

Les Bases de données

« NoSQL »

Part 1

Présenté par

Marie Esther BOUA
Data Scientist



Qu'est ce que une base de données?

Pour quel besoin spécifiques les bases de données ont-elles été créées?

Qu'est ce qu'un SGBD?

Qu'est ce que le modèle relationnel?

Qu'est ce qu'un SGBDR?

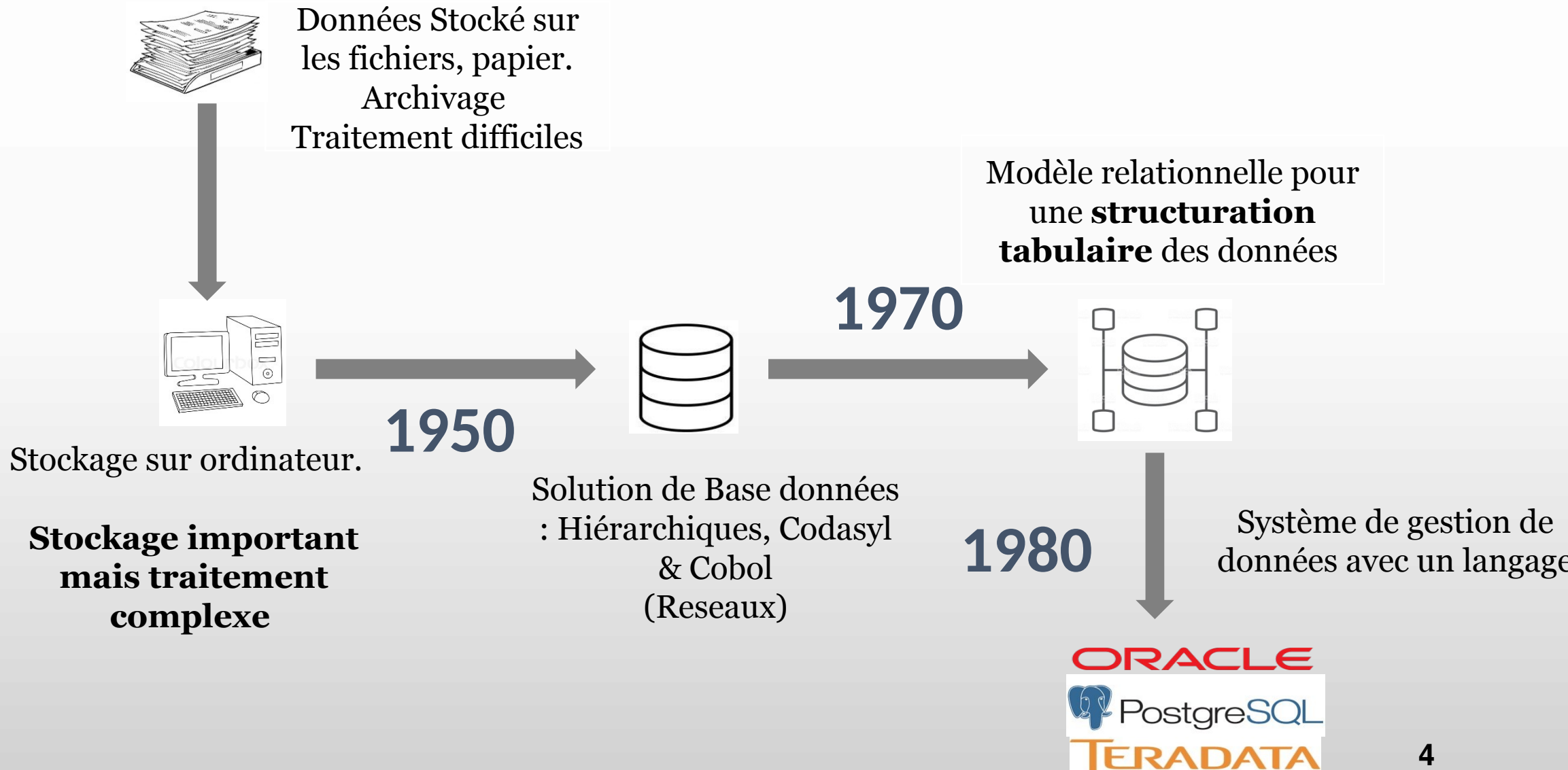
Citez 5 SGBDR connus.



