Bases de Données

Ludovic Deneuville

Une donnée ?

- texte (nom, prénom, ville…)
- numérique (age, prix…)
- date
- liste (compétences, loisirs)
- image
- fichier

Définition d'une donnée

> **i** Définition
>
> Une **donnée** est une information brute composée de 3 éléments :
>
> - un concept (exemple : hauteur d'un monument)
> - un domaine des possibles (nombre entier positif et une unité : mètres)
> - une valeur (pour la tour Eiffel : 324)

Source : Qu'est-ce qu'une donnée ?, Courrier des statistiques N5, Pascal Rivière

Où stocker des données

- Fichier txt
- Fichier csv
- Fichier json, Exemple
- Fichier parquet
- Base de données

Format Parquet

- Stockage orienté colonne
- Contient des métadonnées
- Adapté pour le stockage de données figées

> Articles intéressants
>
> - Parquet pour remplacer csv, Éric Mauvière, Icem7
> - Comment bien préparer son Parquet, Éric Mauvière, Icem7
> - Lire et écrire des fichiers Parquet, UtilitR

Une base de données ?

> **i** Définition
>
> Une **base de données** est un ensemble de données rassemblées et stockées de manière organisée dans un but spécifique.

Usages

Usage transactionnel

- Faire fonctionner le système au quotidien
- OLTP : On-Line Transaction Processing
- Applications et utilisateurs métiers

Usage décisionnel

- Analyser et comprendre ce qui se passe
- OLAP : On-Line Analytical Processing
- Explorer de grands volumes de données historiques pour aider à la décision
- Data warehouse, outils BI

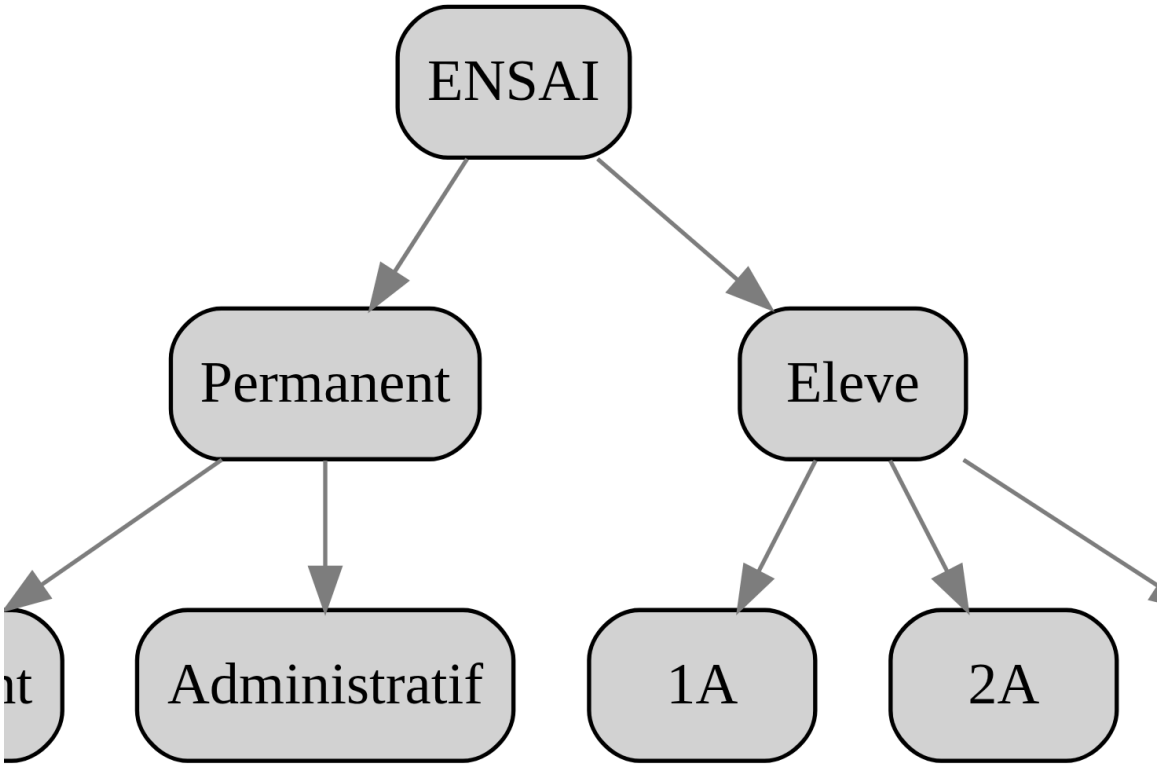
Enjeux

- Accès aux données : langage SQL
- Sécurité des données : seuls les utilisateurs autorisés (droits)
- Partage des données : facile en lecture, mais en écriture ?
- Non-redondance des données
- Cohérence des données : respect des contraintes d'intégrité

