



RT R3.10



A wide-angle photograph of a modern conference room or auditorium. In the foreground, rows of dark blue chairs are arranged facing a stage. On the stage, there is a white podium positioned in front of a large, dark rectangular screen. Above the stage, a complex rig of stage lights is visible, with one light beam directed towards the screen. The room has large windows on the right side, letting in natural light. The ceiling is white with some structural elements.

Prenez place

Veo

Qui parle ?

1. David LACAN
2. Diplômé en Informatique et mathématiques
3. Responsable de l'OF et CFA ARES FORMATION à Roanne
4. Intervenant professionnel depuis 12 ans à l'IUT et au CUR
5. Spécialités : informatique (développement et gestion) et mathématiques
6. david.lacan@univ-st-etienne.fr

De quoi on parle ?

R3.10 Gestion d'un système de bases de données

Description

La ressource lui permet d'acquérir les notions suivantes :

- L'installation et la configuration d'un système de gestion de base de données,
- La gestion des comptes utilisateurs du service et leur accès à l'interface d'administration
- La gestion des droits d'accès aux données par les utilisateurs,
- La sécurité des données et des connexions au service,
- La sauvegarde et la restauration.

La ressource peut aussi aborder les stratégies et méthodes de conception (qu'il s'agisse du schéma de la BDD ou de la structure du réseau pour y intégrer un service de BDD).

La data et le monde qui l'entoure

Introduction

Un monde VUCA

Volatilité



Complexité



Monde
VUCA

Incertitude



Ambiguïté



Introduction

Un monde VUCA

Comment s'adapter à un monde VUCA ?

S'adapter en fonction des feedbacks et du bilan de vos actions

Se focaliser sur la valeur pour atteindre l'objectif



Avoir un objectif à long terme mais des actions à court terme

Travailler en mode itératif et incrémental

Software As A service

Modèle économique actuel : mode SaaS



SaaS

Data center

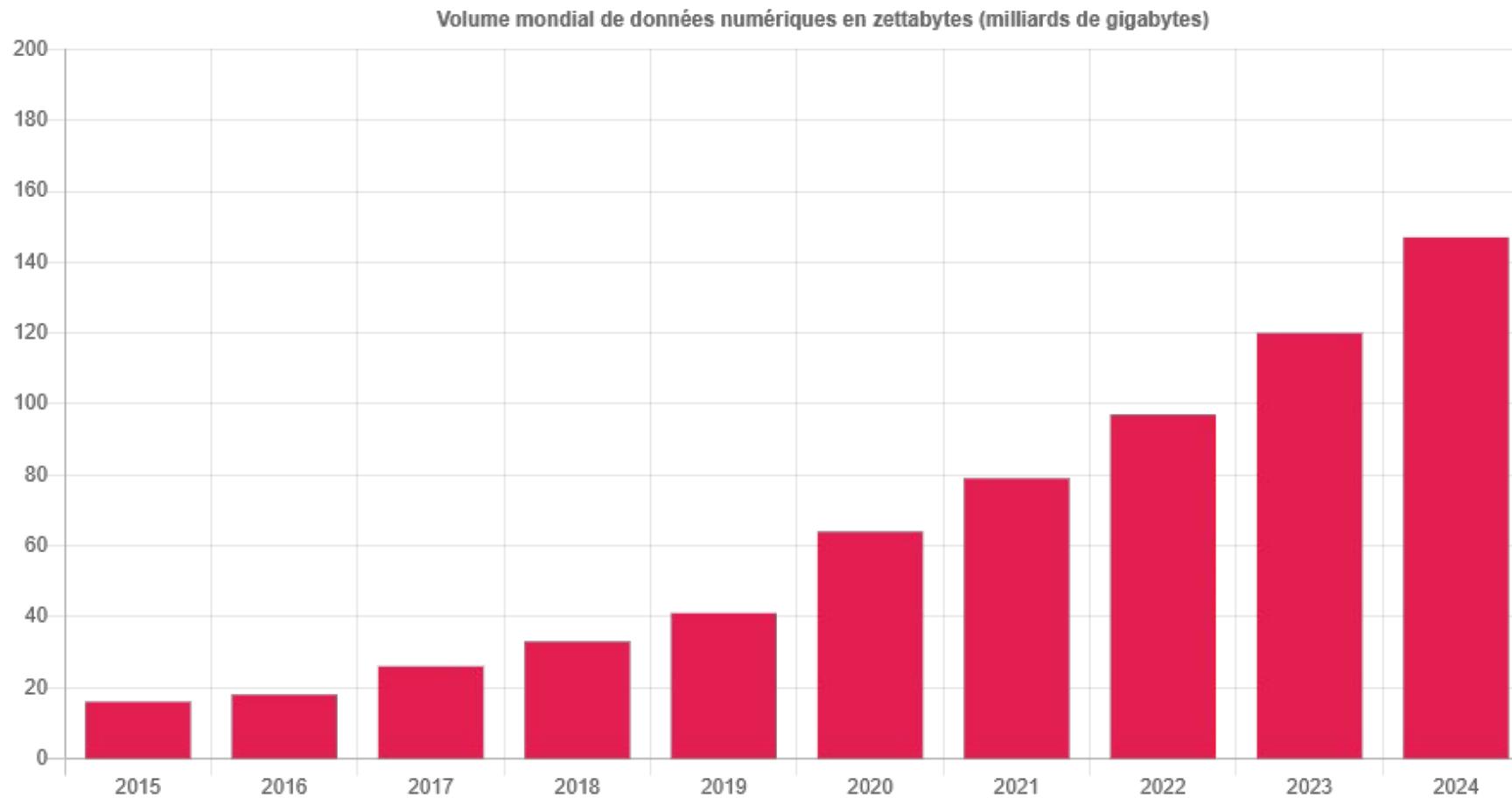
Le mode SaaS génère du big data

Les centres de données et leurs corollaires (centres de diffusion des contenus, centres d'élimination des attaques) représentent la dimension matérielle, physique, de nos sociétés du numérique, fondées sur la massification de l'usage des données. Leur localisation est stratégique mais ils ne représentent pas une manne d'emploi pour les territoires où ils sont implantés.



