

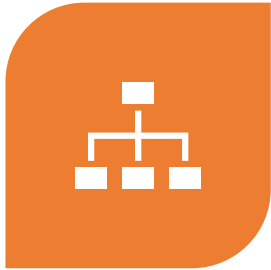
# SQL Avancé

---

Kimba FONDIO & Aurel ATTERE

# Plan

---



ORGANISATION DES  
DONNÉES



CONCEPTION D'UNE  
BASE DE DONNÉES



LANGAGE DE  
DÉFINITION DES  
DONNÉES



LANGAGE DE  
MANIPULATION DES  
DONNÉES

# Organisation des données

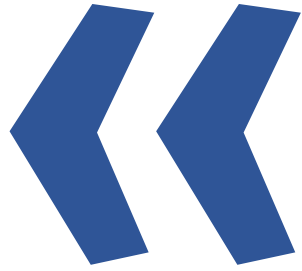
---

Historique des bases de données

Modèle relationnel

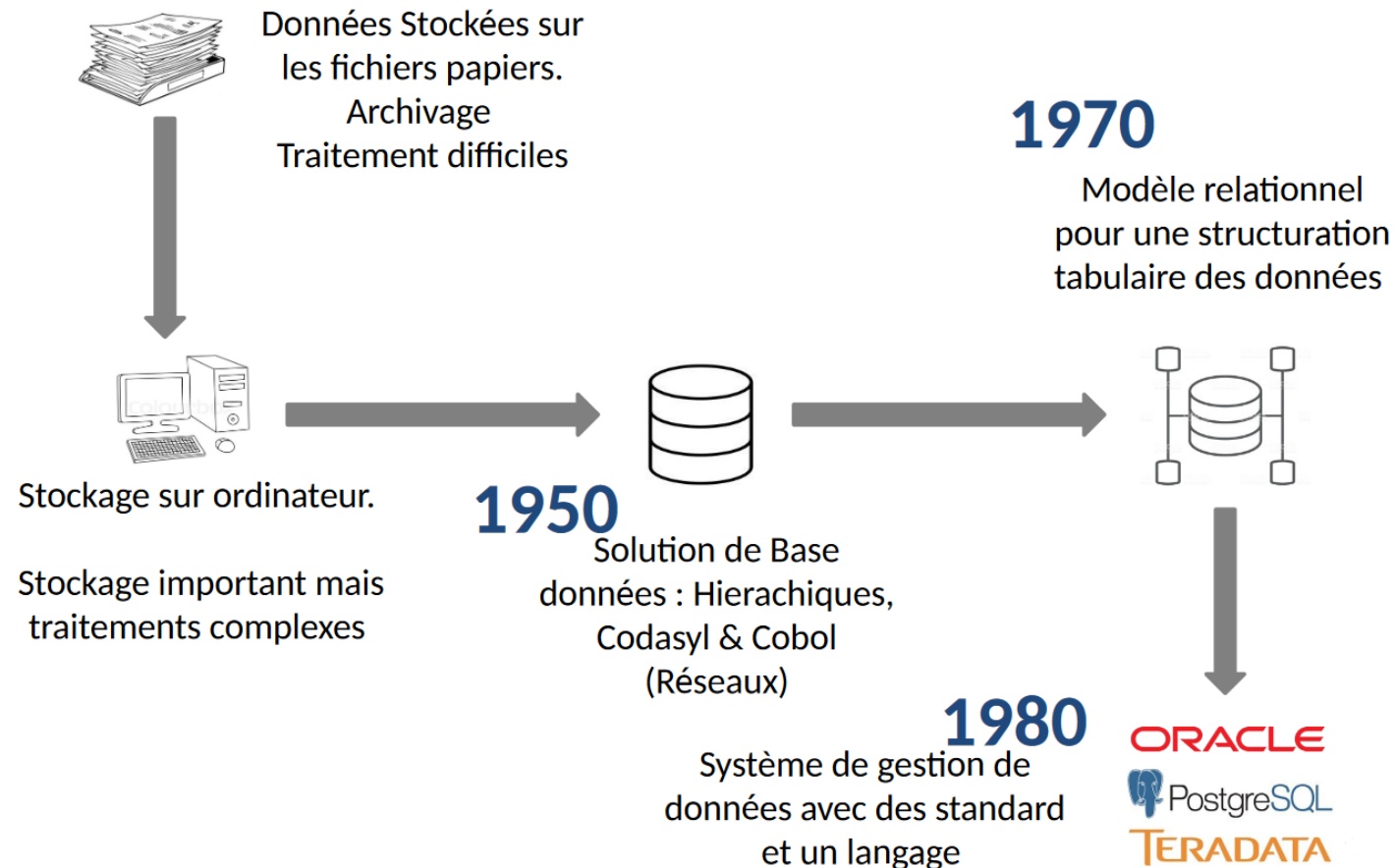
Modèle orienté agrégat

Modèle orienté graphe



Au sens large, la donnée se présente, entre autres, sous forme de données factuelles, de calculs, de texte, de documents, d'images stockés sur un ordinateur.

