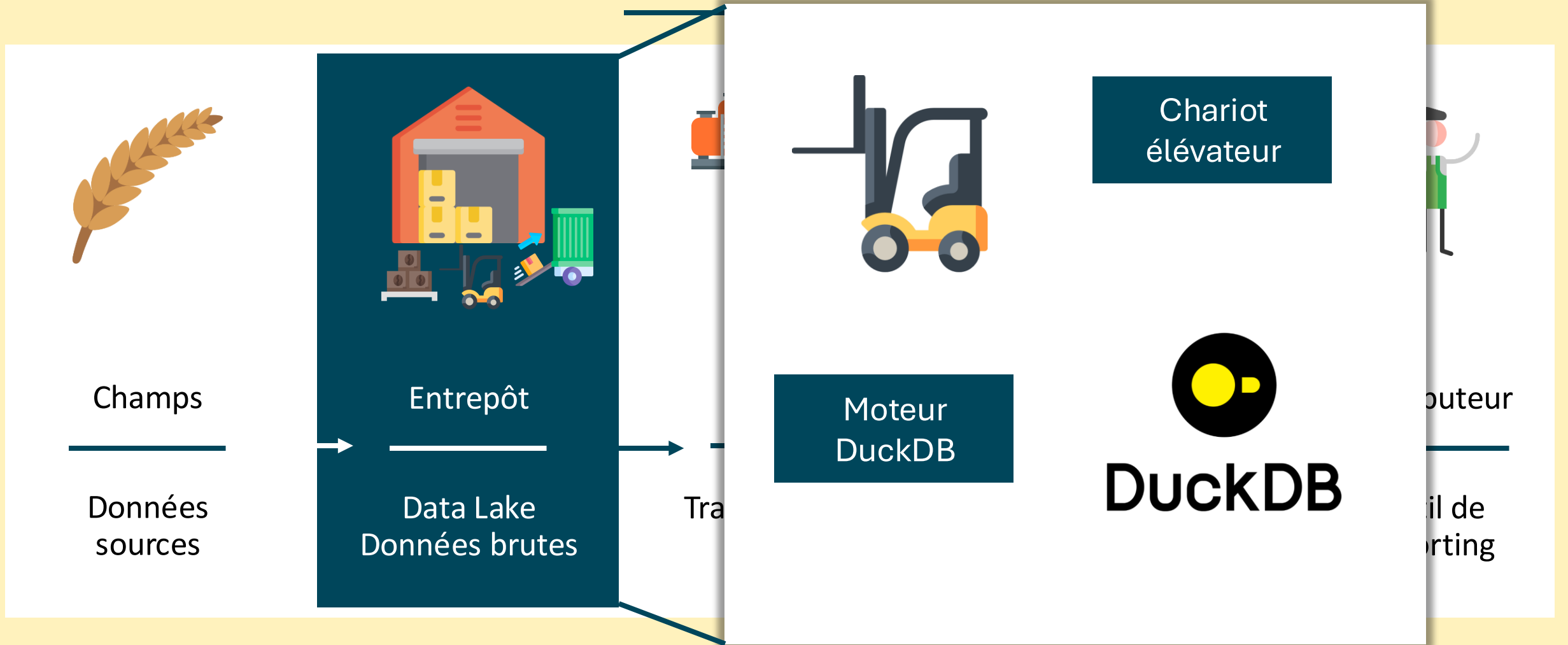
La Brasserie



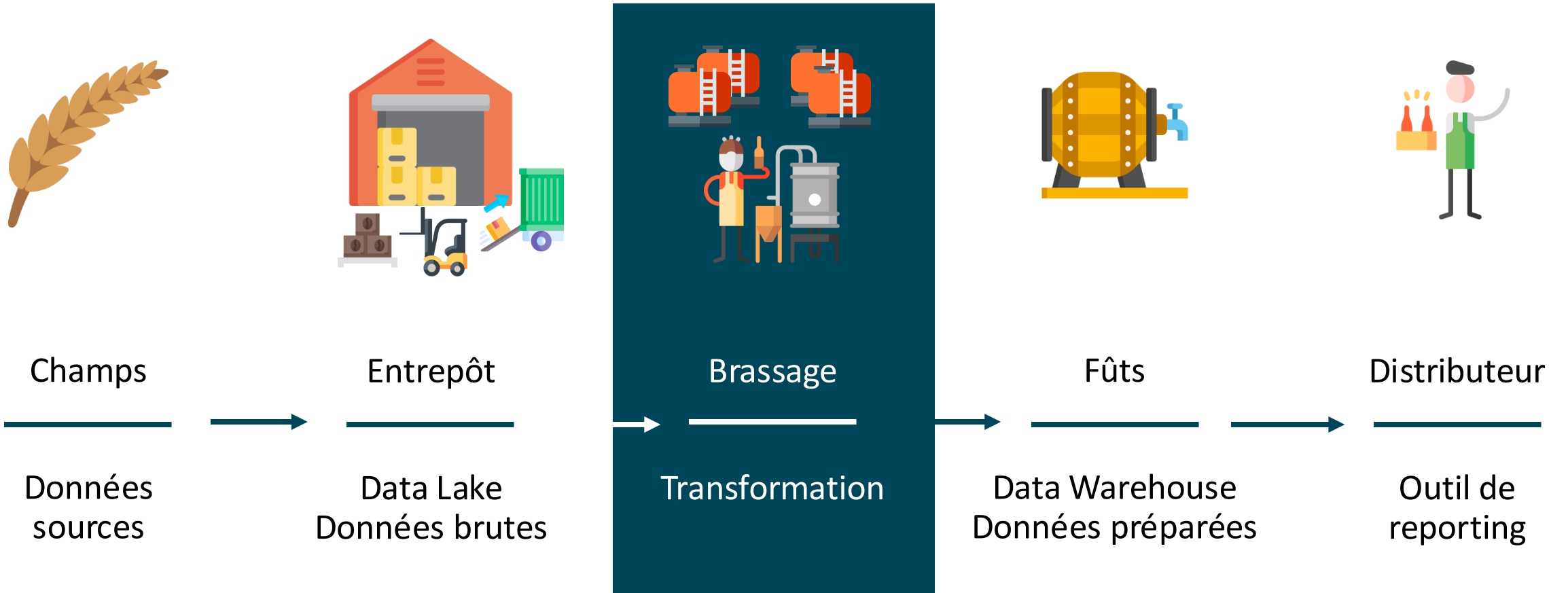
La Brasserie



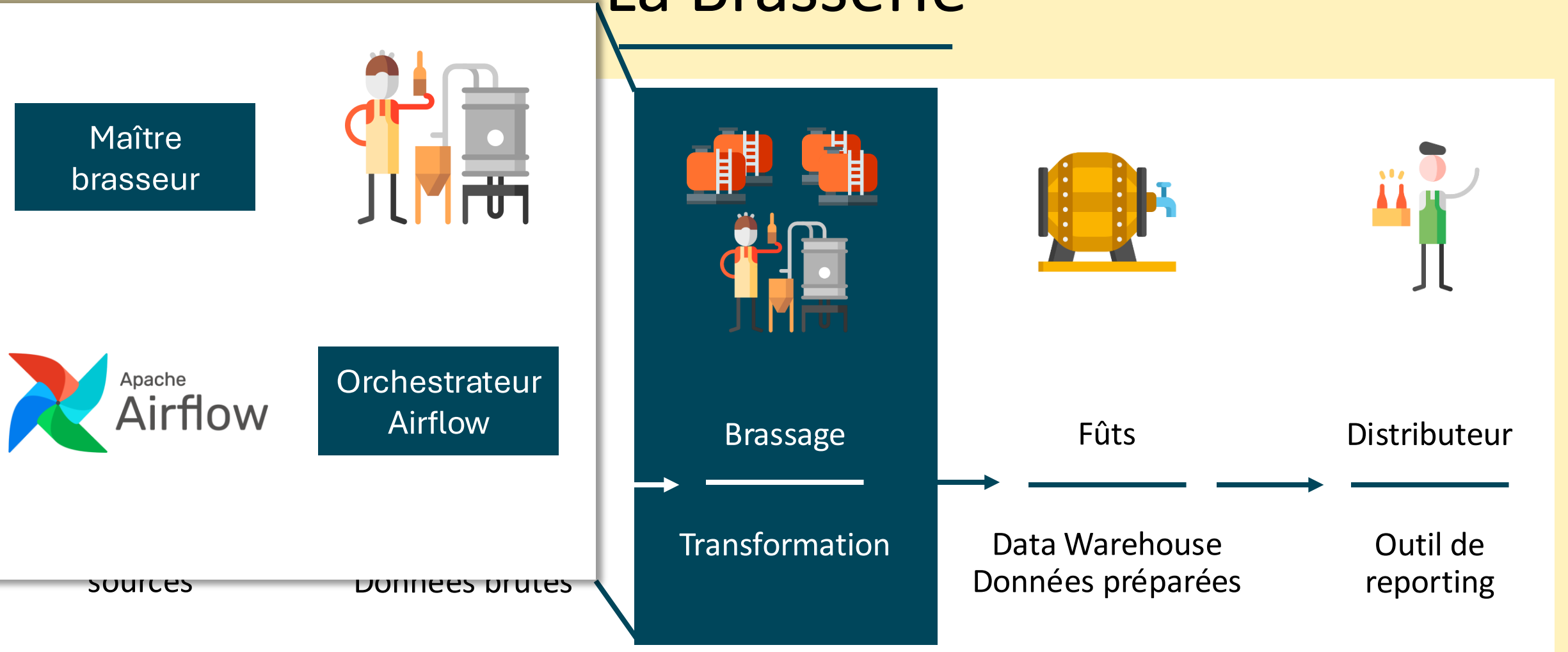
La Brasserie



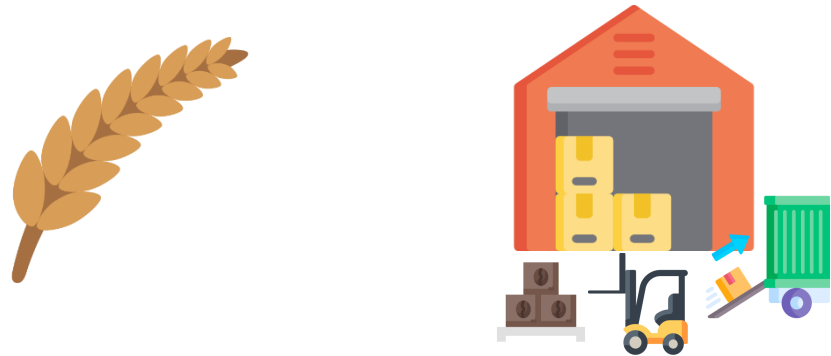
La Brasserie



La Brasserie



La Brasserie



Champs

Entrepôt



Brassage