SaaS

Data center



Evolution des hardwares et des data

Loi	Domaine	Principe	Effet
Moore	Puissance de calcul	Les processeurs doublent en puissance tous les ~2 ans	Calcul plus rapide
Kryder	Stockage magnétique	La densité de stockage double rapidement	Plus de capacité, moins cher
Butter	Stockage flash	Vitesse et capacité flash augmentent vite	SSD de plus en plus rapides
Hilbert (ou Data Growth)	Données numériques	Le volume mondial de données double tous les 2 ans	Explosion du big data

enjeux économiques

Importance pour l'entreprise

- 1) Gain de compétitivité
- 2) Prise de décision



Provenance des données

Internes

opérations quotidiennes de l'entreprise



- Données ventes
- Données financières
- Données des salariés
- Données de production

Externes

Environnement



- Données de marché
- Données démographiques
- Réseaux sociaux
- ...

Type des données

Structurées

Non Structurées

SaaS

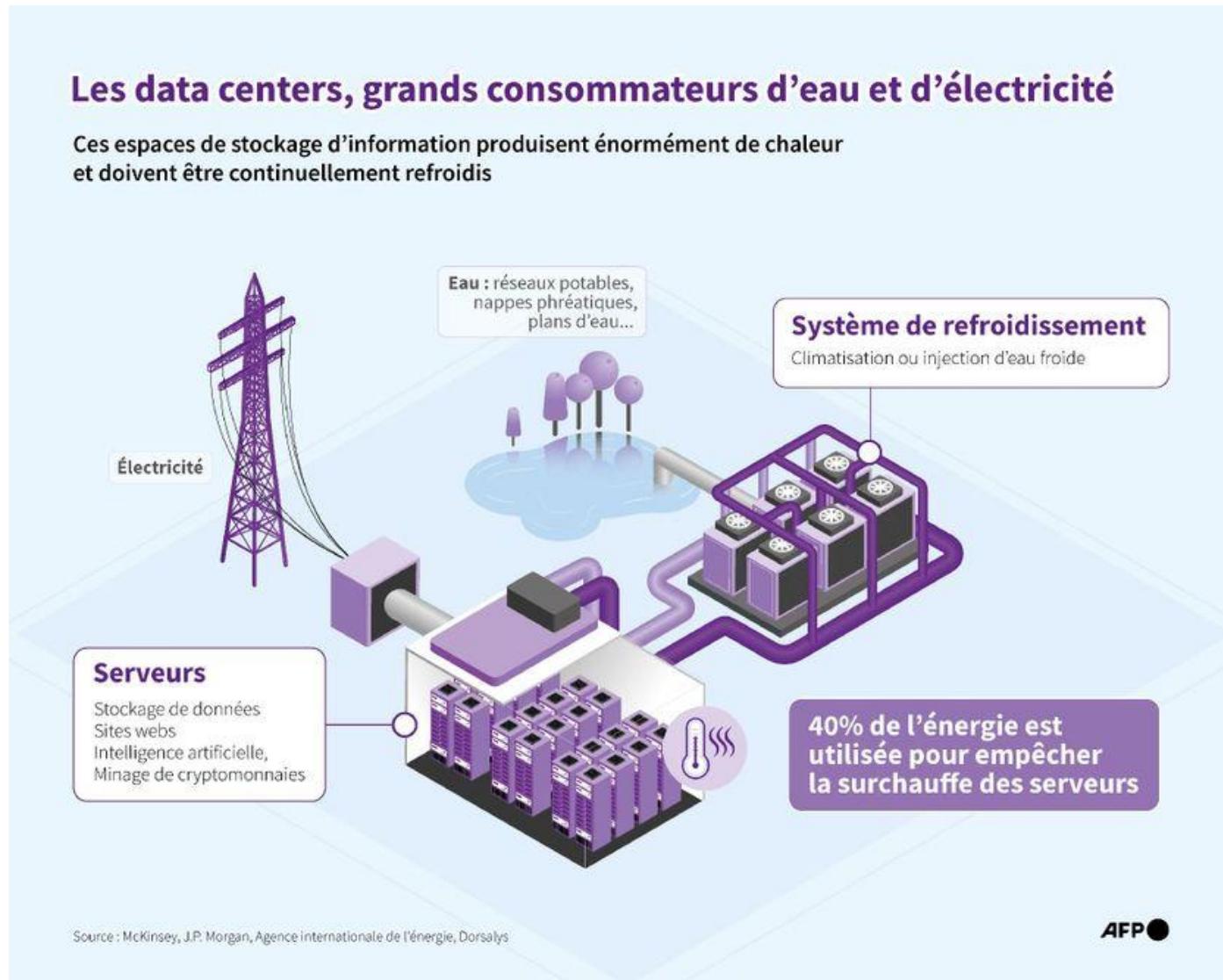
Data center

Document 9. La domination mondiale des marchés des centres de données par les GAFAM