# Historique des bases de données(1)



# Historique des bases de données(2)

---

Années 70	Années 80	Années 90	Années 2000
<ul style="list-style-type: none"><li>• apparition des premiers SGBD<ul style="list-style-type: none"><li>• séparation de la description des données de la manipulation de celles-ci par les applications</li></ul></li><li>• modèles hiérarchique et réseau CODASYL</li><li>• langages d'accès navigationnels</li><li>• SGBD IDMS, IDS 2 et IMS 2</li></ul>	<ul style="list-style-type: none"><li>• modèle relationnel</li><li>• les SGBDR représentent l'essentiel du marché BD (aujourd'hui)</li><li>• architecture répartie client-serveur</li></ul>	<ul style="list-style-type: none"><li>• modèles de données plus riches<ul style="list-style-type: none"><li>• systèmes à objets</li><li>• OBJECTSTORE, O2</li></ul></li><li>• Base de données distribuées</li></ul>	<ul style="list-style-type: none"><li>• Datawarehouse</li><li>• NoSQL</li><li>• NewSQL</li><li>• Big Data</li><li>• Cloud</li></ul>

# Modèle relationnel : Définition

---

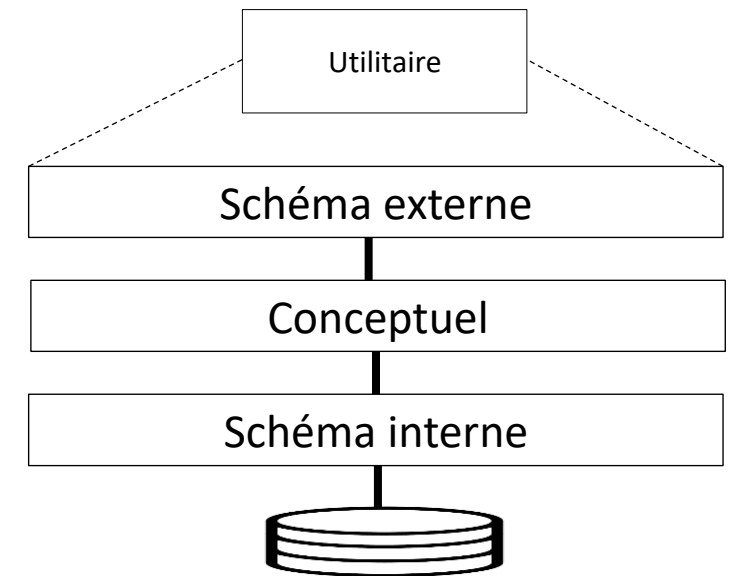
- Base de données : Ensemble structuré de données apparentées qui modélisent un univers réel. Une BD est faite pour enregistrer des faits, des opérations au sein d'un organisme (administration, banque, université, hôpital, ...).
- Une base de données est un ensemble d'informations qui est organisé de manière à être facilement accessible, géré et mis à jour. Elle est utilisée par les organisations comme méthode de stockage, de gestion et de récupération de l'informations.

De façon plus large on dira que : Les bases de données sont conçues pour stocker des données structurées, semi-structurées ou non structurées, et elles peuvent contenir des informations telles que des textes, des chiffres, des images, des vidéos, des enregistrements audio, etc. Elles sont utilisées pour gérer de grandes quantités de données et permettre l'accès rapide et efficace aux informations nécessaires.

Ici on parlera essentiellement des base de données relationnelle, donc d'information structurées.