Données sources

Data Lake
Données brutes

Transformation



Cuve

Pod
K3S



K3S

Intro

Contexte
métier

Architecture

La Brasserie



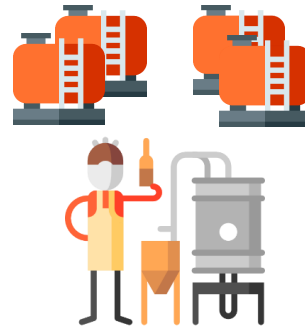
Champs

Données sources



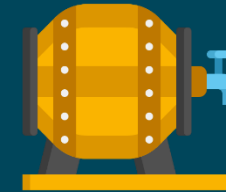
Entrepôt

Data Lake
Données brutes



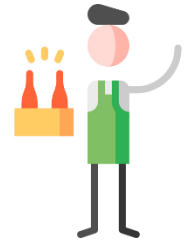
Brassage

Transformation



Fûts

Data Warehouse
Données préparées



Distributeur

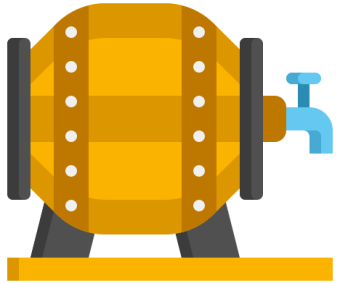
Outil de reporting

Intro

Contexte
métier

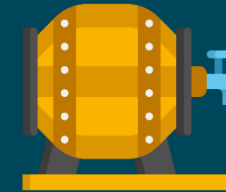
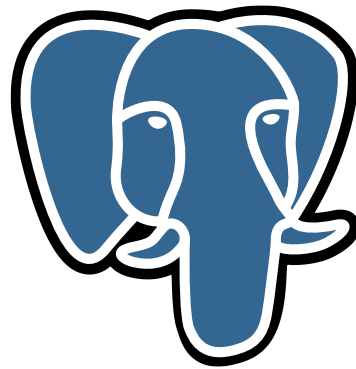
Architecture