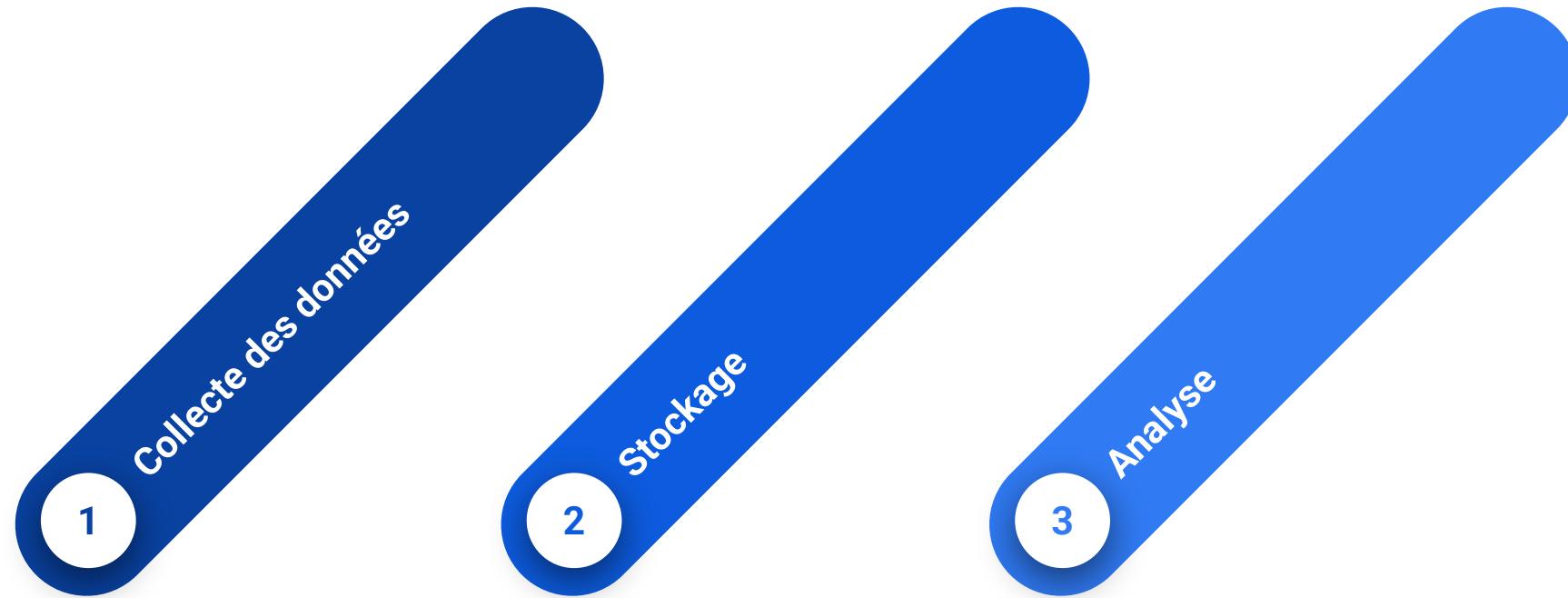
Rang	Monde	États-Unis	Chine	Reste de l'Asie	Europe	Reste du monde
1	Amazon	Amazon	<i>Alibaba</i>	Amazon	Amazon	Amazon
2	Microsoft	Microsoft	<i>Tencent</i>	Microsoft	Microsoft	Microsoft
3	Google	Google	<i>China Telecom</i>	Google	Google	Google
4	<i>Alibaba</i>	Oracle	<i>Huawei</i>	<i>NTT</i>	Oracle	Salesforce
5	Oracle	Salesforce	<i>China Unicom</i>	<i>Alibaba</i>	Salesforce	Oracle
6	Salesforce	IBM	<i>China Mobile</i>	<i>Fujitsu</i>	IBM	IBM

En **gras** : entreprise étatsunienne. En *italique* : non étatsunienne. Source : Synergy Research group, 2024.

Data center



Processus de Gestion



Stockage des données

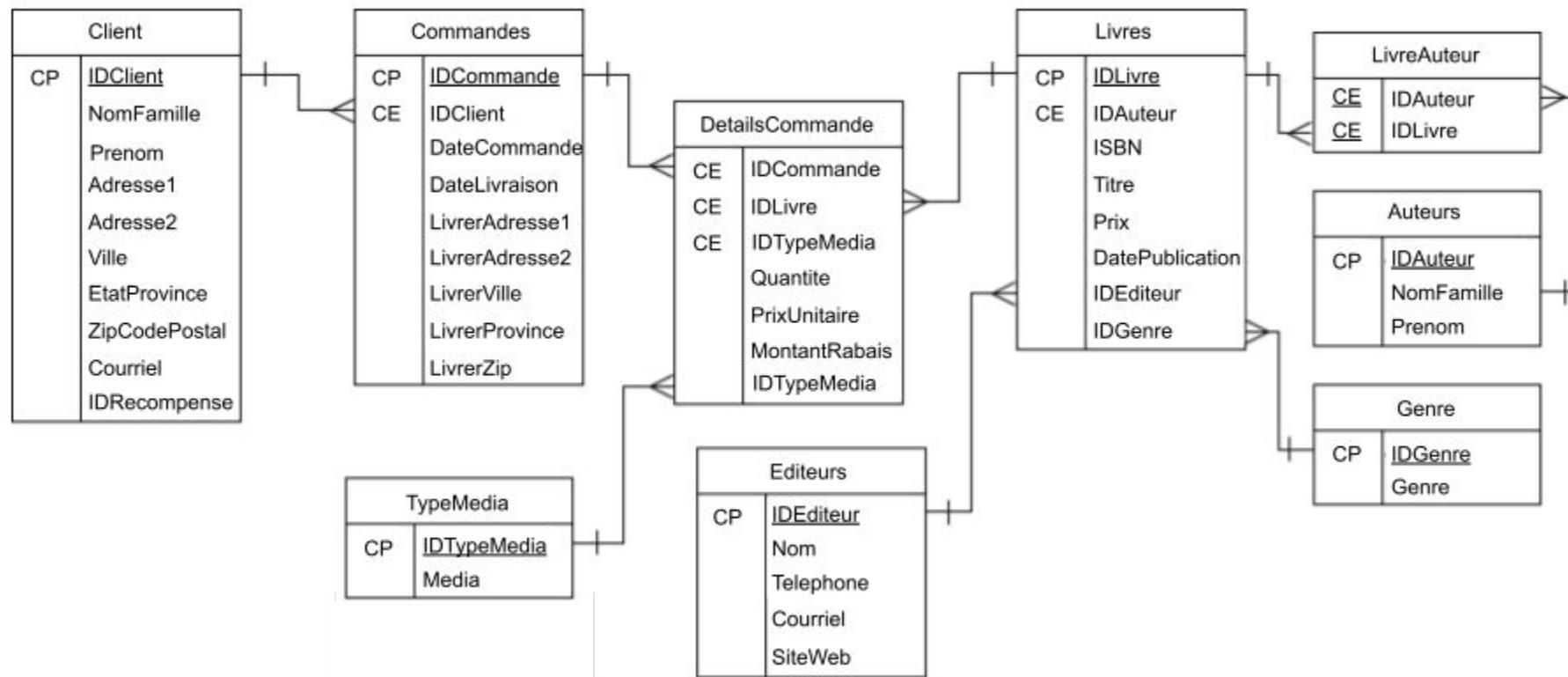
SYSTÈME DE GESTION DE BASE DE DONNÉES

Une base de données est un ensemble organisé et structuré d'informations (ou de données) stockées de manière à pouvoir être consultées, modifiées et gérées facilement.