# Modèle relationnel : SGBDR

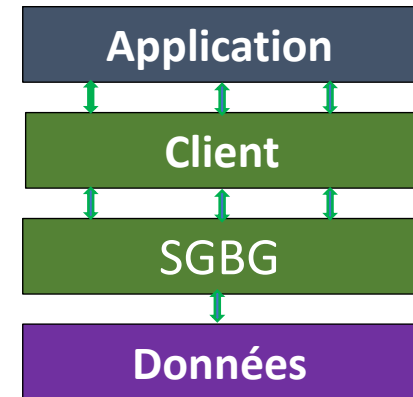
- Les bases de données sont généralement gérées par des Systèmes de Gestion de Bases de Données (SGBD) ou Systèmes de Gestion de Bases de Données Relationnelle (SGBDR), qui sont des logiciels spécifiques conçus pour créer, organiser, stocker, interroger et gérer les données d'une base de données. Les exemples de SGBD couramment utilisés incluent
  - MySQL,
  - Oracle,
  - Microsoft SQL Server,
  - PostgreSQL ...
- On part de fichiers : Les données des fichiers sont décrites dans les programmes ; aux bases de données : Les données de la base de données sont décrites hors des programmes dans la base elle-même. Cette description est faite suivant un modèle bien définie : Le modèle relationnel.
- Il s'agit d'un Modèle de **stockage de données** caractérisé par :
  - Structure Tabulaire des données
  - Les contraintes et les règles de gestion
  - Les Operations et langage de manipulation
  - Séparation logique et physique





# Modèle relationnel : Objectifs du SGBDR

- Un gestionnaire des **transactions** avec les SGBDR
  - Une suite d'opérations sur une ou plusieurs BD
  - Lien SGDBR et application : Client – Serveur
- Le SGBDR doit permettre de :
  - **Décrire les données :**
    - Indépendamment des applications (de manière intrinsèque)
    - Langage de définition des données
    - **DATA DEFINITION LANGUAGE (DDL)**
  - **Manipuler les données :**
    - Interroger et mettre à jour les données sans préciser d'algorithme d'accès
    - Dire QUOI sans dire COMMENT
    - Langage de manipulation des données
    - **DATA MANIPULATION LANGUAGE (DML)**
  - **Contrôler les données :**
    - Intégrité : Vérification de contraintes d'intégrité ex: le salaire doit être compris entre 400F et 20000F
    - Confidentialité
    - Contrôle des droits d'accès, autorisation
    - Langage de contrôle des données
    - **DATA CONTROL LANGUAGE (DCL)**



# Modèle relationnel : Succès du SGBDR

---

- **Partage** : Une BD est partagée entre plusieurs utilisateurs en même temps
  - Contrôle des accès concurrents
  - Notion de transaction
  - L'exécution d'une transaction doit préserver la cohérence de la BD
- **Structuration forte des données** : existences des règles de cohérences et des contraintes d'intégrité référentielle
- **Découplage logique & physique** : Indépendances entre modèle de données, stockage et manipulation
- **Existence de SQL**: langage déclaratif simple au fondement mathématique solide avec des opérations algébrique (insert, drop, select, join, union, intersect....)
- **Objet de manipulation** : Vue, procédure, triggers
- **Requêtes simple et efficace** avec un plan d'optimisation
- **Maturité et Stabilité** avec une documentation et un accompagnement
- **Cohérence** transactionnelle
- **Sécurité** : reprise après panne, journalisation
- **Performance d'accès** : index (hashage, arbres balancés ...)

# Modèle relationnel : Propriétés ACID

- **Atomicité**: exécution et/ ou annulation en bloc des opérations d'une transaction
- **Cohérence** : les contraintes d'intégrité de la base sont respectées
- **Isolation** : toute transaction doit s'exécuter comme si elle était la seule sur le système. Aucune dépendance possible entre les transactions.
- **Durabilité** : lorsqu'une transaction a été confirmée, elle demeure enregistrée même à la suite d'une panne d'électricité, d'une panne de l'ordinateur ou d'un autre problème.



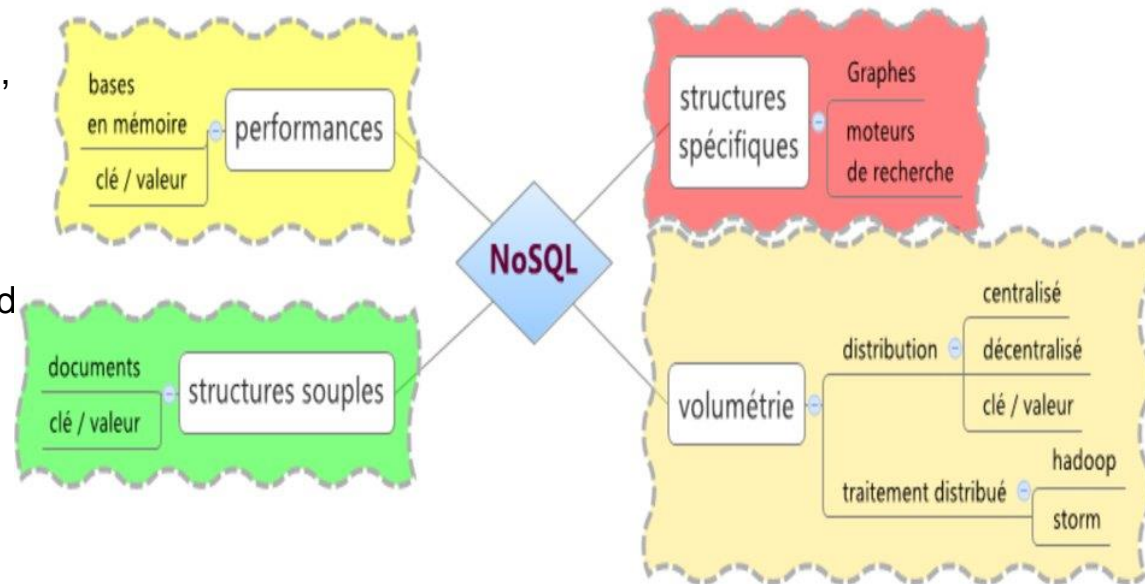
# Modèle relationnel : Limites du relationnel

