



Donjon et python

**La chevaleresse à la
rescousse**



Il était une fois...

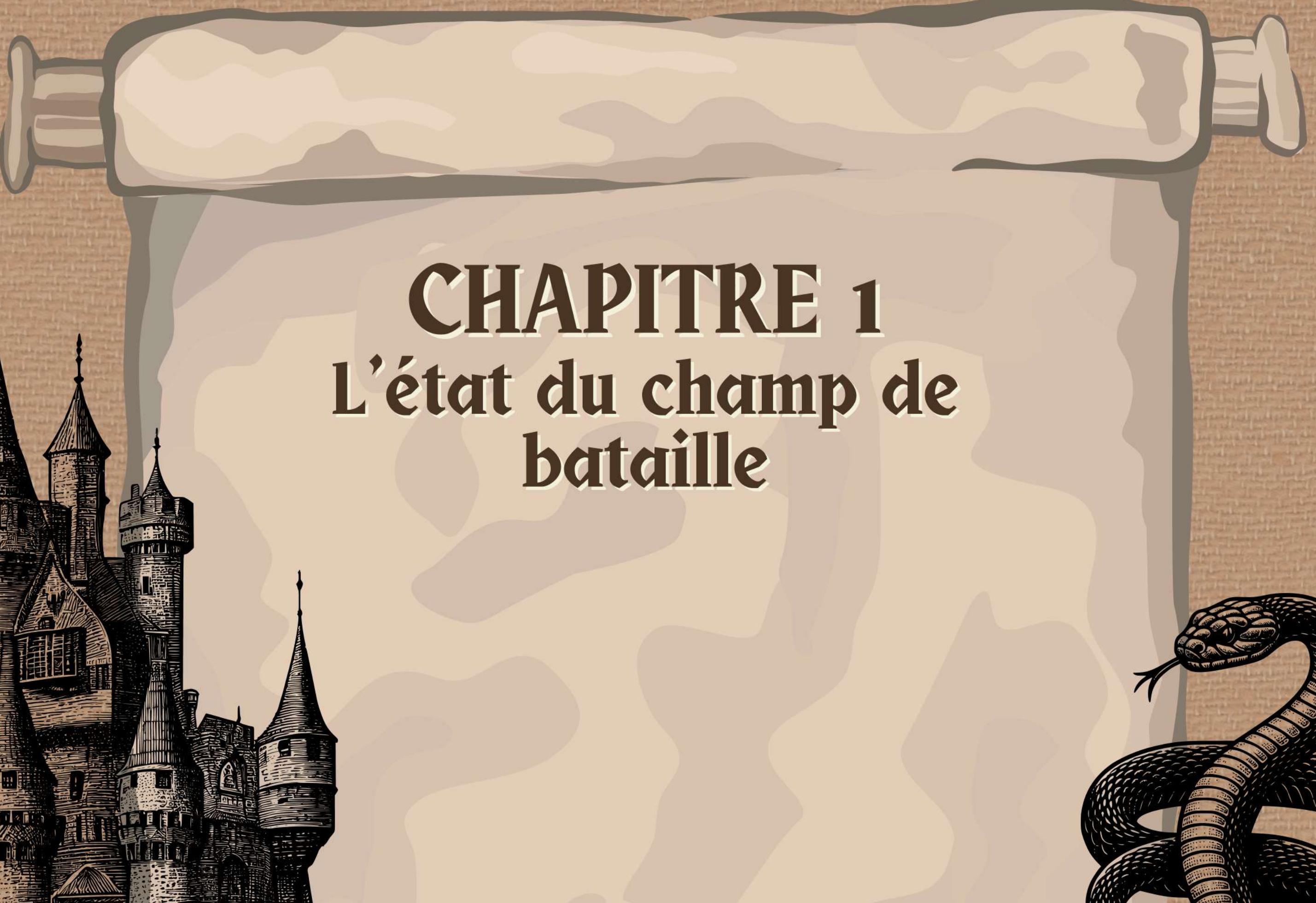




Il était une fois...

Juliette GRILLERE aka
la Chevaleresse Data



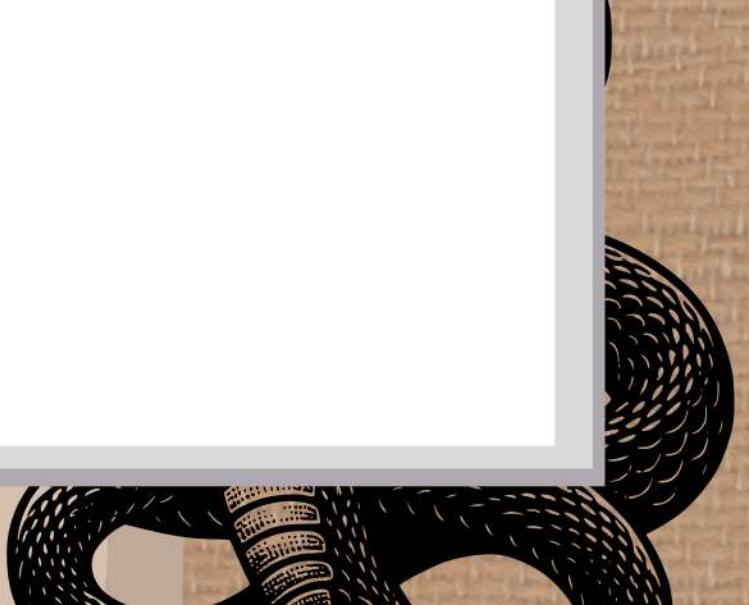
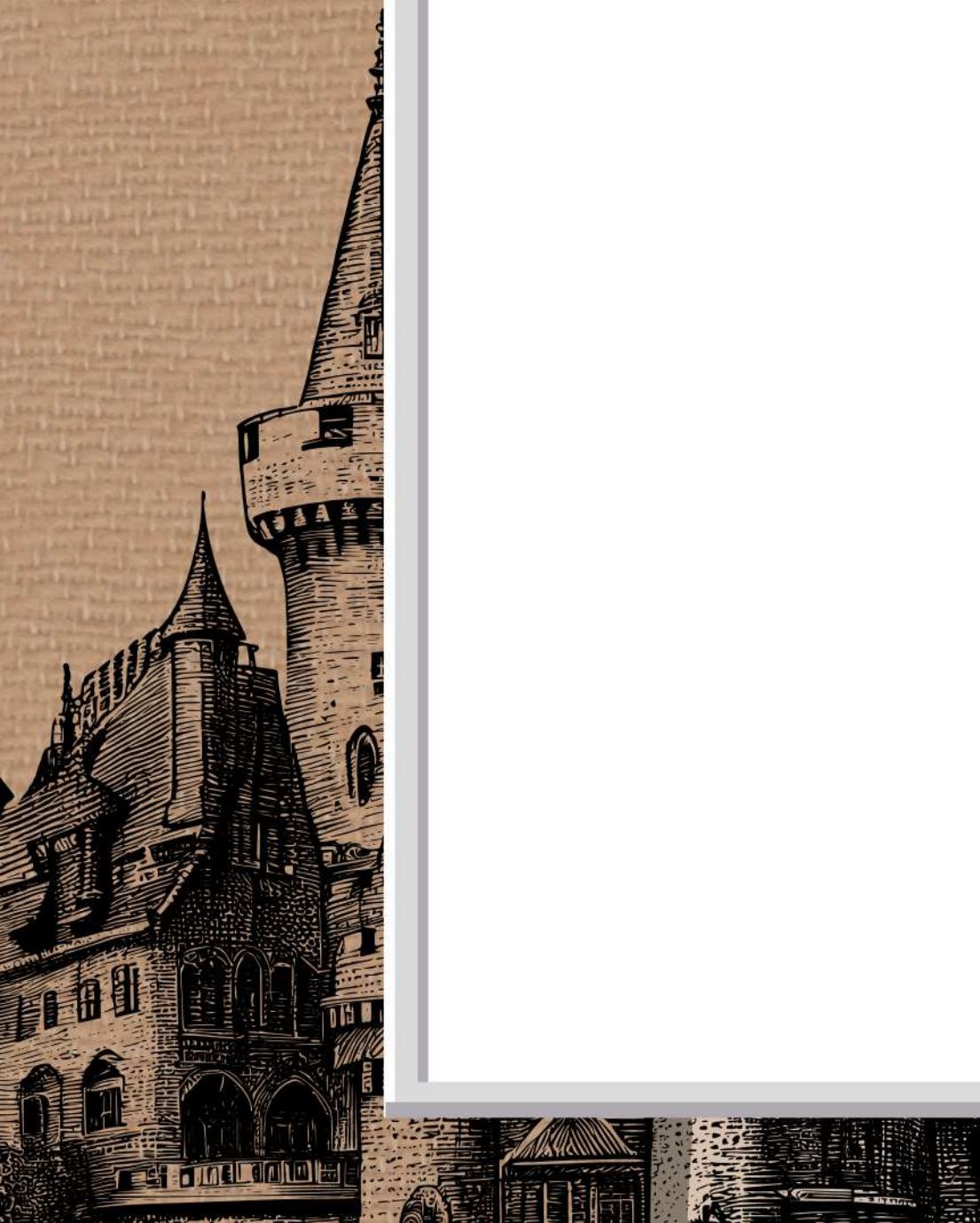
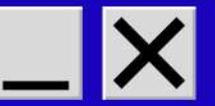