SGBD : fournisseurs

1. Oracle. Oracle Database est un système de gestion de **base de données relationnelle** et **relationnel-objet**, créé par Oracle dans les années 70.
2. Mysql. ...
3. PostgreSQL. ...
4. Microsoft SQL Server. ...
5. MongoDB. ...
6. Redis. ...
7. Elasticsearch. ...
8. SQLite.

SGBD relationnel



SGBD : clé primaire et clé étrangère

Clé primaire (Primary Key)

Une **clé primaire** est un **identifiant unique** pour chaque ligne d'une table.

Elle permet de **distinguer chaque enregistrement** de manière certaine.

→ Caractéristiques :

- Unique (aucune ligne ne peut avoir la même clé primaire)
- Non vide (ne peut pas être NULL)
- Sert souvent à faire le lien avec d'autres tables

Clé étrangère (Foreign Key)

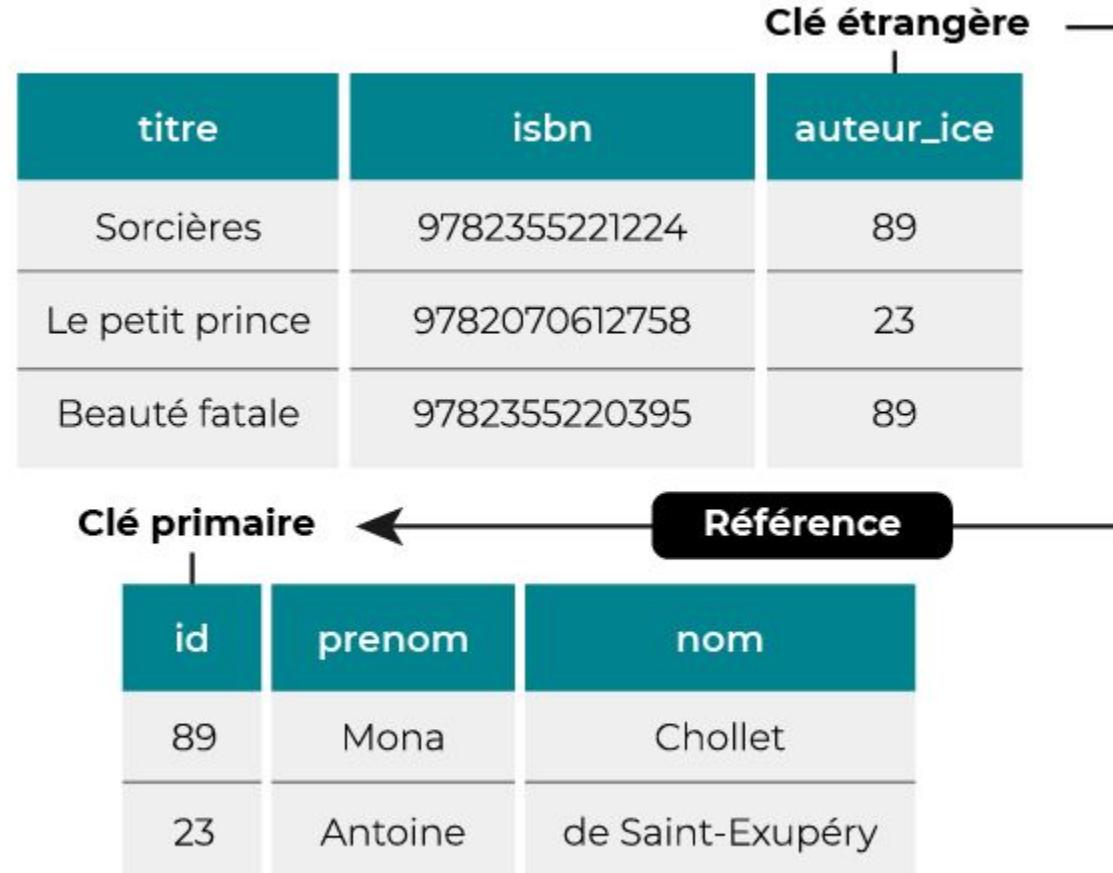
Une **clé étrangère** est un **champ qui fait référence à la clé primaire d'une autre table**.

Elle permet de **créer un lien (relation)** entre deux tables.

→ Caractéristiques :

- Elle établit une **relation** entre deux tables.
- Elle assure la **cohérence** des données (on ne peut pas faire référence à un enregistrement qui n'existe pas).

SGBD : clé primaire et clé étrangère



SGBD : importance des types de données

L'importance du typage des données dans une base de données

Le **typage des données** consiste à **définir la nature (ou le type)** de chaque information stockée dans une base de données. Chaque colonne d'une table a un **type de données précis** : texte, nombre, date, booléen, etc.

- Assurer la cohérence des données
- Optimiser les performances
- Économiser de l'espace
- Garantir l'intégrité des données
- Faciliter le traitement et les calculs

SGBD : ACID

A – Atomicité (Atomicity)

Une transaction doit être **tout ou rien**.

→ Si une partie échoue, **toutes les opérations sont annulées**.

Cela empêche la base d'être laissée dans un état incohérent.

I – Isolation (Isolation)

Les transactions simultanées **ne doivent pas interférer** entre elles.

→ Chaque transaction s'exécute **comme si elle était seule** dans la base.

Cela évite les problèmes d'accès concurrents.

C – Cohérence (Consistency)

Une transaction doit **préserver la validité** des données.