La Brasserie



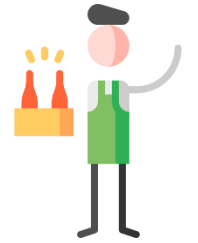
Fûts
étiquetés

BDD
PostgreSQL



Fûts

Data Warehouse
Données préparées



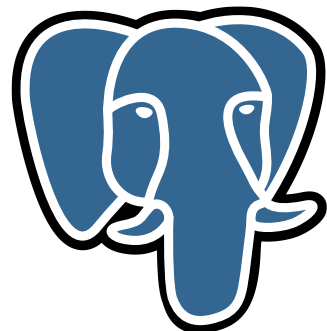
Distributeur

Outil de
reporting

Standards ouverts



S3



PostgreSQL

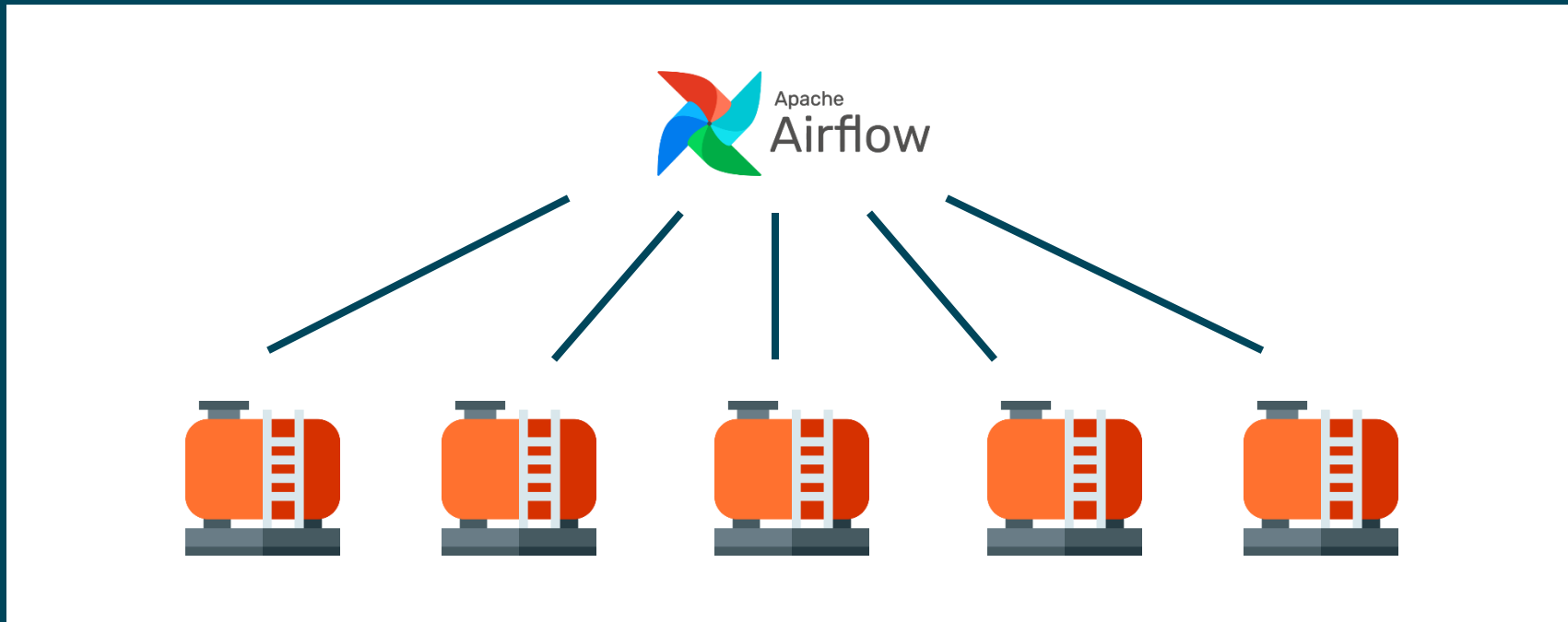


Parquet



Python

Scalabilité



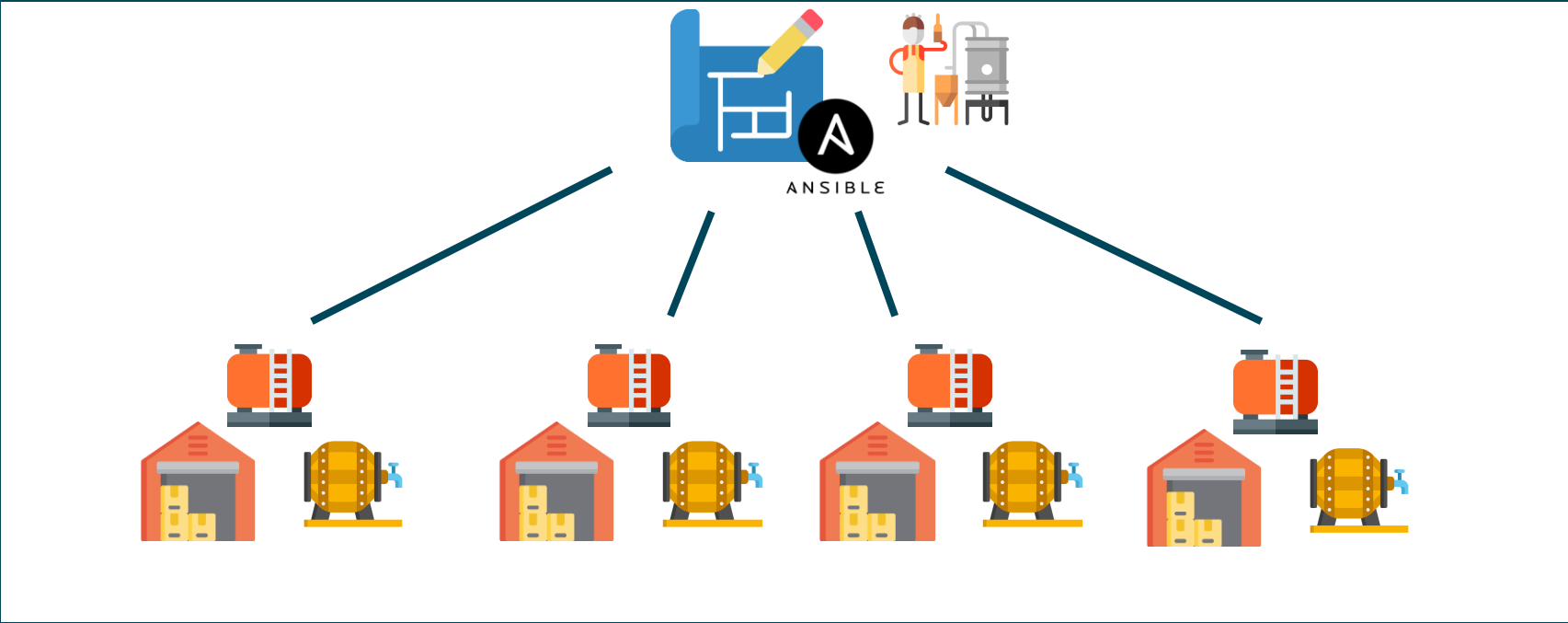
Intro

Contexte
métier

Architecture

Outils

Scalabilité



Résultats

43 M

50 000 lignes /jour

4,5 mois

Time to prod

2 ans

d'historique

0 €

Hors couts marginaux

Chez les géants ?



Microsoft Azure



Google Cloud

Intro

Contexte
métier

Architecture

Outils

Résultats

Cloud
public

Le pattern est toujours le même

Composant	Open source	GCP	AWS	Azure
Data lake	MinIO / Garage	Cloud Storage	S3	OneLake
Warehouse	PostgreSQL	BigQuery	Redshift	Fabric Warehouse / Azure SQL Database
Orchestration	Airflow	Cloud Composer / Dataflow	MWAA / Glue / Step Functions	Data Factory
Compute	K3s + DuckDB	Spark	Spark	Spark
Infra-as-code	Ansible	Terraform	Terraform	Terraform

Intro

Contexte
métier

Architecture

Outils

Résultats

Cloud
public

Chez Microsoft

Notre stack OSS	Azure (Fabric-natif)
Garage / MinIO (S3)	OneLake (intégré Fabric)
PostgreSQL warehouse	Fabric Warehouse (Azure SQL Database)
Airflow	Data Factory dans Fabric
K3s + DuckDB	Spark
Ansible	Terraform