CHAPITRE 1

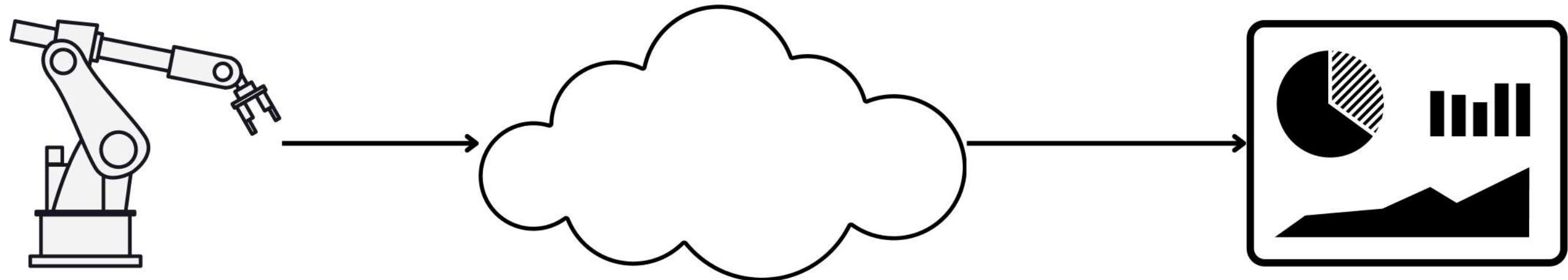
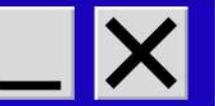
L'état du champ de bataille



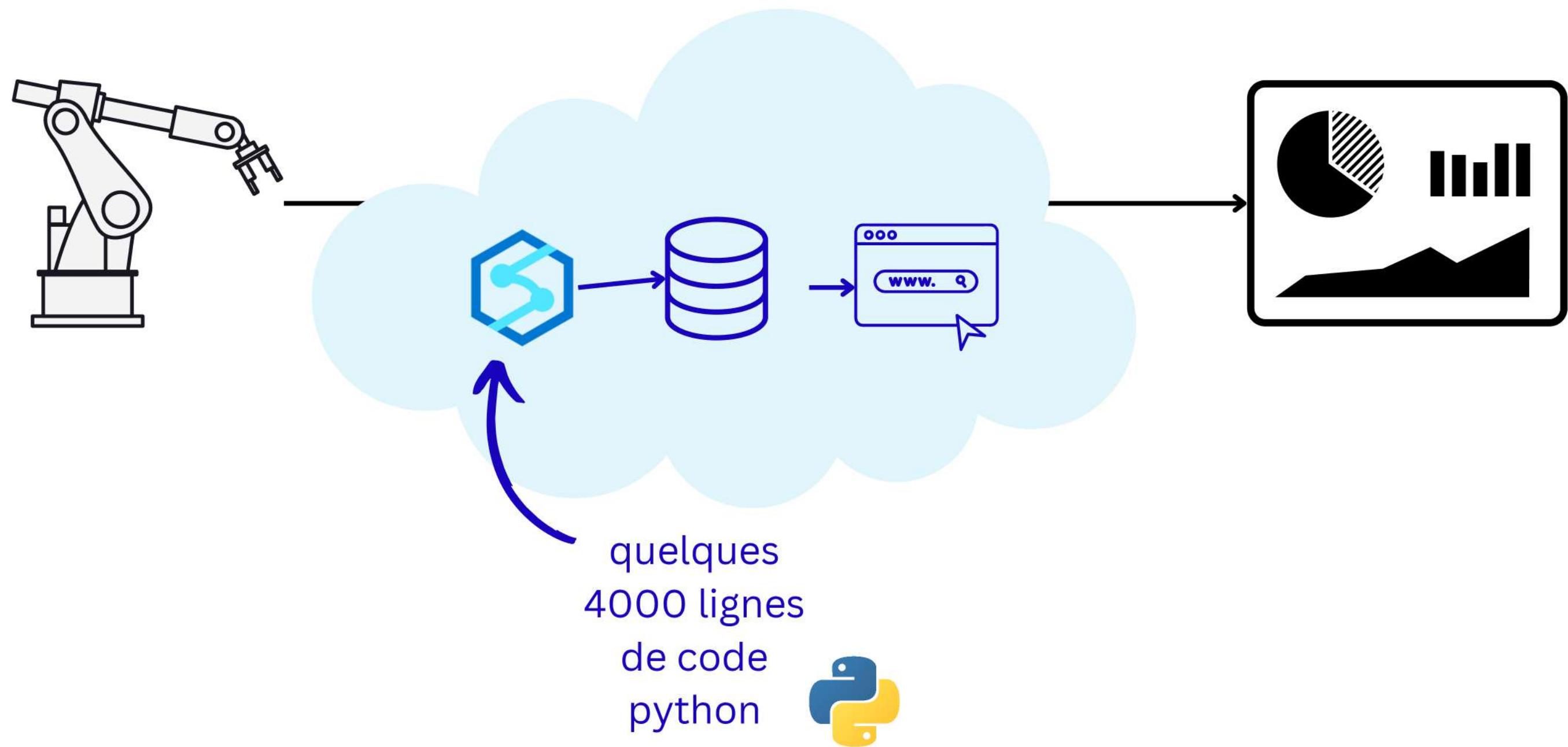
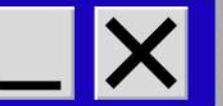
Retour à la modernité