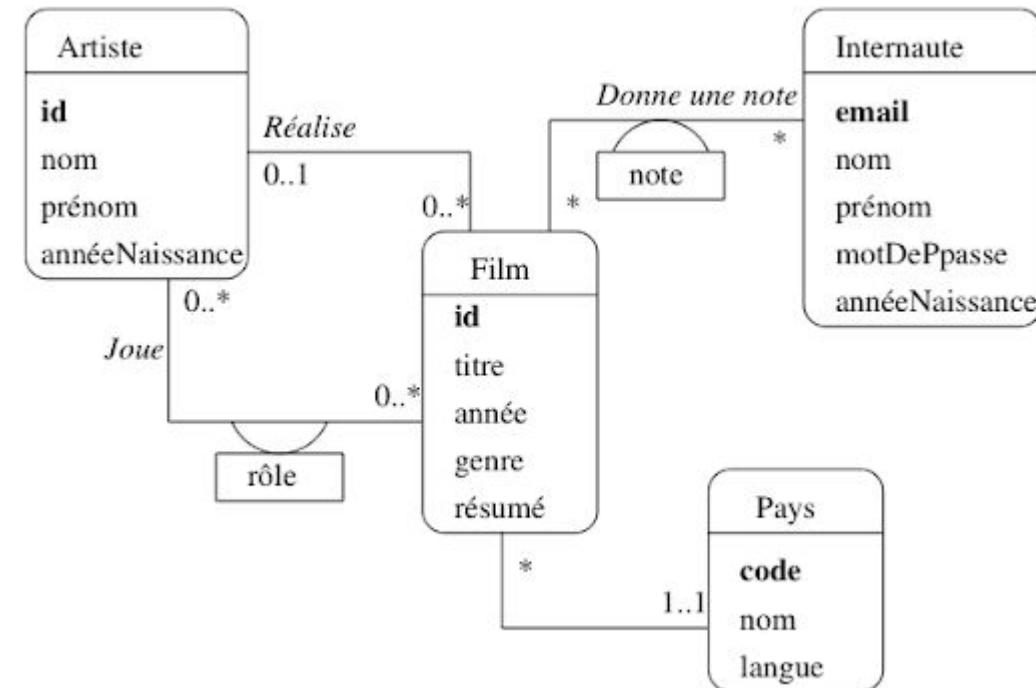
→ Les règles de la base (contraintes, types, clés étrangères, etc.) doivent **toujours être respectées**.

D – Durabilité (Durability)

Une fois validée, une transaction est **définitivement enregistrée**, même en cas de panne.

Les types de relation

- | | |
|------------------------|------------|
| 1) One To One | 1:1 |
| 2) One To Many | 1:n |
| 3) Many To One | n:1 |
| 4) Many to Many | n:m |



Interrogation d'un SGBD

Le Langage SQL

C'est le **langage utilisé pour communiquer avec une base de données relationnelle**.

Objectifs du SQL

- **Créer** des bases de données et des tables
- **Ajouter, modifier, supprimer** et lire des données
- **Interroger** la base pour en extraire des informations précises
- **Gérer** les droits des utilisateurs

Le **SQL** est le **langage universel des bases de données relationnelles**.

Le Langage SQL

Clause	Rôle	Exemple
SELECT	Permet d'exprimer et d'exécuter des requêtes ensemblistes sur la base de données	<code>SELECT essai_comment FROM essai WHERE ref_essai = 8419 ;</code>
INSERT INTO	Permet d'ajouter des lignes ou des tuples dans une table de la base de données	<code>INSERT INTO essai ('ref_essai', 'essai_libelle') VALUES (12, 'ajout de ligne') ;</code>
DELETE FROM	Supprime des lignes dans une table de la base de données	<code>DELETE FROM essai WHERE essai_libelle CONTAINS '%lbc' ;</code>
UPDATE	Met à jour des lignes existantes dans une table de la base de données	<code>UPDATE essai SET essai_libelle = 'texte du jeu' WHERE ref_essai = 12 ;</code>
<Window function> OVER	Permet d'effectuer une opération de fenêtrage. Disponible uniquement à partir de la norme SQL – ISO 2003.	<code>SUM(montant) OVER (PARTITION BY region, categorie, ORDER BY date_vente DESC)</code>

ARES FORMATION

Outils

Outils de BD



BD : Maria DB ou Mysql

accessible en ligne de commande

Interface : phpmyadmin

Outils de BD



1) simulation de serveur (émulateur) :

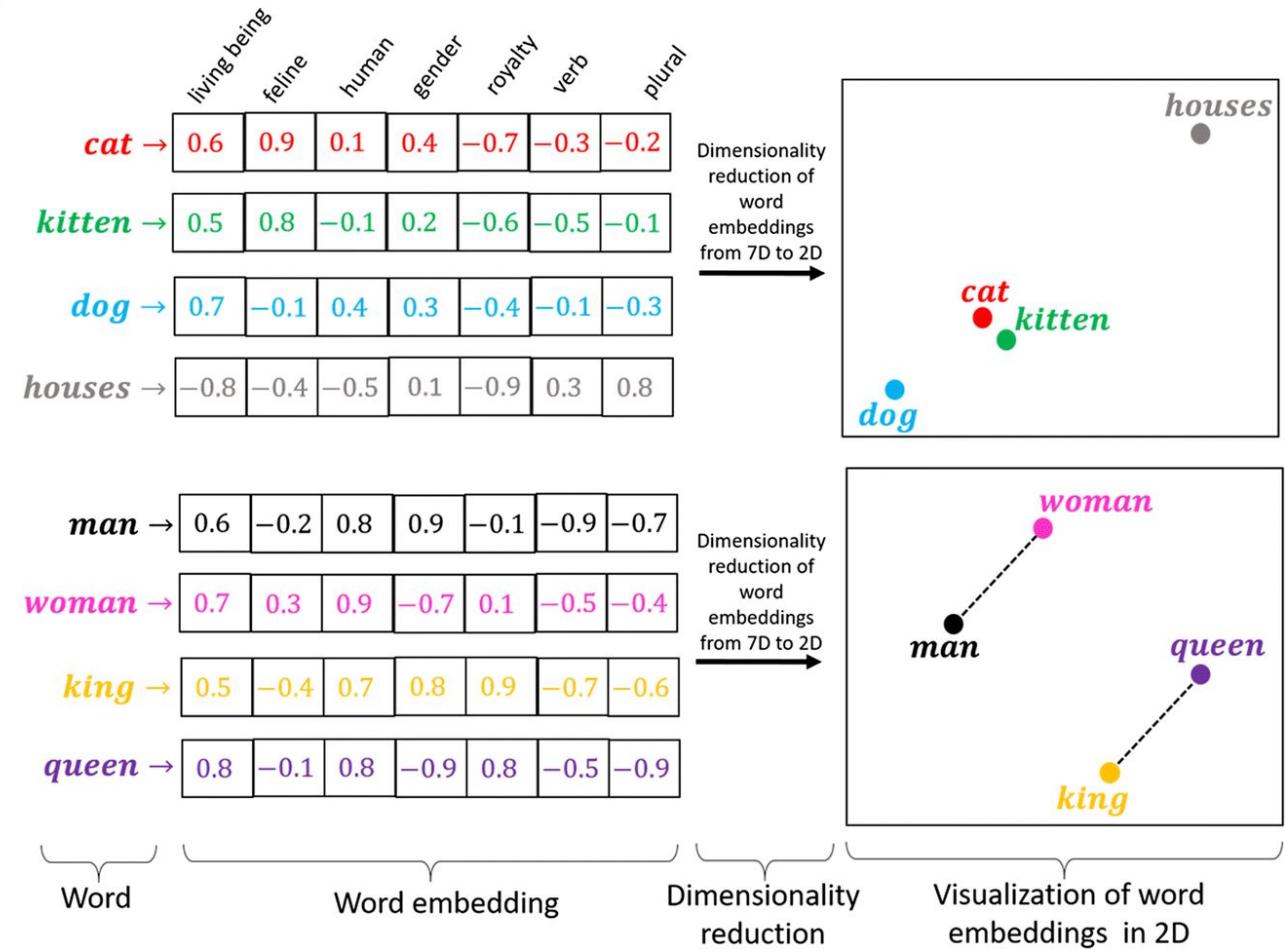
- **easypHP**
- **wamp, mamp**
- **laragon**