---

- **L'ère du Big Data : La révolution du Web et les tendances**
- **Flexibilité**
  - Modéliser les données d'un utilisateur de Facebook avec une BDR
- **Expressivité**
  - Trouver les amis d'un utilisateur
  - Les amis des amis d'un utilisateur
  - Les amies des amis d'un utilisateur
- **Passage à l'échelle**
  - Définir les niveaux d'isolation pour garantir la cohérence et la disponibilité des données
  - Taille fixe des serveurs, serveurs propriétaire et coûteux
  - Traitement centralisé et beaucoup de lecture/écritures sur disque
- **Données transitoires et non structurées**
  - Mécanismes pour stocker des données non permanentes
  - Sensor, applications web 2.0 à BIG DATA

# Autres Modèles : NewSQL - NoSQL

- **NoSQL** : Le NoSQL regroupe de nombreuses bases de données, récentes pour la plupart, qui se caractérisent par une logique de représentation de données non relationnelle et qui n'offrent donc pas une interface de requêtes en SQL
- **NewSQL** : Vise à combiner les avantages des bases de données relationnelles traditionnelles (SQL) avec les nouvelles approches de gestion des données
- **ACID vs BASE**
  - Basically Available
  - Soft state
  - Eventually Consistent



# Modèle orienté agrégat : clé valeur

---

- **Principe**

- Stocke les informations sous forme de paires clé-valeur dans lesquelles une clé sert d'identifiant unique : Simple et efficace
- La valeur associée à une clé peut être une simple chaîne de caractère comme un document, ou encore un objet beaucoup plus complexe pouvant contenir une multitude d'information.
  - **Num : 070895026**
  - **Adresse : Abidjan**
  - **Rue : H151**

- **Avantages**

- Les clés et les valeurs peuvent se présenter sous plusieurs formes, textes, objets simples ou objets composés
- hautement divisibles et permettent une mise à l'échelle horizontale à des échelles que d'autres types de bases de données ne peuvent pas atteindre.

- **Inconvénients**

- Pas de langage standard ou structuré (ex. SQL) pour accéder aux données
- Contraintes d'intégrité à gérer au niveau applicatif ...

- **Cas d'utilisation**

- Applications manipulant des collections d'informations
- Index primaire sur la clé, absence de clé secondaire
- Impossibilité d'accès via le contenu.

# Modèle orienté agrégat : document

- **Principe**

- Stocke et manipule des données de structures complexes
- Similaire au modèle clé-valeur
- Structure arborescente
  - Un champ peut être structuré en sous champs
- Langages d'interrogations riches comme dans les BD classiques
- Index secondaires
- Sérialisation
  - XML / JSON

- **Avantages:**

- faible besoin de jointures
- Lecture rapide si document entier sollicité
- Besoin faible de transactions

- **Cas d'utilisation**

- bibliothèques numériques, collections de produits,
- gestion des historiques d'utilisateurs sur réseaux sociaux



```
{
  "_id": "5cf0029caff5056591b0ce7d",
  "firstname": "Jane",
  "lastname": "Wu",
  "address": {
    "street": "1 Circle Rd",
    "city": "Los Angeles",
    "state": "CA",
    "zip": "90404"
  },
  "hobbies": ["surfing", "coding"]
}
```

## Inconvénients

- Imbrication nécessite beaucoup d'itérations pour trouver une information
- Perte d'autonomie des entités
- Redondance des informations