Types de bases de données

Hiérarchique

- Données organisées comme un arbre
- Relations de type parent-enfant
- Annuaires LDAP, systèmes de fichiers



Réseau

- Organisée en graphe
- Les enregistrements peuvent avoir plusieurs parents et enfants
- Réseaux de transports

Relationnelle

- Organisée en tables
- Lien entre les tables
- Données structurées de manière organisée
- Cohérence des données grâce aux contraintes
- Peu de duplications

NoSQL

Nombreux sous-types :

Type	Description	Exemple
Clé-valeur	Données sous forme de paires clé-valeur. Idéale pour des lectures rapides avec peu de structure.	DynamoDB
Colonnes	Organisée en colonnes, adaptée aux grandes quantités de données avec peu de relations entre elles.	Cassandra
Document	Stocke des documents semi-structurés, souvent en format JSON ou XML, permettant une grande flexibilité.	MongoDB
Graphes	Organisée sous forme de nœuds et de relations, idéale pour des données très connectées ou inter-relées.	Neo4j

Autres

- Graph Database
- Bases NewSQL (relationnelles « modernes » distribuées)
- Bases de données de séries temporelles
- Bases de données vectorielles
- Lakehouse

Base de données relationnelle

Table

- Composée de lignes et de colonnes
- Chaque table représente une entité
- Chaque ligne est un enregistrement
- Chaque colonne représente un attribut de l'entité
  - Types de données : texte, num…

Relations entre les tables

- **Clé primaire** (PK) pour identifier chaque enregistrement dans une table
  - de manière unique
  - sans ambiguïté
- **Clé étrangère** (FK) pour créer des liens avec d'autres tables
  - attribut d'une table A
  - correspond à la clé primaire d'une table B

Clé primaire ?

nom	prenom	elo
Guichard	Pauline	2357
Daulyte-Cornette	Deimante	2371
Navrotescu	Andreea	2338
Sebag	Marie	2448
Millet	Sophie	2366
Skrichempko	Almira	2361

Quel champ choisir comme clé primaire ? 😐

Clé primaire technique

Créons une séquence numérique. 😎

id_joueuse	nom	prenom	elo
1	Guichard	Pauline	2357
2	Daulyte-Cornette	Deimante	2371
3	Navrotescu	Andreea	2338
4	Sebag	Marie	2448

id_joueuse	nom	prenom	elo
5	Millet	Sophie	2366
6	Skrichempko	Almira	2361

Clé primaire

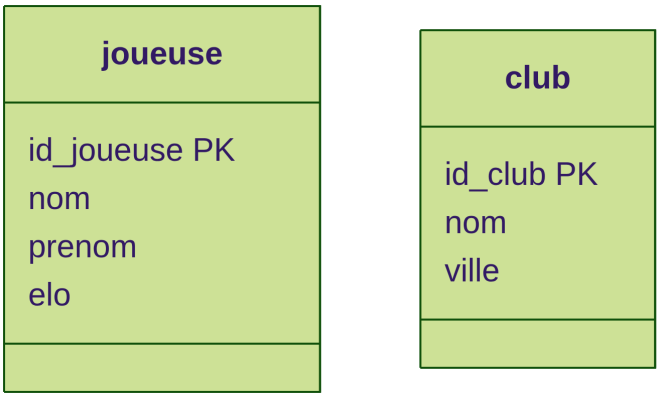
- Attribut (ou ensemble d'attributs)
- Unique
- Non nulle
- Immuable

> Tip
>
> Sauf mention contraire, nous utiliserons une clé primaire "technique" : séquence.

Clé étrangère ?

Comment relier ces 2 tables ?

- Une joueuse joue pour un seul club
- Un club peut avoir plusieurs joueuses



---

club

id_club	nom	ville
1	C'Chartres Echecs	Chartres
2	Bischwiller	Bischwiller
3	Asnieres - le Grand Echiquier	Asnieres
4	Tours des Hauts-De-France	Saint-Quentin