2) créer un serveur en VM

- **debian**
- **apache**
- **php**
- **mysql, mariaDB**
- **phpmyadmin**

Les bases de données vectorielles

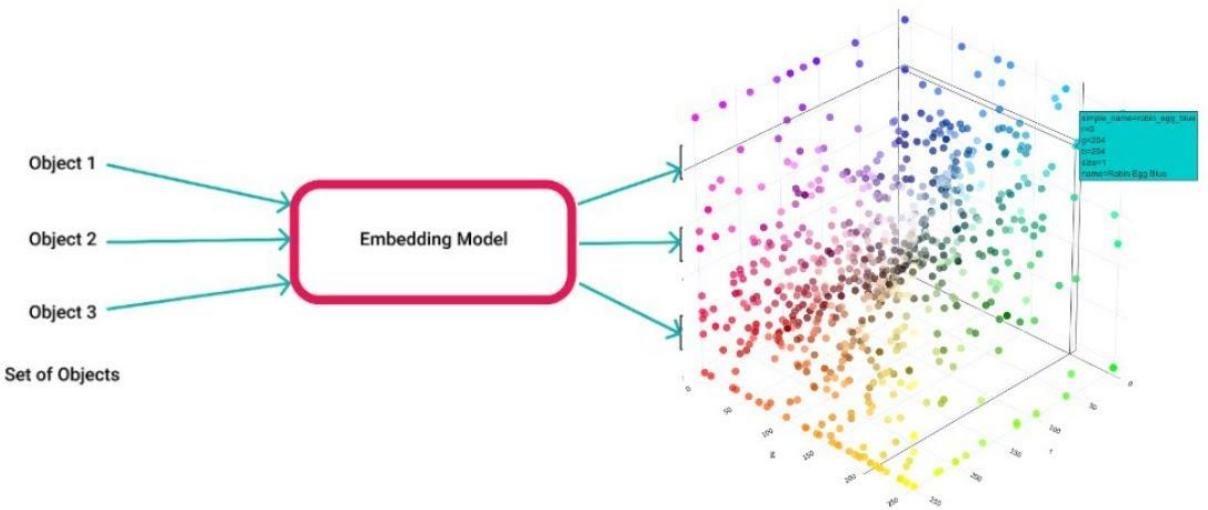
La notion de vecteur



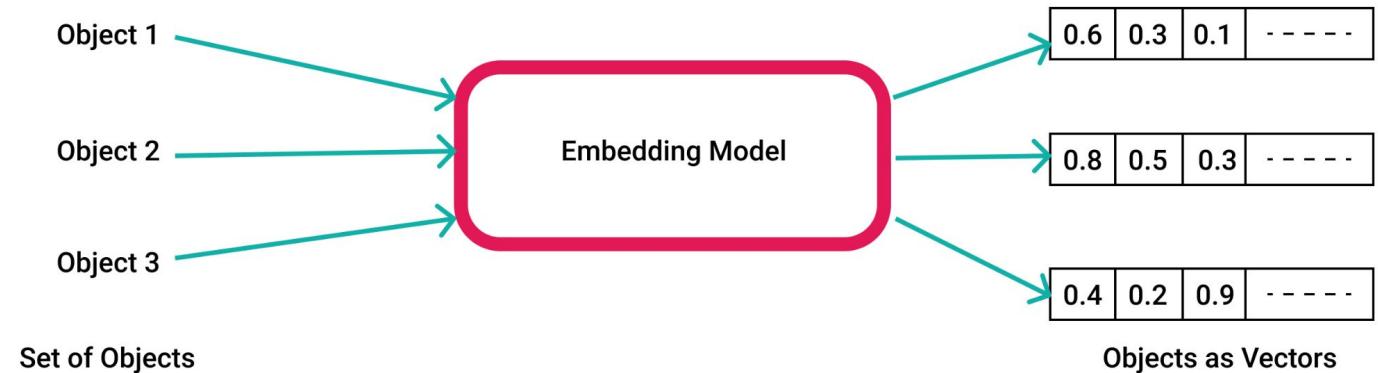
La notion de vecteur



INTRODUCTION TO VECTOR EMBEDDINGS

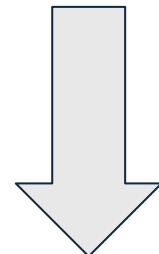


La notion de vecteur



Pourquoi je vous parle
de cela ?