Caractéristiques SGBD

- Indépendance physique
- Indépendance logique
- Manipulabilité
- Rapidité des accès
- Administration centralisée
- Limitation de la redondance
- Vérification de l'intégrité
- Partageable de la donnée
- Sécurité des données

Historique SGBDR

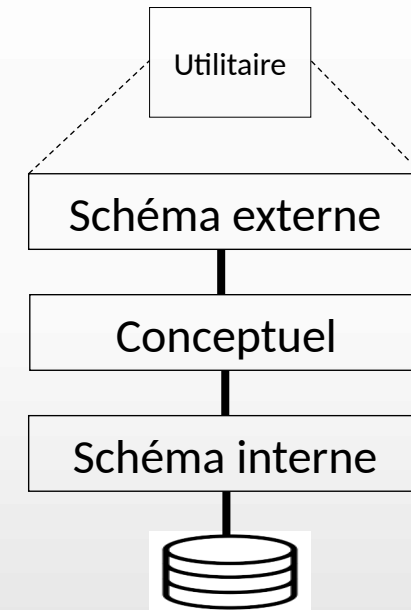




Modèle relationnel

Modèle de stockage de données caractérisé par :

- Structure Tabulaire des données
- contraintes et règles de gestion
- Operations et langage de manipulation
- Séparation logique et physique



SGBDR = SGBD basé sur le modèle relationnel



Caractéristiques SGBDR

- ❖ Accès simple via un langage de requêtes SQL
- ❖ Indépendance entre stockage physique et vision logique des données
- ❖ Existence de contraintes d'intégrité
- ❖ Forte cohérence transactionnelle



Caractéristiques SGBDR

❖ Propriétés ACID

- **Atomicité:** exécution et/ ou annulation en bloc des opérations d'une transaction
- **Cohérence :** les contraintes d'intégrité de la base sont respectées



Caractéristiques SGBDR

❖ Propriétés ACID

- **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.



L'ère du Big Data

1,85 milliard d'utilisateurs
Facebook actifs/ jour

1,78 Milliards de page
web

100 Milliards de message
Whatsapp / jour

5,5 milliards de
recherches *Google*/ jour

293 milliards d'emails
envoyés / jour (hors
Spam)

44 000 Milliards de Go
en 2020 (90% des
données en 10 ans)

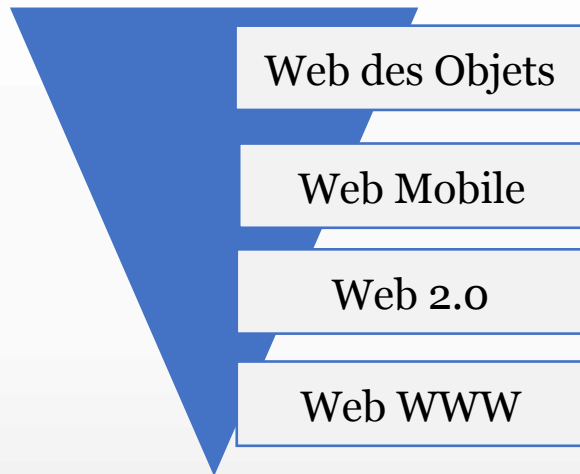
1 milliard d'heures de
vidéos *Youtube*/ jour

210 millions d'utilisateurs
actifs chaque jour sur
Snapchat





La révolution du Web et les tendances



**1 carte Sim par
personne dans
le monde**

**60 % d'utilisation
d'internet par le mobile**

**80 milliards
d'objets connectés**

**1 500 Milliards \$
généré par le e-
commerce**

**2 internautes sur 3
sont sur les réseaux
sociaux**



Une véritable montée en charge : défis de stockage, accessibilité et temps de réponse



Les nouveaux besoins

Les 6 facteurs clés du **SPRAIN** :

- **Scalability (Scalabilité)** : Assurer la montée en charge
- **Performance** : Gestion rapide d'objets complexes et hétérogènes dynamique
- **Relaxed consistence (consistance relâchée)**: une forte disponibilité (au profit de la cohérence)