# Modèle orienté agrégat : colonne

- **Principe**

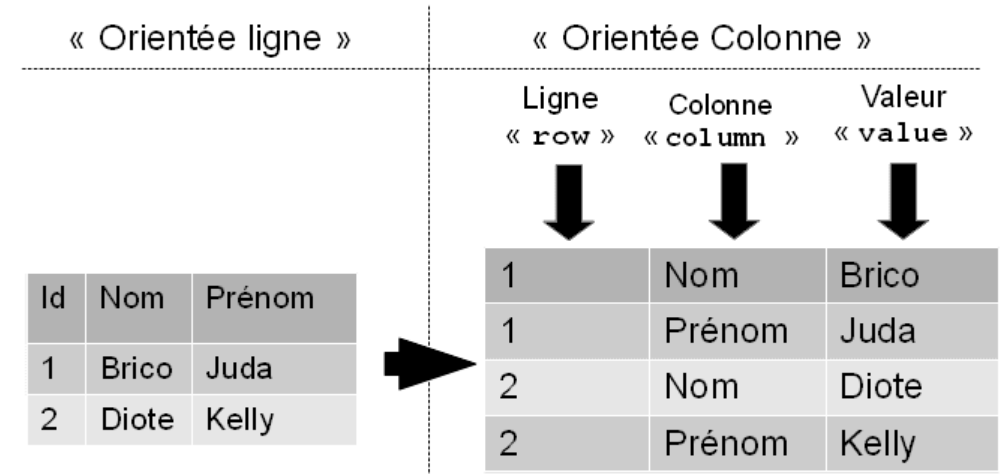
- Structure “hybride”
  - Clé valeur à 2 dimensions
  - Stockage par rowkey identique au modèle document
- Structuration
  - Colonne : couple clé/valeur associé à une version
  - Super colonne: imbrication de colonne
  - Rowkey : identifiant unique d'une ensemble de colonne
  - Keyspace : ensemble de familles de colonnes

- **Avantages:**

- Colonnes dynamiques
  - valeurs non atomiques et diversité des types
  - valeur complexes structurées
- Facilité de distribution des données et de la gestion de leur historique

- **Cas d'utilisation**

- e-commerce, gestion de catalogue, gestion des commandes
- Gestion des logs, Google Analytics



## Inconvénients

- Modèle flou : termes parfois déroutant
- Difficulté des opérations de jointures



# Modèle orienté graphe

- **Principe**

- Exploiter la structure d'un graphe pour stocker et manipuler les données
- Stocker de grande masse données et les interactions (corrélations) entre elles
- permet d'analyser des informations interconnectées, de comprendre, d'évaluer et d'exploiter les processus et les connexions présentes entre les données

- **Avantages:**

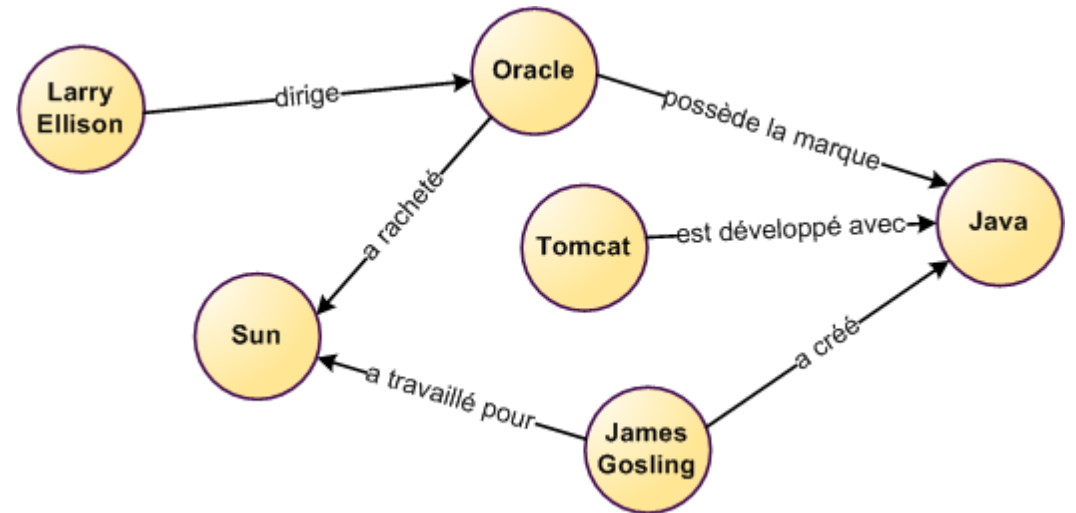
- Opérations de la théorie du graphe
- Requêtes basées sur la gestion de chemins

- **Inconvénients :**

- Distribution non triviale
- Complexité de la gestion des graphes multiplex

- **Cas d'utilisation**

- Réseaux sociaux
- Routes
- Réseaux électriques
- Web social





Merci