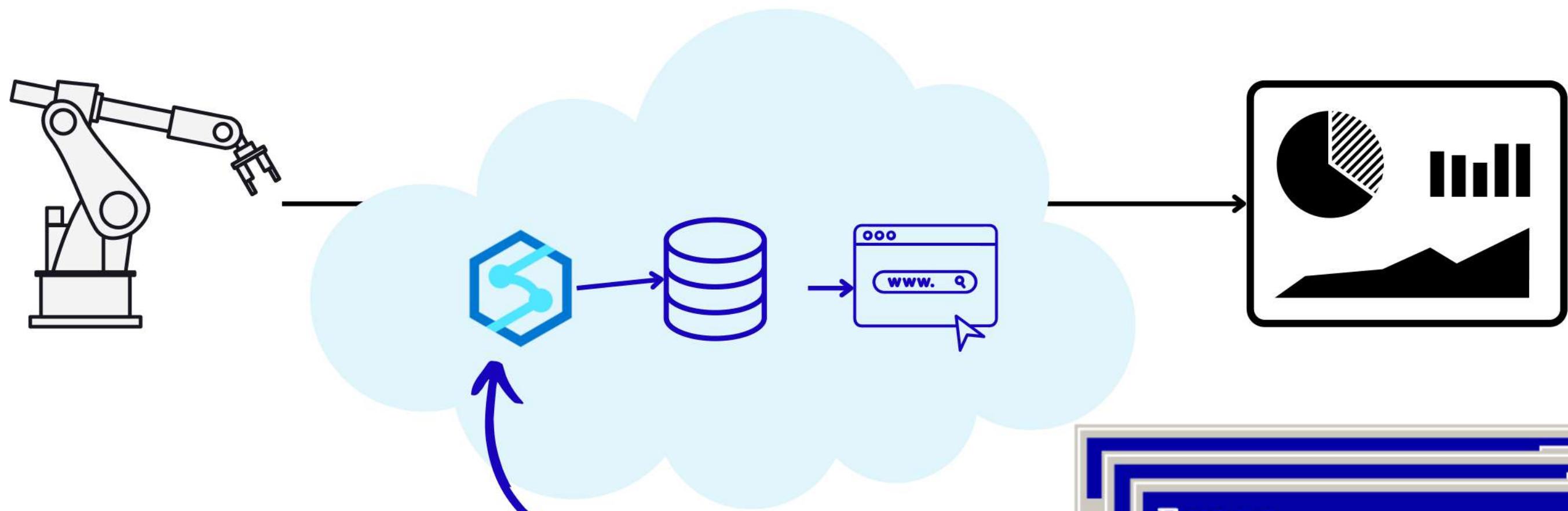
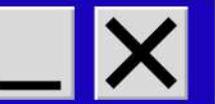
Retour à la modernité



