



RER collectif#2 : Introduction aux BDD



Cadre

Comprendre ce qu'est une base de donnée, connaître les diff types BDD, les terminologies, diff types de données



Mots-clés

▼ Données

les chiffres, les mots et caractères existant sous forme brute.

▼ Information

ensemble de données collectées, structurées de manière collective dans un contexte spécifique

▼ Connaissance(s) (ou modèle)

on les obtient lorsqu'on traite des données, pour en tirer des patterns dans un contexte particulier

▼ Meta-connaissances

connaissances sur les connaissances, soit la sagesse.

▼ Données structurées

données stockées relativement à d'autre dans dans des champs prédefinies afin d'etre facilement interprétées par un SGBD

▼ Données non structurées

Tout ce qui ne respecte pas les caractéristiques d'une donnée structurée. Exemples : fichier texte, image, vidéo

▼ Données semi structurées

L'ensemble des dopnnées non-structurées décrites par des méta-données structurées

▼ Actions principales

Stocker

collecter, enregistrer, recueillir, mémoriser...

Organiser

apporter une structure à la donnée relativement à d'autres

Traiter

manipuler

▼ Base de données ou BDD

ensemble des données structurées, semi-structurées et non-structurées, indexées pour faciliter leurs recherches

▼ BDD relationnelles

données structurées reliées entre elle par des champs communs. Ces champs communs s'incarnent par les notions de clés primaires et étrangères reliant les tables entre elle

▼ BDD non relationnel

C'est l'inverse. On retrouve un format clé-valeur (lié à un index). Surtout pour la collecte de données non-structurées

▼ SQL ou Structured Query Language

Langage de programmation des BDD relationnelles. Il permet de définir, protéger et manipuler les données de manière simple et schématique

▼ NoSQL ou Not Only Structured Query Language

Langage de programmation du Big Data, pour les données semi-structurées et non-structurées ; utilisé dans des BDD non-relationnelles

▼ MySQL

Le SGBD le plus répandue du marché, traite les données structurées gratuitement et en Open source.

▼ Big Data

Souvent traduit par « données massives » définit par la règle des 3V = Volume, Vitesse, Variété à des données

variées arrivent dans un volume croissant et à une vitesse plus élevée. Il permet d'obtenir des réponses plus complètes car le volume d'informations est plus important et indirectement réponses complètes signifie plus de confiance dans les données. (merci Briand)

▼ Cloud computing

Pratique visant à gérer des données informatiques à distance, chez des hébergeurs ou des serveurs, dans des centres de calcul géographiquement dispersés.

▼ Champs

C'est un attribut - soit une caractéristique ou une propriété - décrivant les instances des colonnes des BDD relationnelles - ou plus simplement la colonne d'une table.

▼ Enregistrement (ou tuple)

Constitue les lignes d'une BDD ; consiste en un ensemble de valeurs organisées selon des attributs spécifiques et possédant un type défini

▼ Tableau

table formée d'un ensemble d'enregistrements (lignes) et d'attributs (colonnes)

? Problématiques

Quels type de bdd pourraient répondre aux besoins de

Comment stocker, traiter et organiser des données

| l'entreprise ?

| structurées et non
structurées?



Hypothèses

- ▼ La BDD permet de stocker la donnée client de l'entreprise - *Adeline*

Vrai

- ▼ La BDD relationnelle est majoritairement utilisée pour extraire les données entre tableaux - *Loïc*

*Vrai, via **les champs communs***

- ▼ Les activités de l'entreprise peuvent impacter le type de BDD - *Seydou*

*Cela dépend du **type de données collectées** plutôt que des activités*

- ▼ Le type de BDD ne permet de n'utiliser qu'un seul type de donnéé - *Adrien*

*Vrai, uniquement en terme de **données structurées et non-structurées***

- ▼ On a plusieurs types de BDD - *Solenn*

*Vrai, **BDD relationnelles et non-relationnelles***

- ▼ Une BDD permet de stocker un nombre important de données - *Axel*

*Tout est relatif. Il faut **prédéfinir la taille** de la **BDD** dans une **SGBD***

- ▼ Une BDD permet de recollecter les données non-ordonnées d'une entreprise pour les replacer de manière ordonnée - *Briand*

Vrai

- ▼ les différents types de BDD présentent des contraintes différentes en terme d'opération qui peuvent être conduites sur les données - *Etienne*

*Vrai, car on ne peut effectuer le même type de transaction sur une **BDD relationnelle** et **BDD non-relationnelle***

- ▼ La BDD ne peut traiter la donnée si cette dernière n'a pu être stockée et organisée au préalable - *Adeline*

Vrai

- ▼ Une BDD doit contenir des données uniques à un tableau - *Tatyana*

*Faux, certes un tableau doit contenir au moins un **champ** à valeur unique (**clé primaire**) mais on peut retrouver des **champs communs** à différentes tables*

- ▼ Toutes les entreprises n'ont pas besoin de BDD relationnelles - Jean-Paul

Faux, pour des besoins organisationnelles, hiérarchiques, salariales...

- ▼ MySQL est le type de base de données le plus simple à exploiter - Nicolas

Faux, c'est un logiciel et non un type de BDD. Il peut être ou pas le plus simple à utiliser mais il s'agit du plus répandu

- ▼ les BDD relationnelles fonctionnent avec des champs-clé - Aude

Vrai, on les appelle clés primaires et clés étrangères

Plan d'action

Avantages et inconvénients des BDD relationnelles et non relationnelles

by Etienne Wagner

	Base de données relationnelles	Base de données non relationnelles
Avantages	<ul style="list-style-type: none">- Les données sont structurées - Plus sécurisées car moins tolérantes - Système de gestion des droits utilisateurs permettant un ajustement aux données	<ul style="list-style-type: none">- Accepte les données semi-structurées ou non structurées - Meilleures performances (plus rapide à l'écriture et à la lecture) - Distribution des données sur plusieurs nœuds ce qui permet une plus grande extensibilité et une meilleure tolérance aux erreurs - Plus grande flexibilité en utilisant un modèle de données affranchi des contraintes du modèle relationnel - Une administration et une prise en main plus simple
Inconvénients	<ul style="list-style-type: none">- Faible extensibilité - Non procédural - Ignore les notions de variable dimensionnée	<ul style="list-style-type: none">- Moins sécurisée - La manipulation des données

et indicée, de format, les affectations, les entrées-sorties - Peu flexible car la structure de la base de données est contrainte par les choix réalisés lors de sa création (nombre d'attributs, type de données, nom des relations) cela implique une faible portabilité vers des matériels différents, ou leur représentation est différente - Incompatible avec les données non structurées (un enregistrement qui ne renseigne pas de valeur pour certains attributs n'est pas intégré à la base de données) - Demande une formation préalable à la logique relationnelle et à l'utilisation des bases SQL	est plus ardue
---	----------------

Rechercher caractéristiques des données structurées et non structurées et leur différences

<u>Données Structurées</u>	<u>Données non-structurées</u>	<u>Données semi-structurées</u>
Accès en libre-service	Expertise en data science nécessaire	Expertise en data science nécessaire
Types de données sélectionnées	Nombreux types différents en conglomérats	basées sur XML et RDF
Schéma à l'écriture	Schéma à la lecture	Plus flexible que données structurées et moins que les données non-structurées
Généralement stockées dans des data warehouses	Généralement stockées dans des data lakes	
Format prédéfini	Format natif	
Mise à l'échelle laborieuse	mise à l'échelle simple	Mise à l'échelle de difficulté relative