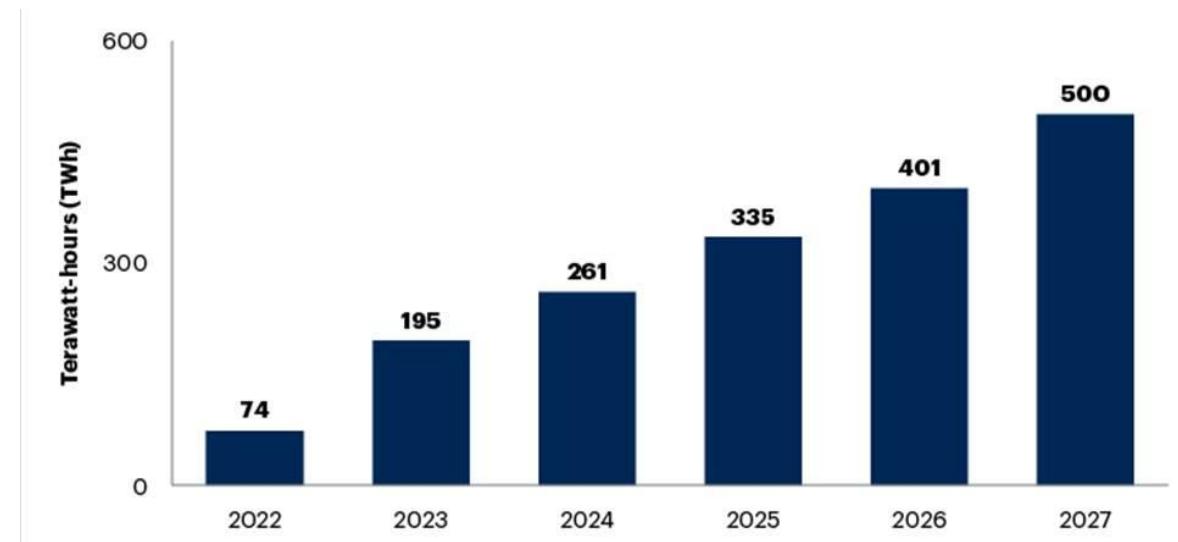
Pourquoi on parle de cela ?



Quel est le coût d'utilisation de l'IA ?

Petite parenthèse électrique

Les datas centers sont constitués de différents éléments pour fonctionner : serveurs, câbles, connexion internet, pare-feu... Et certains éléments d'un data center sont particulièrement énergivores. C'est notamment le cas des serveurs informatiques et des systèmes de refroidissement. Près de **80 % de la consommation d'énergie d'un data center** provient de ces deux éléments.



Petite parenthèse électrique

La **consommation d'énergie d'un data center** dépend avant tout de la taille de la structure. Sites électro intensifs, ils consomment en moyenne 5,15 MWh/m²/an en France. Ainsi, un data center de 10 000 m² de surface présente une consommation équivalente à celle d'une ville de 50 000 habitants, comme Laval ou la Roche-sur-Yon, par exemple.

La notion de token



Soit la phrase :

Je suis heureux de participer à cette formation. Elle est géniale en tout point.

La notion de token

	Tokens	Characters
GPT-4o & GPT-4o mini	17	80

Je suis heureux de participer à cette formation. Elle est géniale en tout point.

	Tokens	Characters
GPT-3,5 & GPT-4	19	80

Je suis heureux de participer à cette formation. Elle est géniale en tout point.

	Tokens	Characters
GPT-3 legacy	26	80

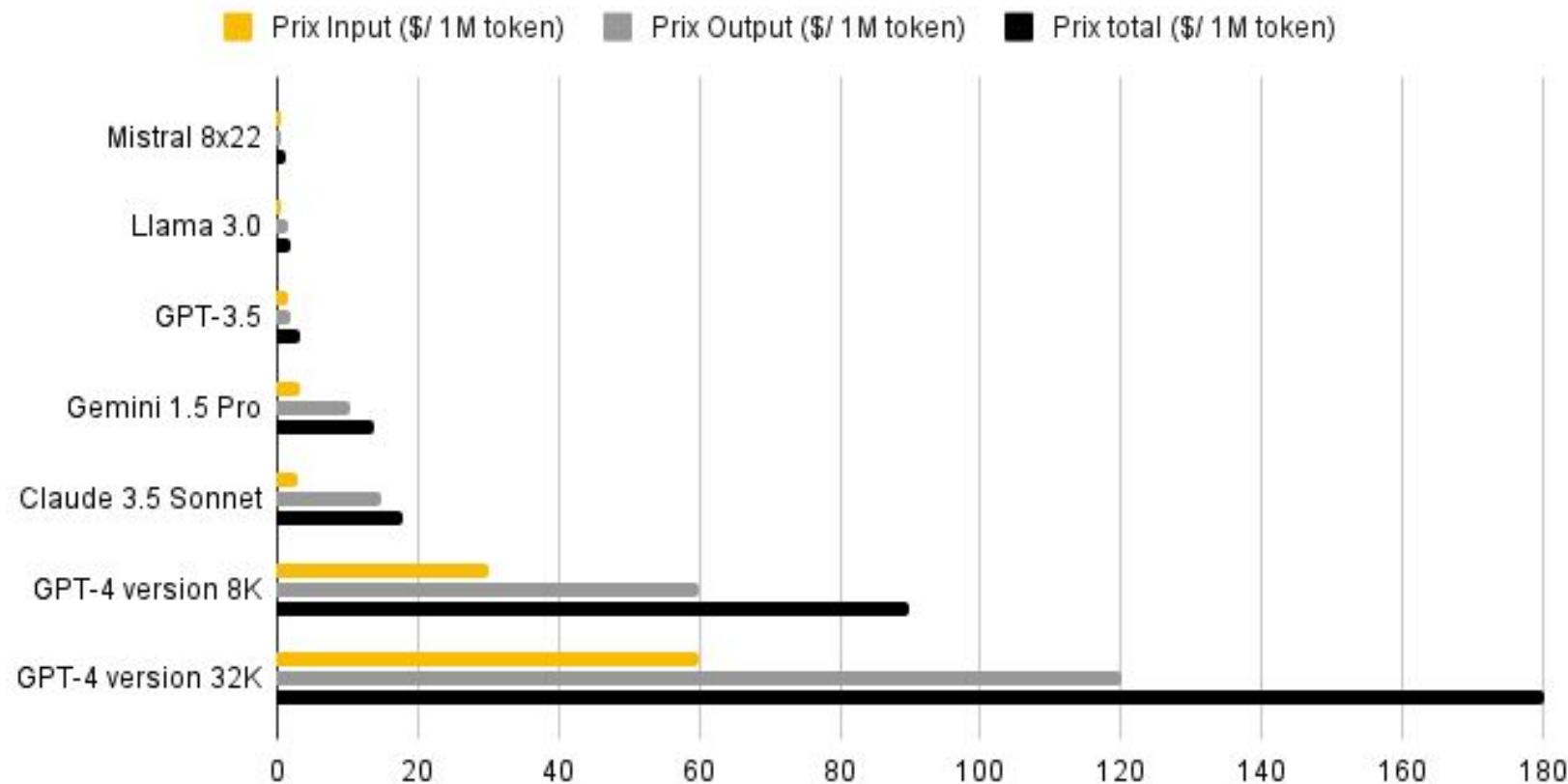
Je suis heureux de participer à cette formation. Elle est géniale en tout point.