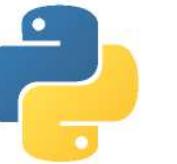
Retour à la modernité

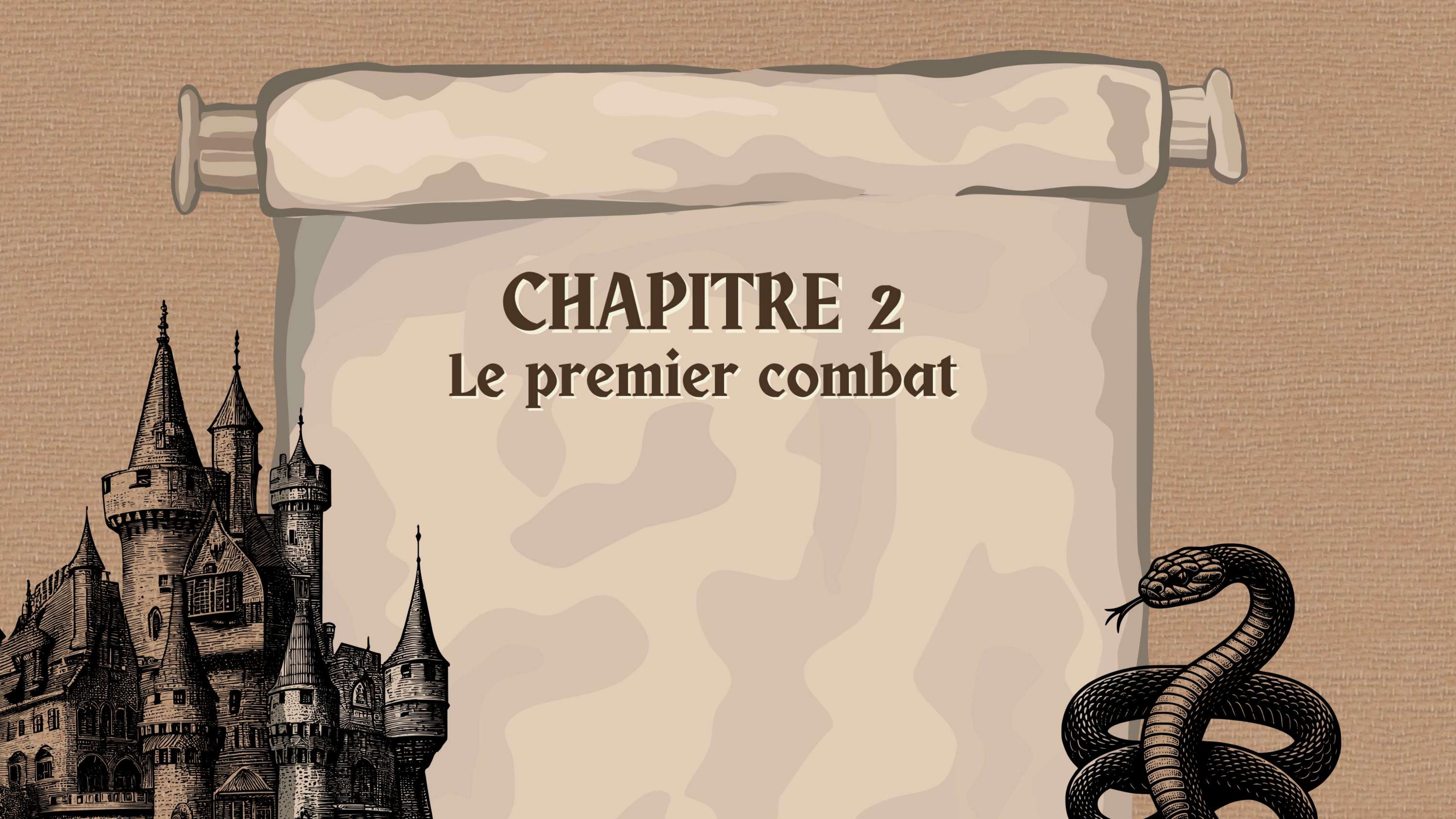


Retour à la modernité



quelques
4000 lignes
de code
python



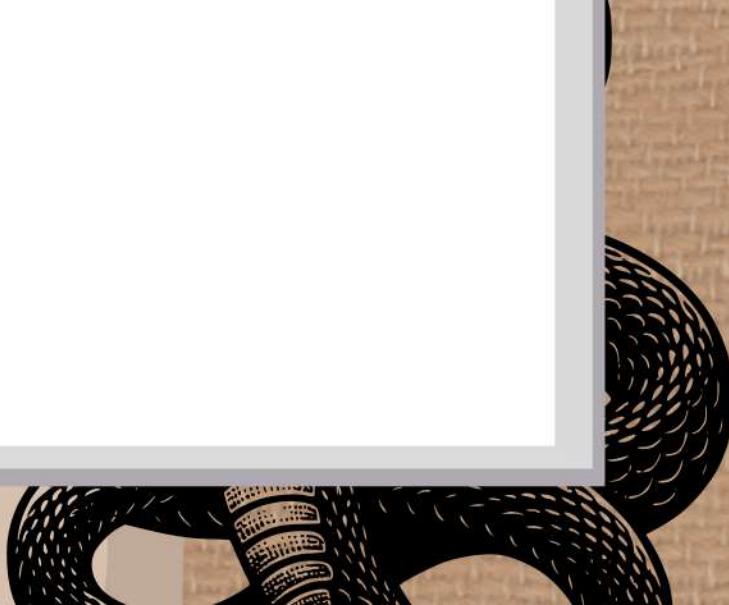
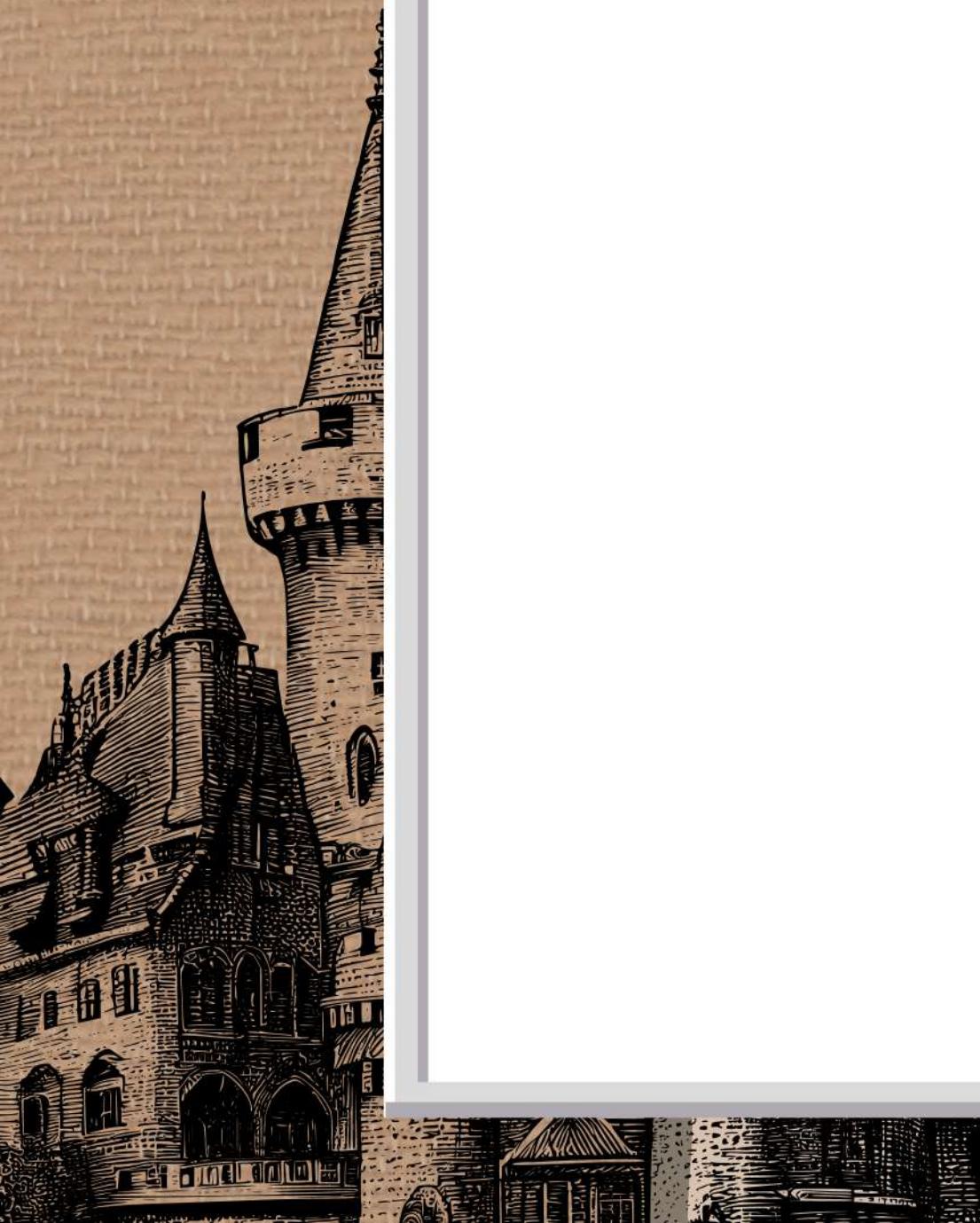
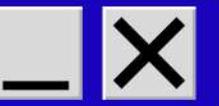


CHAPITRE 2

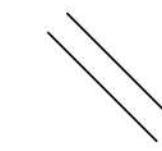
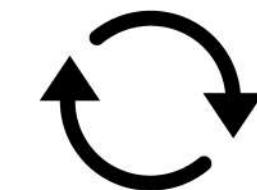
Le premier combat



Retour à la modernité



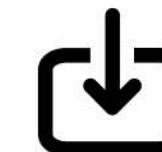
Optimisation de



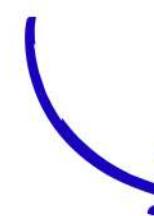
Paralléliser certains traitements



Découper le code, le rendre plus
lisible



Optimiser la façon dont sont importées
les données (lecture des fichiers)



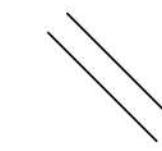
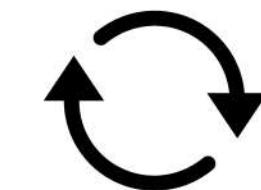
Coût réduit de 20%



Temps de traitement à
45min au lieu de 1h

Mieux mais pas satisfaisant

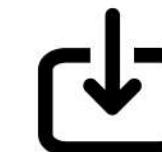
Optimisation de



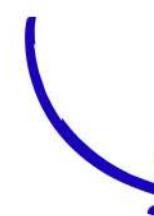
Paralléliser certains traitements



Découper le code, le rendre plus
lisible



Optimiser la façon dont sont importées
les données (lecture des fichiers)