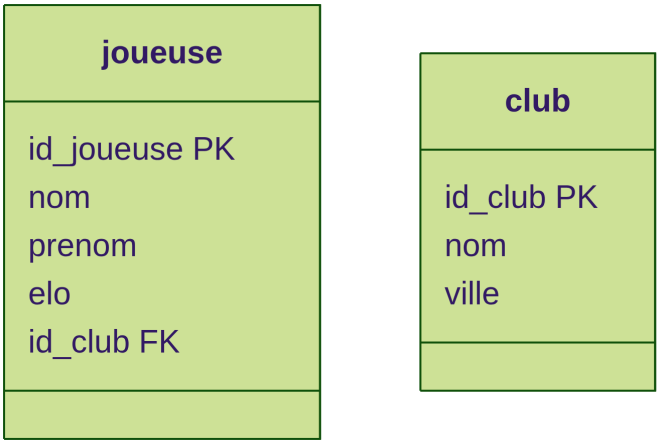
joueuse

id_joueuse	nom	prenom	elo
1	Guichard	Pauline	2357
2	Daulyte-Cornette	Deimante	2371

id_joueuse	nom	prenom	elo
3	Navrotescu	Andreea	2338
4	Sebag	Marie	2448
5	Millet	Sophie	2366
6	Skrichempko	Almira	2361

Clé étrangère

En ajoutant une clé étrangère.



---

id_joueuse	nom	prenom	elo	id_club
1	Guichard	Pauline	2357	3
2	Daulyte	Deimante	2371	1
3	Navrotescu	Andreea	2338	2
4	Sebag	Marie	2448	2
5	Millet	Sophie	2366	3
6	Skrichempko	Almira	2361	

🔥 Remarques

Une clé étrangère peut être nulle.

Une clé étrangère référence une valeur existante de clé primaire.

Question

Et si vous supprimez, le club de C'Chartres Echecs ? 😇

⚠️ Warning

```
SQL Error [23503]: ERROR: UPDATE or DELETE on table "club" violates the foreign key constraint "fk_id_club" on table "joueuse". Detail: Key (id_club)=(1) is still referenced from table "joueuse".
```

Intégrité référentielle

- Garantit que les relations entre tables restent cohérentes
- Empêche les références invalides
- Empêche la suppression d'une référence utilisée
  - Évite les "valeurs orphelines"

Autre question

Si une joueuse peut adhérer à plusieurs clubs ? 😯

- *id_club* devient une liste de clubs ?

Table d'association

Création d'une table d'association : **adhesion** 😅



---

id_club	id_joueuse
1	2
2	3
2	4
2	6
3	1
3	5
4	1
4	2

Associations

Il existe plusieurs types d'associations entre tables.

Association 1:1

Une entité dans la table A est associée à une seule entité dans la table B.

Exemple :

- Une personne a une seule carte vitale
- Une carte vitale appartient à une seule personne

Association 1:n

Une entité dans la table A peut-être associée à plusieurs entités dans la table B.

Exemple :

- Un club est composé de plusieurs joueuses
- Une joueuse adhère à un seul club

Association n:n

Une entité dans la table A peut-être associée à plusieurs entités dans la table B.

Et inversement, une entité dans la table B peut-être associée à plusieurs entités dans la table A.

Exemple :

- une joueuse participe à plusieurs tournois
- Un tournoi regroupe plusieurs joueuses

Association récursive

Une table peut avoir des relations avec elle-même.

Dans ce cas, la table contient une clé étrangère qui référence sa propre clé primaire.

Exemple :

- *employe (id_employe, nom, prenom, #id_chef)*
- *id_chef* clé étrangère qui référence *id_employe*

Association multiples

Une table A peut avoir plusieurs relations avec une table B.

Exemple :

- *equipe (id_equipe, nom)*
- *match (id_match, #id_equipe1, #id_equipe2, score)*

SGBD

> ! Définition
>
> Un **Système de Gestion de Bases de Données** est un logiciel qui permet de créer, gérer et manipuler des bases de données.
>
> Il fournit des outils pour stocker, récupérer, modifier et sécuriser les données.

- **PostgreSQL**
- MySQL, MariaDB

- Oracle Database
- SQLite

PostgreSQL

- Open Source, Libre
- Communauté active
- Stable et performant
- Basé sur le modèle client-serveur
- Extension PostGIS (Données géographiques)

Outil de Gestion et Administration

Il est nécessaire d'avoir un **outil pour gérer et administrer** une base de données.

> Tip
>
> - PostgreSQL : Installé sur un serveur, sert de moteur de base de données
> - pgAdmin : Interface graphique pour interagir avec PostgreSQL

- pgAdmin
- DBeaver ou cloudBeaver

Kesako

- Base de données : endroit où sont stockées les informations
- SGBD : logiciel qui permet de créer, gérer et manipuler les bases de données
- Outil de gestion du SGBD : application graphique qui facilite l'interaction avec un SGBD
- SQL : Langage pour donner les instructions à la base de données

---

Next step : SQL