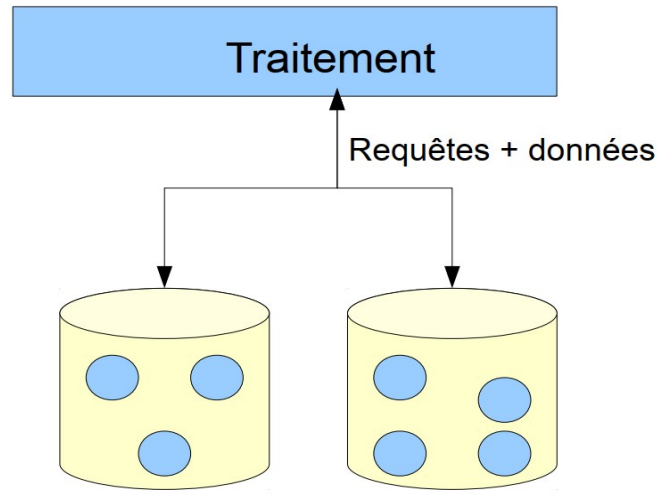
Les nouveaux besoins

Les 6 facteurs clés du **SPRAIN** :

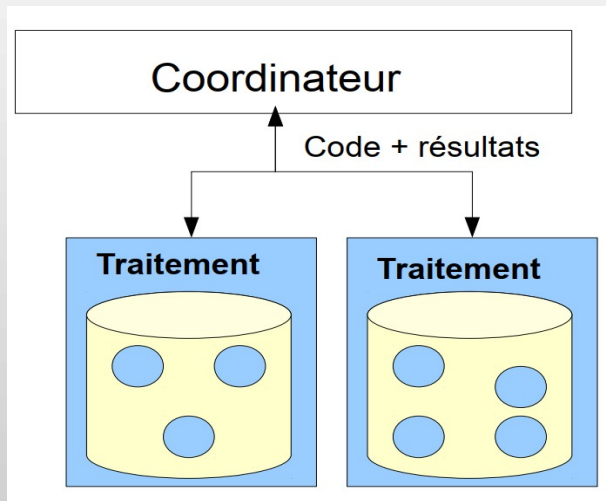
- **Agility (Agilité)**: Persistance polyglotte et flexibilité des schémas
- **Intrication** : Big Data – Total Data
- **Neccesity** : Open Sources



Les Limites des SGBDR



- Distribution et fragmentation des données
- Indexation des données
- Multiplication des ressources : Data center, grille, cloud
- Niveau de concurrence (cohérence faible)
- Dénormalisation des schémas : OLTP à OLAP (modèle étoilé, flocon de neige)





Les Limites des SGBDR

- **Rigidité** : Ajout de nouvelles informations avec des contraintes (non adapté à la web 2.0)
- **Complexité** : Requête long et plusieurs jointures (amis de amis sur facebook)
- **Rapidité** : Traitement centralisé et beaucoup de lecture/écritures sur disque
- **Ressources limitées** : Taille fixe des serveurs, serveurs propriétaire et coûteux





Qu'est ce que le NoSql ?

- ✓ **No SQL = Pas de SQL** : En 1998 par Carlos Strozzi utilise ce terme pour la première fois en désignant les BD relationnelles Open Source qui n'utilisent pas du SQL .
- ✓ **No SQL = Not Only SQL** : En 2009 face aux faiblesse du SQL (BDR) et sans solution sur le marché, les GAFAM (Google, Amazon, Facebook , Apple, Microsoft) proposent des solutions non – relationnelle

« 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 »



Pourquoi le NoSql?

- **Flexibilité** : Les solutions NoSQL sont sans schéma fixe, ainsi il s'adaptent à tout type de données. Les objets complexes peuvent être mappées sans trop de problèmes.
- **Elasticité** : Ajout de nouvelle données sans altération des anciens ou définition de format/type



Pourquoi le NoSql?

- **Stockage et traitement distribué** : Une découverte de Google et Facebook de parallélisation des tâches dans un environnement clustérisé (Map Reduce)
- **La volumétrie** : Bon niveau de performance face à la volumétrie .
- **Disponibilité et cohérence relâchée** :

ACID → BASE

MERCI