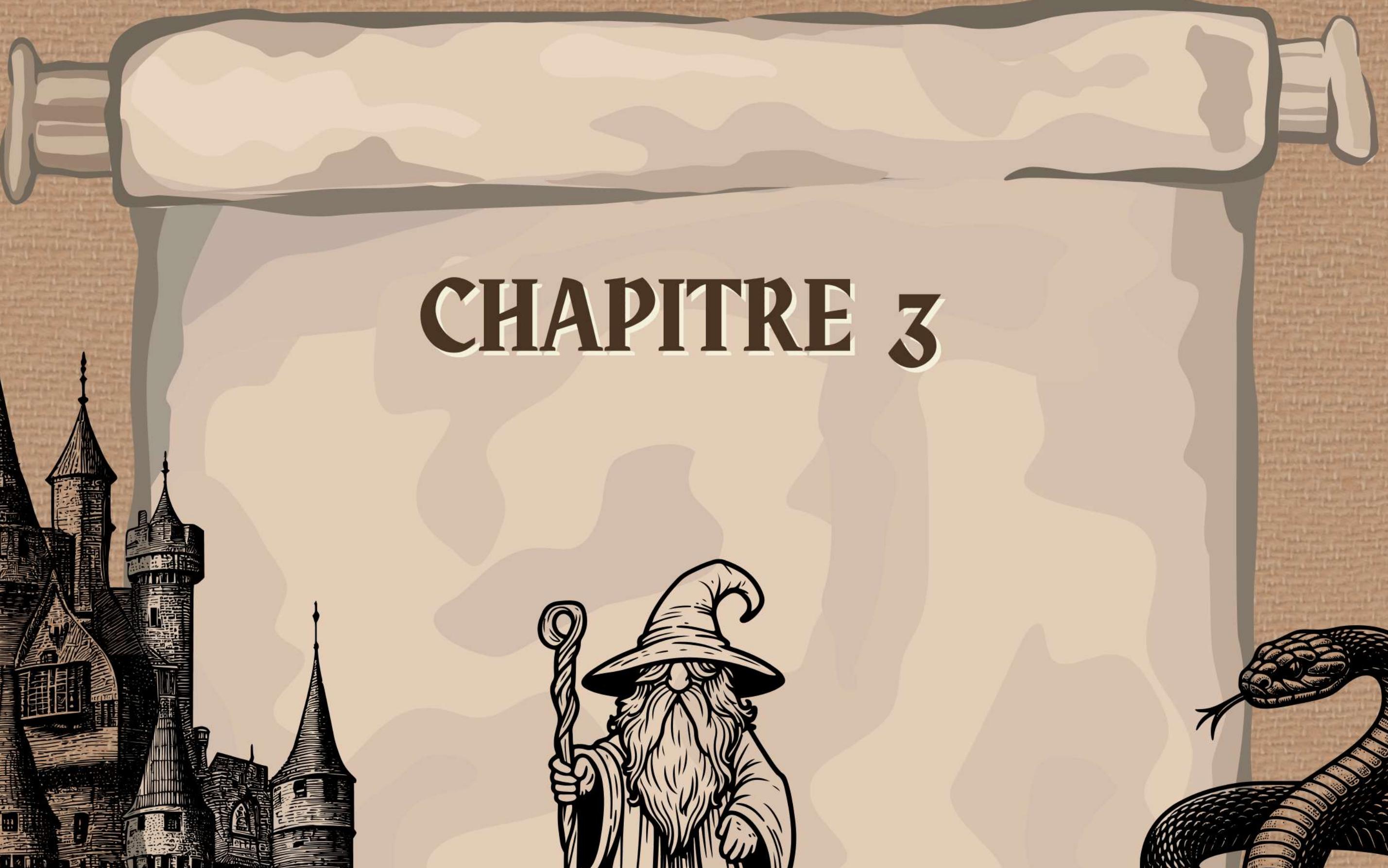


Coût réduit de 20%



Temps de traitement à
45min au lieu de 1h

Mieux mais pas satisfaisant



CHAPITRE 3

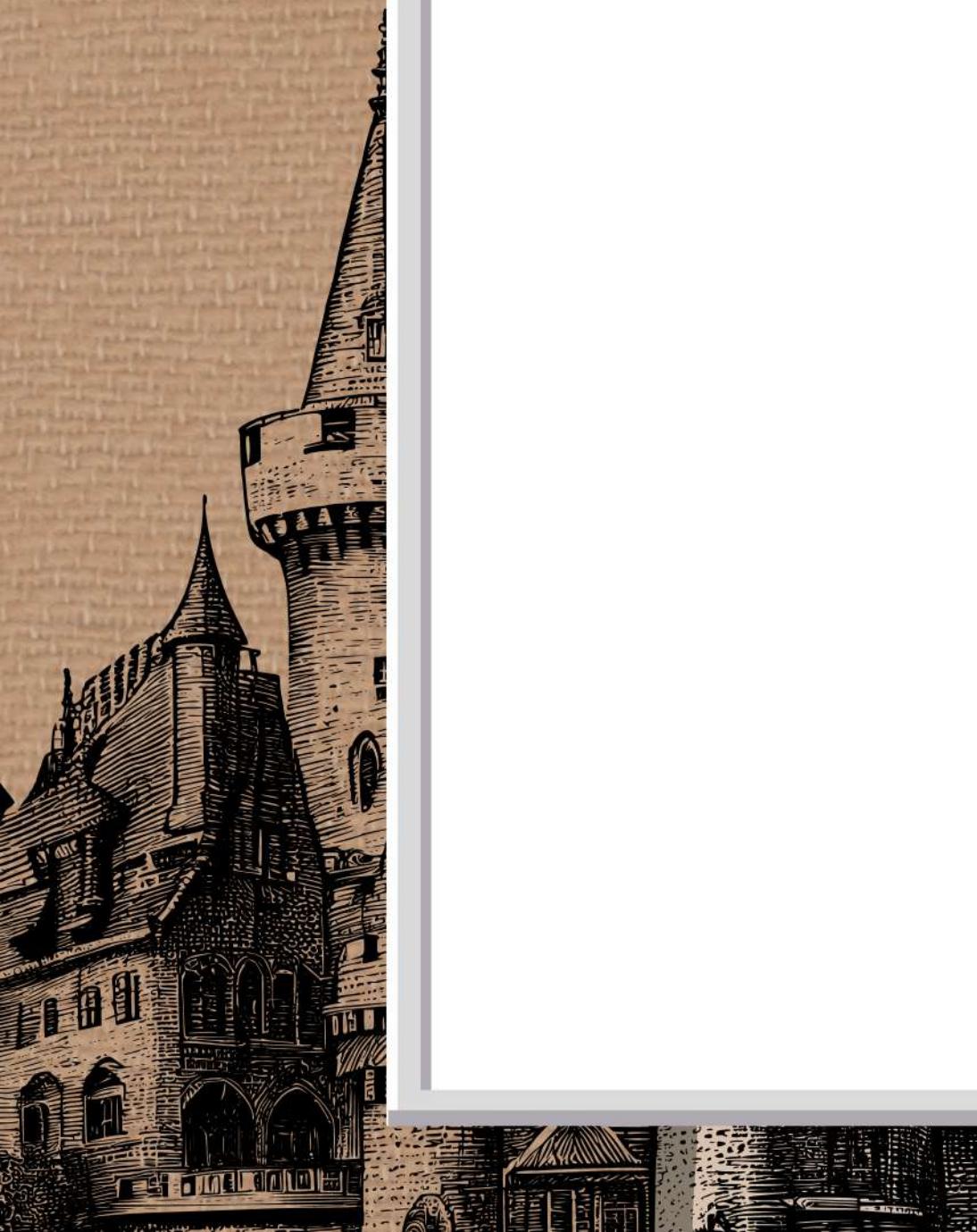
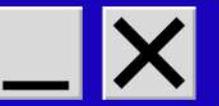


CHAPITRE 3

Trouver un substitut au python



Retour à la modernité



Azure Synapse

- Traitement en **batch**
- Coûte environ 1000€/mois



Azure Stream analytics

- Traitement en **stream**
- Coût estimé à 100€/stream/mois

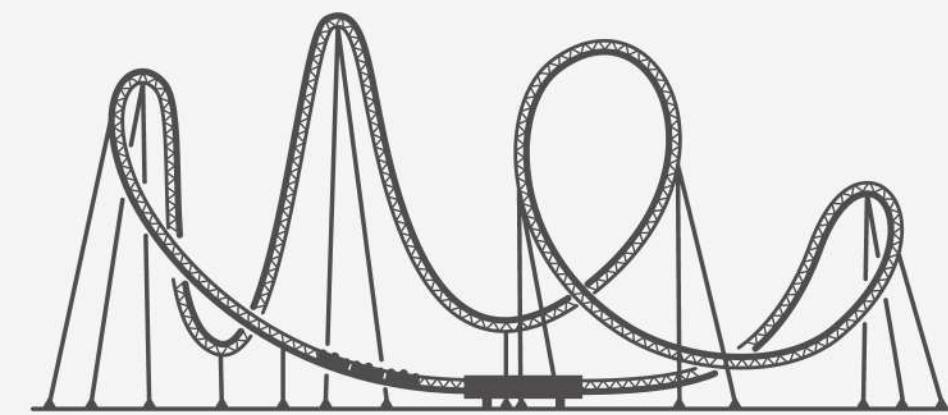


Les enjeux : stateful VS stateless

Stateful



On sait tout !



Stateless



On ne connaît l'état
qu'à l'instant T



Les enjeux : ranger la donnée

Base de données



Dédiée à un usage précis



Donnée structurée



La donnée est optimisée pour l'analyse/le reporting



La donnée est historisée

Datalake



Ensemble des données d'un projet (voire de l'entreprise)



La donnée peut être non structurée



La donnée n'est pas (ou peu) transformée

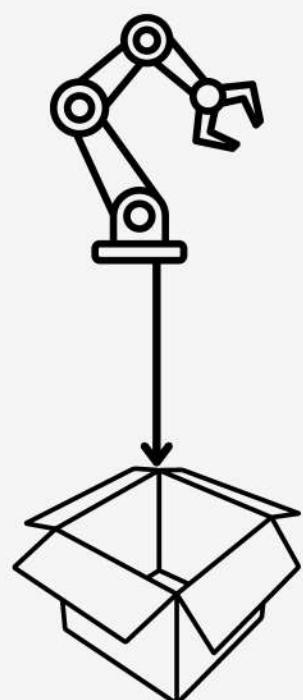
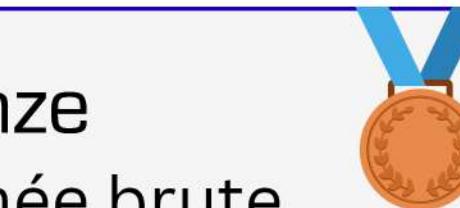