Poste	€/mois
Microsoft Fabric F8	481,66
Azure Monitor	36,44
TOTAL	518,10

6220 € / an

Intro

Contexte
métier

Architecture

Outils

Résultats

Cloud
public

Ce qu'on ne voit pas

Axe	OSS auto-hébergé	Azure Fabric	Google Cloud
Coût cloud (€/an)	0	~6 220 €	Variable
Prévisibilité des coûts	Prévisible — on scale, on maîtrise	Opaque — forfait bloquant au dépassement	Opaque — explose hors budget
Time-to-prod initial	4,5 mois	~3 mois	~3 mois
Charge humaine en run	~1 jour / mois	~½ jour / mois + FinOps	~½ jour / mois + FinOps
Souveraineté juridique	Totale	Cloud Act applicable	Cloud Act applicable
Lock-in	Standards ouverts	API + formats propriétaires	API + formats propriétaires
Réversibilité	Prouvée (rclone sync)	Coût de sortie élevé	Coût de sortie élevé

Intro

Contexte
métier

Architecture

Outils

Résultats

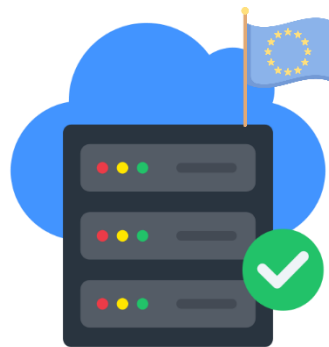
Cloud
public

Hébergement



Google Cloud

Cloud



SecNumCloud
(cloud privé)



Auto-hébergé

Intro

Contexte
métier

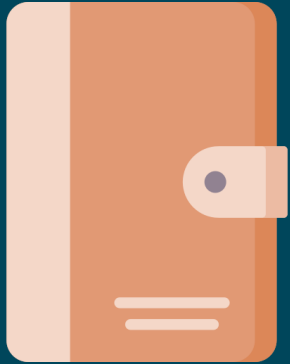
Architecture

Outils

Résultats

Cloud
public

Les pièges



Courbe
d'apprentissage



Maintenance



Effet tunnel

Intro

Contexte
métier

Architecture

Outils

Résultats

Cloud
public

Comparaison
des cloud

GO / no GO



Données sensibles

Volume stable

Ops interne

Infra existante

Budget contraint



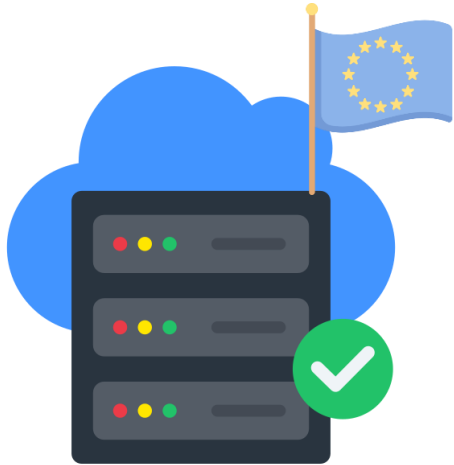
Pas d'OPS

Charges variables

Équipe trop petite

Faible volume

Alternatives



Clouds
européens



Hybridation

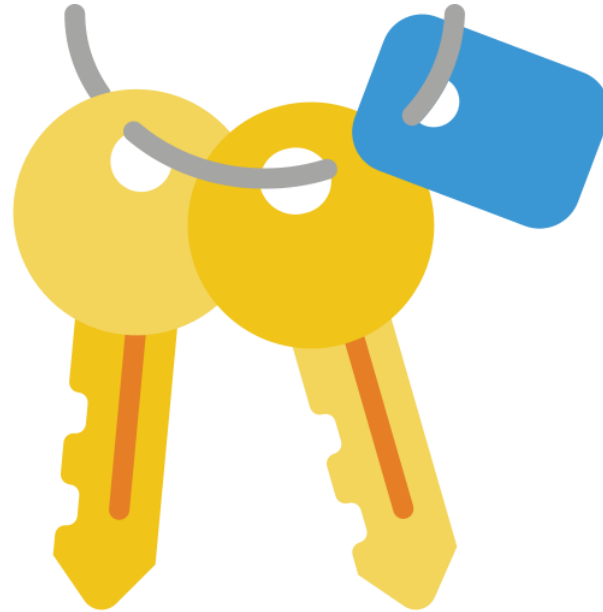
Reprenez le contrôle

1. Le cloud privé n'est pas une utopie
2. La souveraineté c'est une architecture, pas un fournisseur
3. L'écosystème existe, et il est local

À qui je donne les clés ?



Cloud
Provider



Equipe
Technique

Questions



Jonathan FRITSCH

Data Architect &
Fondateur de Datakhi

LinkedIn



Nathan LECLERCQ

ML Ops chez
Datakhi

LinkedIn



Datakhi

