

# Bases de données

## 1 Introduction

L'objet des bases de données (abrégié BD) est la gestion d'informations, qui peuvent se révéler nombreuses et variées dans des applications réelles.

Il s'agit de trois choses :

1. Stocker les informations de manière persistente,
2. Permettre l'ajout, la suppression et la modification de ces informations,
3. Rendre facilement (de manière aisée pour des non-spécialistes) et efficacement accessible ces informations.

On entend par facilement "de manière aisée pour des non-spécialistes". On utilise pour cela un langage d'interface simple.

Les systèmes de gestion de bases de données (abrégié SGBD) assurent le stockage physique des informations, et les opérations de création / ajout / modification et recherche, pourvu qu'elles soient formulées dans un langage adéquat, appelé **langage de requête**. Pour cela, les informations doivent être organisées dans une base de données, que l'on peut définir comme un ensemble d'informations structurées.

La structure d'une base de données est définie par ses **tables**, et les **colonnes** de ces tables. Chaque table est identifiée par son nom, et contient un nombre fixé de colonnes. Chaque colonne est identifiée par son nom, et fixe un **domaine** de valeurs possibles : entiers, chaînes de caractères, ou flottants. Ce sont les lignes de ces tables, appelées **enregistrements**, qui contiennent ces informations.

**Remarque:** La définition de base de données n'interdit pas la multiplicité des lignes. En pratique, on s'interdira de stocker deux fois la même ligne dans le même tableau.

Si l'on fixe l'ordre des  $n$  colonnes d'une table, et que l'on note  $(D_i)_{i \in \llbracket 1, n \rrbracket}$  les domaines de ces colonnes, chaque enregistrement est un  $n$ -uplet de  $D_1 \times \dots \times D_n$ .

Le nombre de lignes n'est pas limité, et leur ordre n'est pas significatif, de sorte qu'une table puisse être vue comme un sous-ensemble de  $D_1 \times \dots \times D_n$ , que l'on appelle aussi une relation sur  $D_1 \times \dots \times D_n$ .

**Remarque:** Techniquement, il ne s'agit pas réellement d'un sous-ensemble à cause de la multiplicité possible, mais plutôt d'un multi-ensemble. Mais on évitera toujours de telles redondances via l'identification d'une clé (voir plus loin).

La gestion d'une base de données consiste donc en ces trois points :

1. la conception de la base de données, c'est à dire décider de la structure la plus adéquate pour une application donnée ;
2. l'acquisition des données et leur enregistrement dans la base ;
3. la valorisation des données, c'est à dire l'extraction d'informations pertinentes.

## 2 Le modèle entité/association (E/A)

Le modèle entité-association est un outil qui permet de représenter la structure d'une base de données. Il est utile en phase de conception, ou pour visualiser une base de données existante.

**Exemple:** On cherche à enregistrer un groupe de personnes, l'endroit où elles habitent, la capitale du pays d'où elles viennent et leur âge.

Nom	Prénom	Pays de résidence	Capitale	Date de naissance
DURAND	Paul	Pays-Bas	Amsterdam	26/09/1981
WILES	Andrew	Royaume-Uni	Londres	11/04/1953
...	...	...	...	...

**Remarque:** On a ici une redondance entre le pays de résidence et sa capitale : cela pose des problèmes d'espace occupé, mais surtout si on doit faire une correction (imaginons qu'un pays change de capitale!), il faut s'assurer que le changement soit fait partout. Il serait ici plus adapté de stocker dans une autre table les différents pays et leur capitale, et ne stocker alors que le pays dans la table précédente.

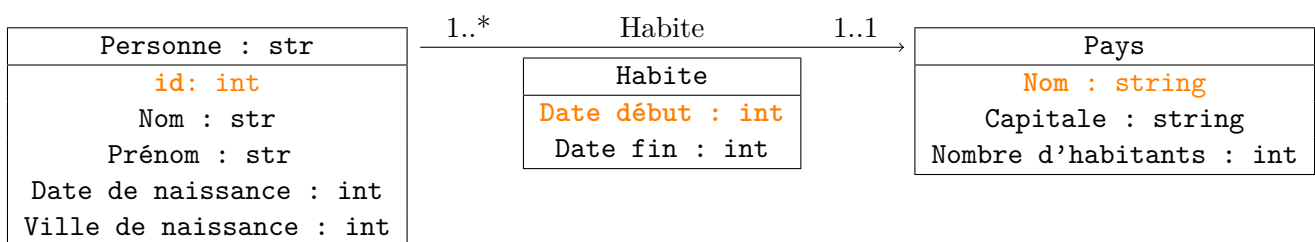
**Remarque:** Il faut aussi éviter les doublons, qui pourront poser des problèmes de dénombrement, de cohérence et de mise à jour. Pour cela, il faut identifier ce qui caractérise un enregistrement, c'est à dire la **clé** de cet enregistrement, et imposer que deux enregistrement aient deux clés différentes.

La clé d'une table est donc un sous-ensemble de ses colonnes, telle que deux lignes de la table ne coïncident pas sur ces colonnes. Il y a parfois plusieurs clés possibles. Celle qu'on indique au SGBD est appelée clé primaire.

Une bonne clé est :

- composée de données stables dans le temps, peu sujettes à des modifications ;
- composée de données dont on dispose pour chaque enregistrement ;
- la plus petite possible.

Si parmi les données aucune ne satisfait ces contraintes, on ajoutera un identifiant artificiel qui servira de clé.



REPRÉSENTATION DE DEUX ENTITÉS ET D'UNE ASSOCIATION ENTRE ELLES. Les lignes en orange désignent les clés.

Dans le modèle entité-association, on différencie les tables qui modélisent les objets de celles qui modélisent les liens entre ces objets. Les premiers sont appelés entités, et les seconds associations.

On définit un type d'entité par la donnée de son nom, ses attributs avec leurs domaines, et sa clé. On définit une association binaire par son nom, les deux entités qu'elle relie (A et B), des cardinalités  $c_A$  et  $c_B$ , éventuellement des attributs. Chaque clé de A apparaît autant de fois qu'indiqué par  $c_A$  dans la table de la relation.

Toutes ces données se représentent comme ci-dessus. *Reprendre la notion de cardinalité lundi...*