La donnée n'est pas forcément fiable

Les enjeux : ranger la donnée

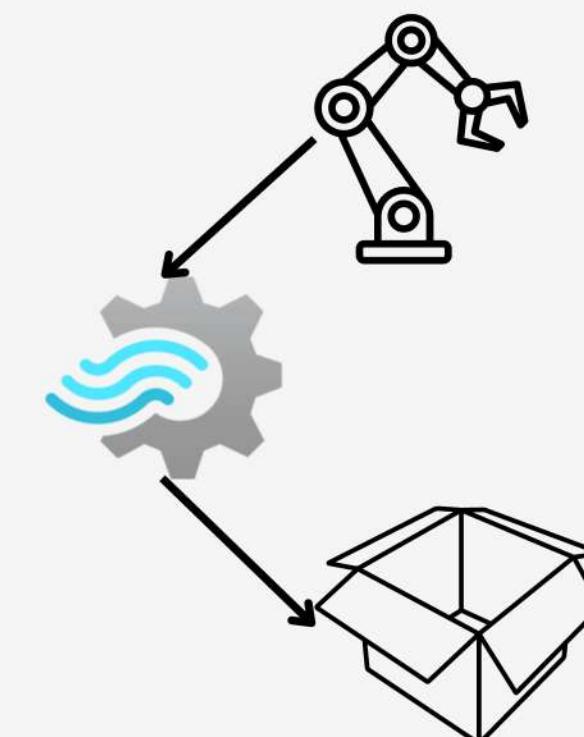
Bronze

Donnée brute



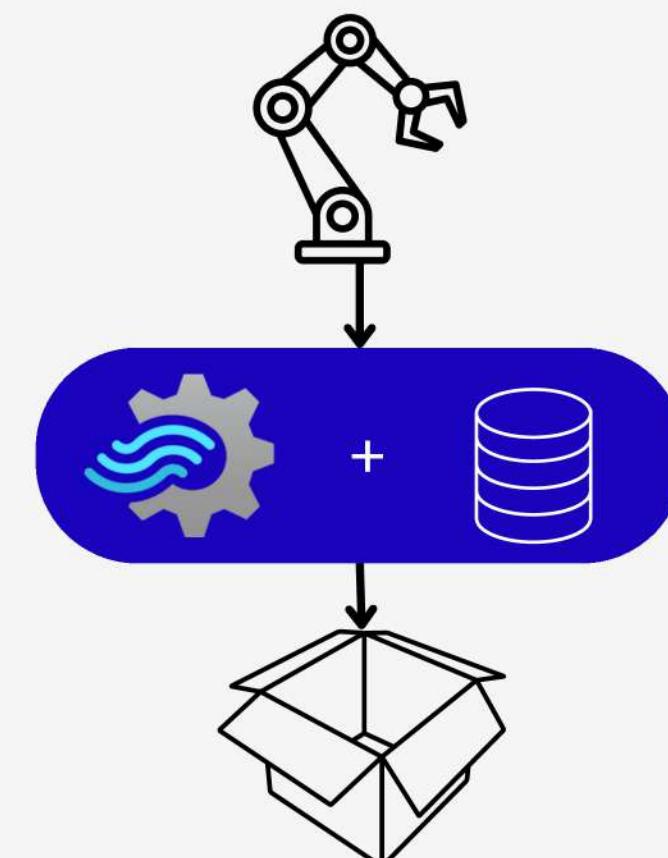
Silver

Donnée validée



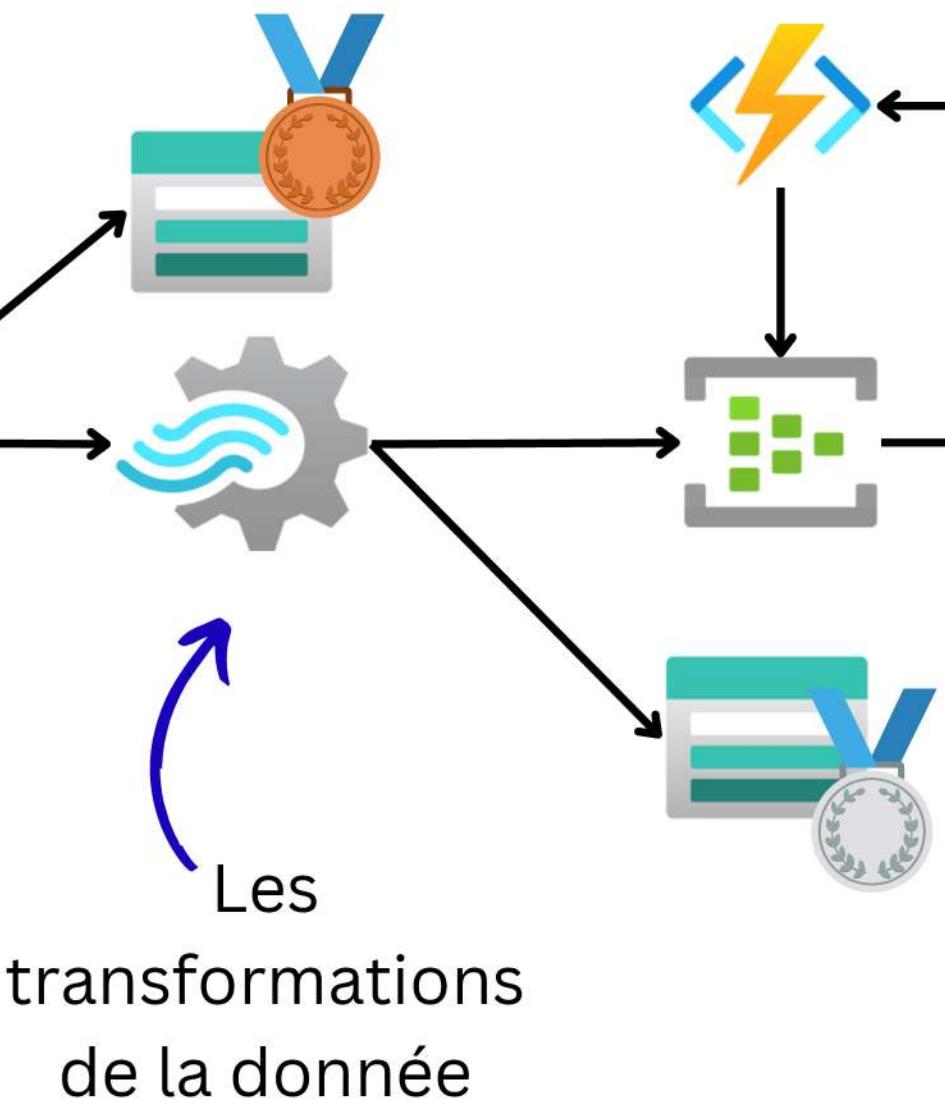
Gold

Donnée enrichie



Du Kafka
déguisé en
ressource

Azure



Les
transformations
de la donnée

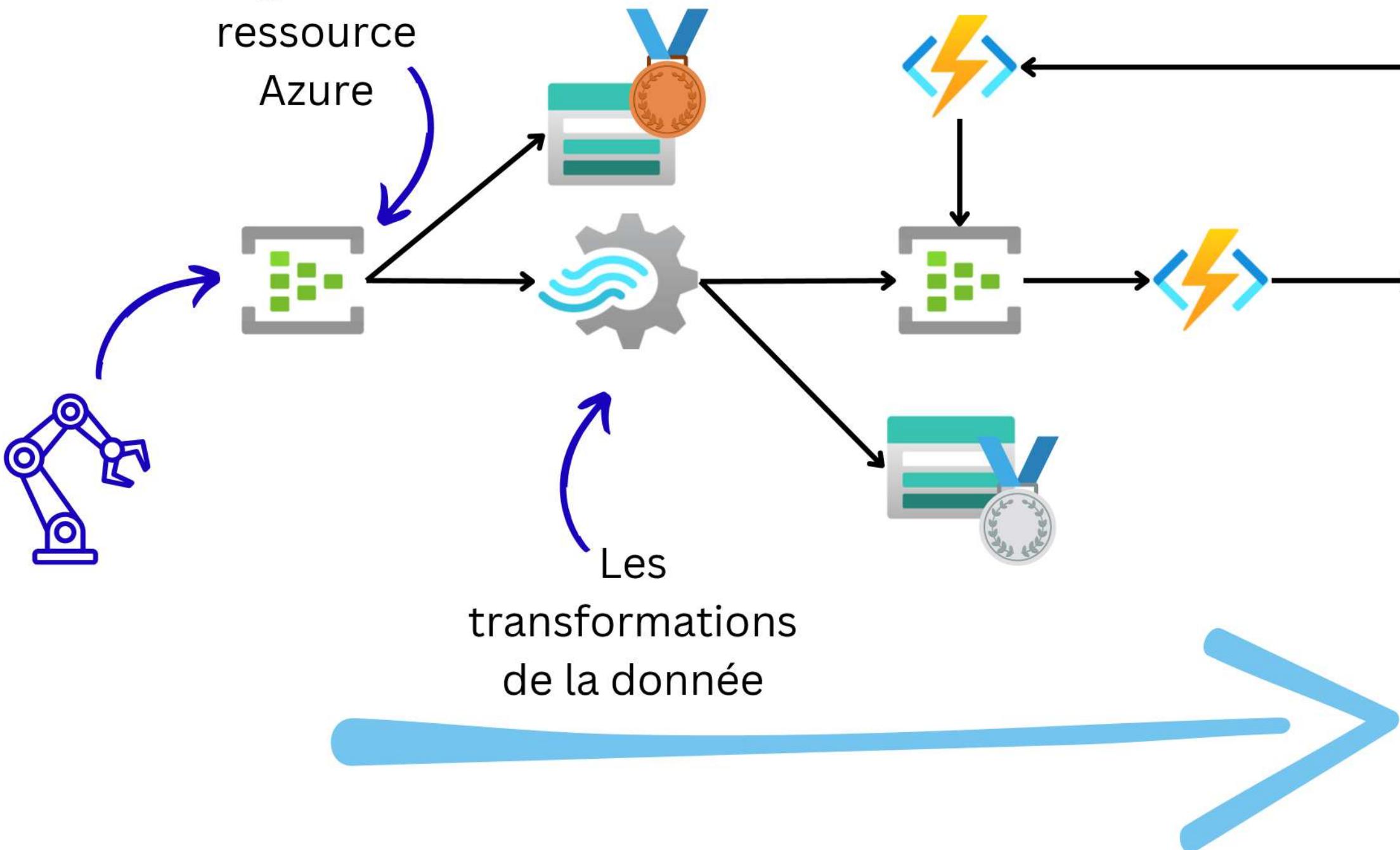
Pour
optimiser
l'affichage
sur le
dashboard