Quel est le coût d'utilisation de l'IA

- Utilisation non-payante
- Utilisation avec abonnement
- Utilisation en API en fonction du nombre de token entrée / sortie

Quel est le coût d'utilisation de l'IA

Coût des différents LLM pour 1 M de tokens (en dollars)





Quel est le coût d'utilisation de l'IA

Model	8B		70B		405B	
	Input	Output	Input	Output	Input	Output
AWS	\$0.22	\$0.22	\$0.99	\$0.99	\$5.32	\$16.00
Azure	\$0.30	\$0.61	\$2.68	\$3.54	\$5.33	\$16.00
Databricks	-	-	\$1.00	\$3.00	\$5.00	\$15.00
Fireworks.ai	\$0.20	\$0.20	\$0.90	\$0.90	\$3.00	\$3.00
IBM	\$0.60	\$0.60	\$1.80	\$1.80	\$5.00	\$16.00
Octo.ai	\$0.15	\$0.15	\$0.90	\$0.90	\$3.00	\$9.00
Snowflake	\$0.57	\$0.57	\$3.63	\$3.63	\$9.00	\$9.00
Together.AI	\$0.18	\$0.18	\$0.88	\$0.88	\$5.00	\$15.00

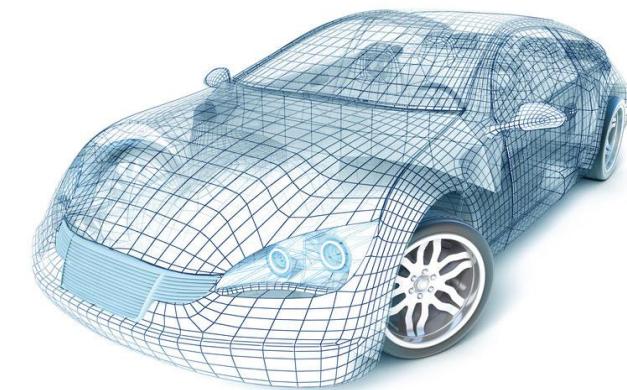
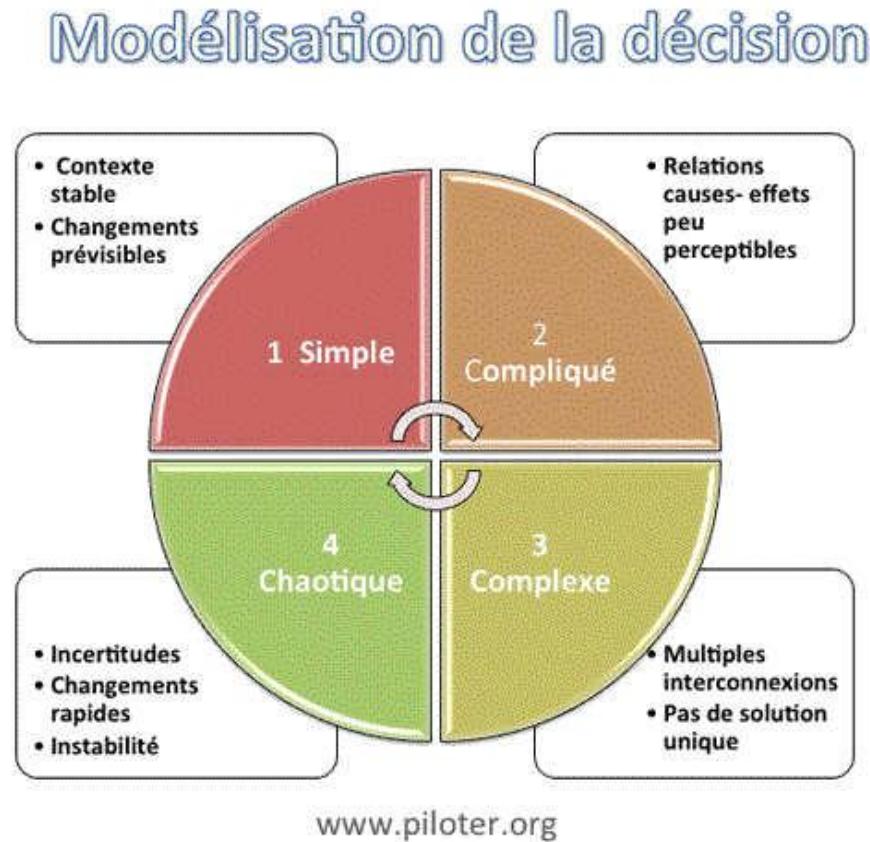
Que va t-on faire
réellement ?

Mise en place d'un SGBD

- Mettre en place l'environnement
- Prise de connaissance d'une étude de cas
- Modélisation
- Implémentation
- Interrogation
- Mise en place des outils de sauvegarde et de restauration

Modélisation

La modélisation est le processus de création d'une représentation abstraite d'un système réel ou d'un concept, souvent dans le but de mieux comprendre, analyser, simuler ou prédire son comportement.



ARES FORMATION

<https://github.com/david-alpha/R310--2025>

Support

ARES FORMATION

Conclusion

Merci et qu'en avez vous pensé ?

