

TD1 : Introduction aux bases de données

Données persistantes

Les variables créées dans un programme sont détruites à la fin de l'exécution du programme. Par exemple en Python :

```
s = "abc def"  
v = input()  
s += v
```

La chaîne de caractère `s` est détruite à la fin de l'exécution. Les données rentrées par l'utilisateur aussi. Il n'est pas possible de les récupérer si on ne les a pas sauvegardé sur l'ordinateur. On dit que ces données ne sont pas persistantes.

A l'inverse si le code python avait travaillé sur un fichier sauvegardé sur l'ordinateur et l'avait modifié et enregistré, alors dans ce cas les données sont persistantes car elles sont récupérables sur l'ordinateur même après l'extinction de ce dernier.

Les bases de données sont des données persistantes structurées pour faciliter la gestion des données.

Cas d'étude : les employés d'une entreprise

Considérons une entreprise dans laquelle on désire stocker des données concernant les employés et les départements dans lesquels travaillent ces employés. Nous allons tout au long de ce cas d'étude examiner différentes façons de représenter ces données au moyen de tableaux que nous allons manipuler avec votre tableur favori. Ces exercices doivent vous permettre d'appréhender la modélisation de données persistantes et de comprendre les difficultés liées à une représentation tabulaire des données, telles qu'elles le sont dans les bases de données.

À l'intérieur de l'entreprise, quelques règles de gestion sont mises en œuvre. Vous devrez veiller à les respecter lors de cette étude :

- Un employé peut appartenir à plusieurs départements
- Un employé n'a qu'un seul poste dans un même département
- Un employé n'a qu'un seul manager même s'il occupe plusieurs postes
- Deux départements, situés dans la même ville, ne peuvent pas avoir le même nom

Les données stockées sont les suivantes :

- EMPNO : numéro de l'employé
- ENAME : nom
- JOB : poste
- MANAGER : numéro du manager
- SAL : salaire en €
- HIREDATE : date d'embauche

- COMMISSION : commission en €
- DEPTNO : numéro du département
- DNAME : nom du département
- LOC : localisation du département

Télécharger les fichiers struct1.xls, struct2.xls et struct3.xls.

Structure 1 : un tableau

Le fichier cas1.xls présente de manière partielle une seule feuille de calcul avec la relation suivante :

TABLEAU_1 (EMPNO, ENAME, JOB, SAL, MANAGER, HIREDATE, COMMISSION, DEPTNO, DNAME, LOC)

Question 1 : Où travaille l'employé DOE ? Où travaille l'employé SCOTT ? Quelle règle de cette situation illustre ces deux employés ?

Question 2 : la règle 2 est elle vérifiée pour l'employé SCOTT ? Et pour l'employé ADAMS ?

Question 3 : Suite à une erreur de saisie, l'employé ADAMS ne travaille pas dans le département RESEARCH mais dans le département SALES de BOSTON comme ANALYST. Corriger les données.

Question 4 : est ce que la règle 4 est elle vérifiée ? Corriger cela en renommant le département 50.

Question 5 : Faites les mises à jour nécessaires pour renommer en REAL ESTATE le département dans lequel travaille l'employé TURNER. Dans le pire des cas, combien de lignes devez-vous parcourir en lecture ? Combien de lignes avez-vous dû modifier ? À quoi ce dernier nombre correspond-il ?

Structure 2 : deux tableaux

On restructure les données précédentes en créant deux tableaux. Dans le premier tableau on stocke des informations sur les employés et dans le second tableau on stocke les informations sur les départements.

Question 6 : Quel est l'intérêt, par rapport à la question précédente, d'avoir divisé ainsi les données ?

Question 7 : Comment le tableau 1 fait il référence au tableau 2 ? Est ce que l'on est sûr de quel département on parle en utilisant DNAME ?

Question 8 : Corriger le problème précédent en modifiant le tableau 1. Quel attribut des départements est utilisé pour faire la référence ?

Question 9 : Faites les mises à jour nécessaires pour renommer en REAL ESTATE le département dans lequel travaille l'employé TURNER. Dans le pire des cas, combien de lignes devez-vous parcourir en lecture ? Combien de lignes avez-vous dû modifier ?

Définition : on dit que DEPTNO est une *clé primaire* dans le tableau 2 : c'est un *attribut* qui doit être unique dans un tableau. Dans le tableau 1, DEPTNO joue le rôle d'une *clé étrangère* : c'est un attribut qui fait référence à une clé primaire d'un autre tableau.

Question 10 : l'attribut DNAME du tableau 2 peut il être une clé primaire ? L'attribut LOC du tableau 2 peut il être dans ce cas d'études une clé primaire ? Et en toute généralité si on continue à remplir la base de données ?

Question 11 : quels sont les attributs qui peuvent être des clés primaires pour le tableau 1 ?

Question 12 : Le salaire de SCOTT a augmenté et est passé à 3500. Modifier le tableau en conséquence.

Question 13 : Y a-t-il des redondances dans cette structure ?

Question 14 : Imaginer comment restructurer les données pour n'avoir à faire qu'une seule modification pour la question précédente ?

Structure 3 : trois tableaux

Dans cette troisième structure des données il y a trois tableaux : un pour les données sur les employés, un pour les départements et un pour les affectations. Chaque ligne du tableau 3 des affectations dit que l'employé EMPNO travaille dans le département DEPTNO en tant que JOB.

Question 15 : Remplissez les 3 tableaux en fonction des données fournies dans les structures précédentes.

Question 16 : Identifiez dans chacun des 3 tableaux, les clés primaires et étrangères.

Question 17 : En quoi cette nouvelle structure permet de modifier le salaire d'un employé sans oublier des lignes ?

Question 18 : Où travaille DOE ?

Question 19 : Faites les mises à jour nécessaires pour déplacer le département dans lequel travaille l'employé JONES à NEW YORK.

Question 20 : Y a-t-il encore des données redondantes ?

Bilan

Question 21 : Quel est le danger des redondances ? Comment restructure-t-on les données pour éviter les redondances ? Pourquoi cependant cette restructuration complexifie la recherche dans les données ?

Exercice de modélisation 1

On désire concevoir une base de données relatives à la présentation de films dans des cinémas. Chaque cinéma a un nom, est situé dans une ville et a un certain nombre de salles. Chaque film a un titre et une durée.

Établir une structure de base de données évitant les redondances. Préciser les clés primaires et étrangères.

Exercice de modélisation 2

La société Shinkan-Zen gère le transport de marchandises commandées par des clients via des camions qui les déplacent d'un dépôt vers un autre dépôt.

Établir une base de données qui permette de répondre aux questions suivantes :

- où est la marchandise ? Dans un dépôt ? Dans un camion ?
- quel est le numéro du client à qui appartient tel marchandise ?
- est-ce que le camion est plein ? (les marchandises ont un volume et le camion aussi)

- sur quel trajet est tel camion ?

Identifier les clés primaires et étrangères.

Exercice de modélisation 3

Lors des jeux olympiques, des athlètes concourent à des épreuves qui sont couverts par des journalistes.

Établir une base de données qui permette de répondre aux questions suivantes :

- dans quel épreuve concoure actuellement un athlète ?
- quel épreuve couvre actuellement un journaliste ?
- à quels épreuves un athlète ou un journaliste a été présent ?

Identifier les clés primaires et étrangères.