Du Kafka
déguisé en
ressource

Azure

Les
transformations
de la donnée

Pour
optimiser
l'affichage
sur le
dashboard

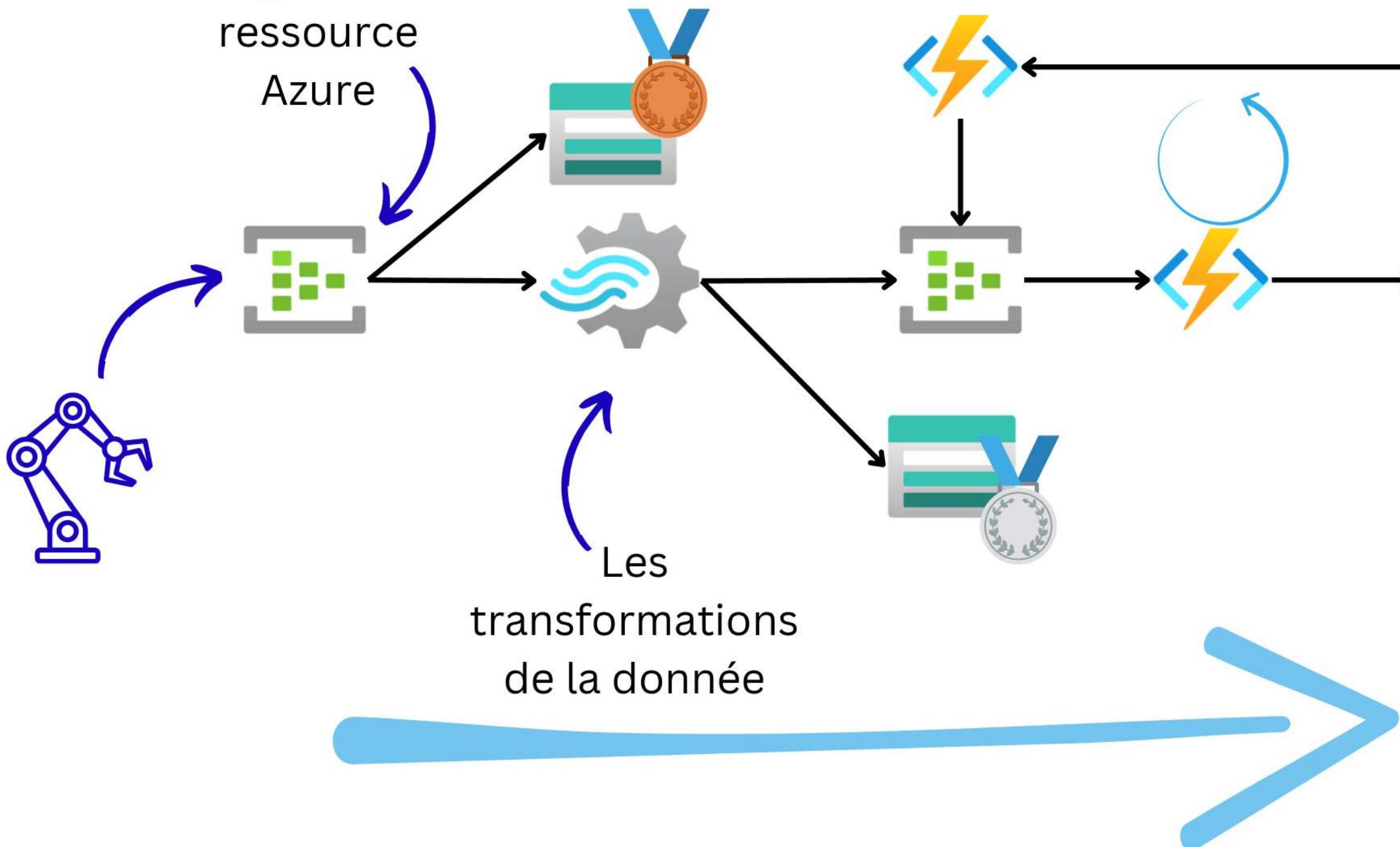


Du Kafka
déguisé en
ressource

Azure

Les
transformations
de la donnée

Pour
optimiser
l'affichage
sur le
dashboard

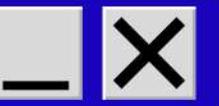




**Et ils vécurent
heureux
Le banquet de fin**

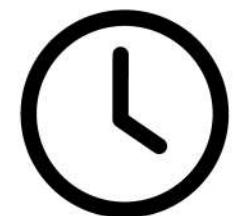


Retour à la modernité

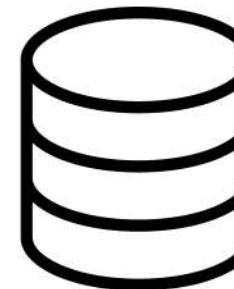




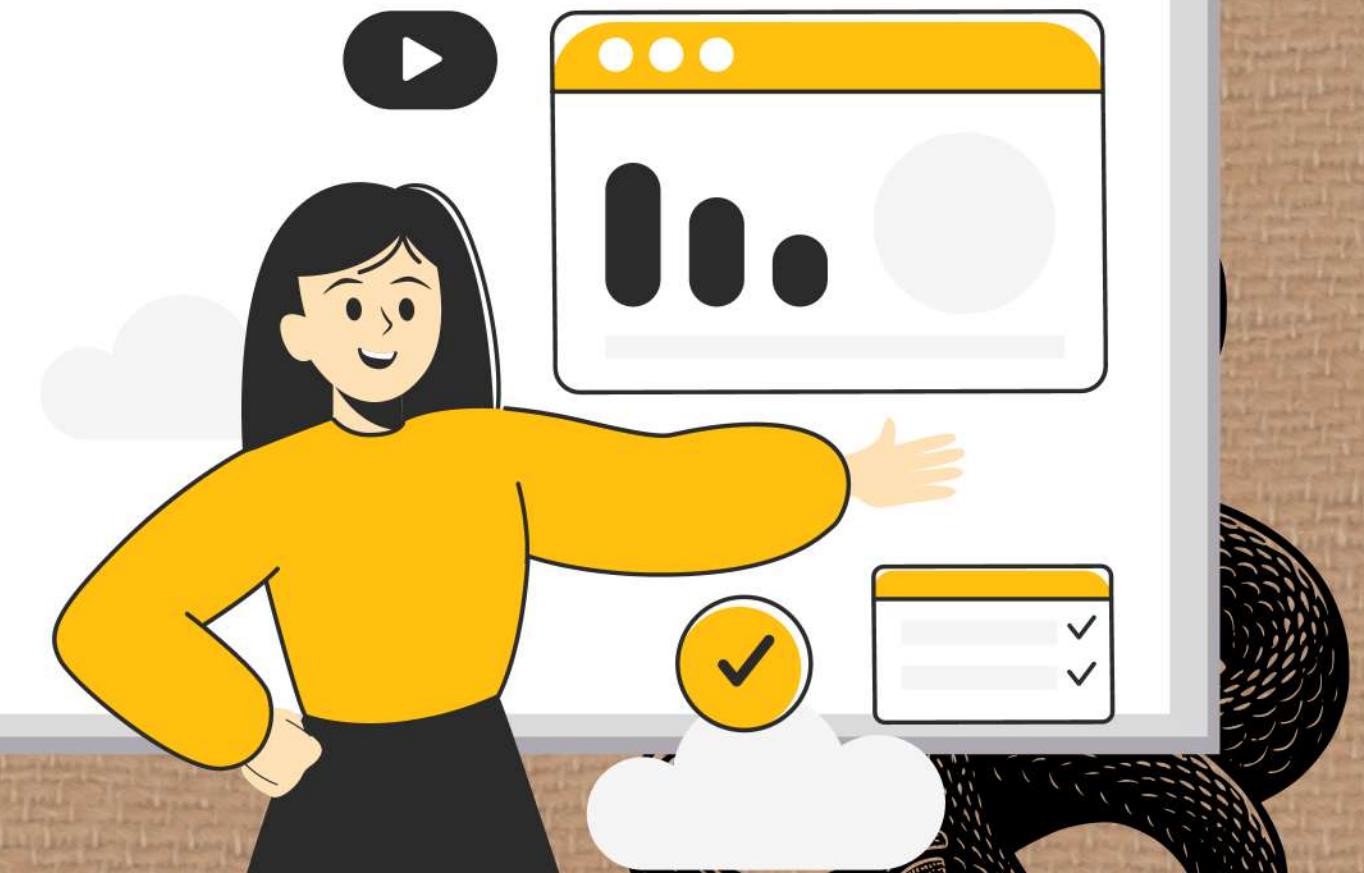
Des coûts divisés par 2 pour chaque environnement



Un temps de traitement qui passe de 45min à moins d'une minute



Un meilleur accès aux données, que ce soit pour de la manipulation ou du débogage



Ce qu'on retient

Analyse du besoin :

stream ou batch

Partir sur le bon outil



Affichage en temps réel = streaming

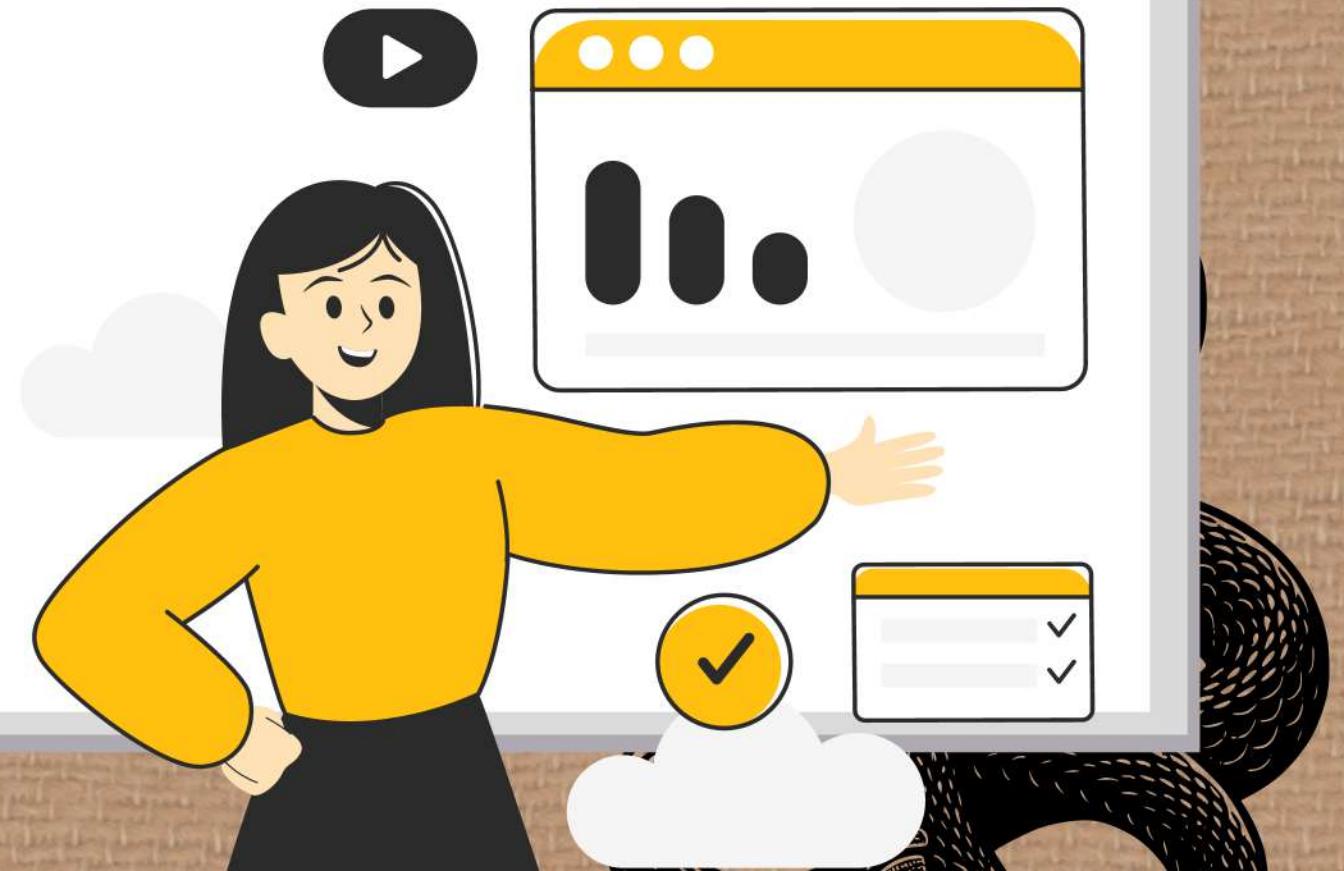


Analyse de données à posteriori = batch



Stateful/stateless

Bien ranger les données





Fin.



Hann

igui

m

nba

muff

am

Fin.

