

Persistance des données I

DON2

Denis Boigelot, Geneviève Cuvelier, Selim Rexhep,
Yannick Voglaire



Haute École Bruxelles-Brabant
École Supérieure d'Informatique

Année académique 2020 / 2021

Plan du cours

- 0 – Présentation
- 1 – Introduction
- 2 – Dépendance fonctionnelle
- 3 – Schéma conceptuel
- 4 – Projection et sélection
- 5 – Jointure
- 6 – Agrégat
- 7 – Sous-requête
- 8 – Fichiers

Table des matières

1 La persistance des données

2 Fichiers

3 Base de données

Persistance des données

Ce cours traite de la 'persistance des données'. De quoi s'agit-il exactement ?

Persistance des données

« En programmation , la gestion de la persistance des données (en anglais : persistence) et parfois des états d'un programme réfère au mécanisme responsable de la sauvegarde et de la restauration des données. Ces mécanismes font en sorte qu'un programme puisse se terminer sans que ses données et son état d'exécution ne soient perdus. »

[https://fr.wikipedia.org/wiki/Persistance_\(informatique\)](https://fr.wikipedia.org/wiki/Persistance_(informatique))

Persistance des données

Mémoire rémanente Une mémoire est dite rémanente ou persistante, en opposition à volatile si la mémoire n'est pas détruite lors de l'extinction du système.

Exemple

Votre feuille de papier sur laquelle vous prenez actuellement des notes est une mémoire persistante.

Mémoire secondaire

Mémoire secondaire La mémoire secondaire, aussi appelée mémoire de masse, mémoire de stockage est une mémoire rémanente qui sert à stocker l'information à long terme.

Quelques mémoires secondaires

- ◇ HDD,
- ◇ SSD,
- ◇ Cartes SD,
- ◇ Disque optique (CD, Blu-ray...),
- ◇ ...

Persistance des données

Remarque : on parle de **mémoire secondaire** pour marquer la différence avec la **mémoire centrale** de l'ordinateur.

Voyez-vous bien la différence entre les deux ?

Fichiers

Fichier La plupart du temps, les données sont organisées sous forme de fichier sur une mémoire secondaire.

Exemples

- ◇ Son : MP3, FLAC...
- ◇ Image : JPG, PNG, ...
- ◇ JAVA : .Class, .java
- ◇ ...

Fichiers

L'organisation des fichiers sur le dispositif de mémoire secondaire est prise en charge par le système d'exploitation, au moyen d'un **système de gestion de fichiers** (en abrégé : **SGF** en anglais **FS** pour File System).

Le but de ce cours n'est pas d'étudier le fonctionnement des SGF, ceci sera fait dans le cours de système d'exploitation.

Fichiers

La mémoire est donc organisée en fichiers, gérés par le système d'exploitation. Si l'utilisateur ou un programme quelconque souhaite y enregistrer des données, il lui suffit : d'ouvrir le fichier, d'y inscrire les données, de fermer le fichier.

Mais est-ce vraiment aussi simple ?

Fichiers

Essayons d'imaginer le mode de fonctionnement d'une entreprise devant gérer de grandes quantités de données quotidiennement : une banque, un supermarché, une école,...

Citons quelques problèmes posés par la gestion des données dans de telles institutions.

Fichiers

- 1 Volume et performances** : Une très grande quantité de données doivent être traitées (enregistrées, lues, modifiées) chaque jour, le plus rapidement possible.

Exemple : Selon vous, combien de transactions (paiement par carte, virement) le secteur bancaire doit-il gérer par an ? ¹

1. Pour un aperçu des chiffres du secteur bancaire :
<http://dashboard.febelfin.be/fr>

Fichiers

2 Accès aux données : Comment faire pour obtenir rapidement une vue sur des données répondant à des conditions bien précises ?

Exemple : La banque XY souhaite avoir un rapport sur le nombre de clients habitant Liège et ayant effectué une transaction de plus de 10.000 euros lors des 5 derniers jours.

Fichiers

- 3 Sécurité** : Les données à traitées sont souvent confidentielles (données de paiement, données médicales, ...) , et doivent être protégées contre les tentatives d'accès indésirables.

Exemple : Comment s'assurer que seuls les employés du département *X* de la banque puisse avoir accès aux données des comptes internes ?

Fichiers

- 4 Accès concurrents** : Les données doivent souvent être accessibles simultanément par un grand nombre d'utilisateurs, sans risque de corrompre les données en questions.

Exemple : Les détails des produits vendus par une entreprise doivent être accessibles par tout client potentiel. Le prix de ces produits doit pouvoir être mis à jour facilement par les employés de l'entreprise.

Fichiers

- 4** **Intégrité des données** : Les données doivent être protégées contre les erreurs de l'utilisateur, ainsi que contre les éventuelles pannes du système informatique.

Exemple : L'employé de la banque XY peut se tromper en encodant les données d'un client. Le serveur peut tomber en panne précisément au moment où le programme de la banque met à jour les données d'un utilisateur.

Fichiers

- 5 Accès par des non informaticiens** : Une solution facile d'accès doit souvent être mise en œuvre pour que les données puisse être consultées par des non informaticiens.

Exemple : Les détails des produits vendus par une entreprise doivent être accessibles par tout client potentiel. Le client doit pouvoir trier les données selon ses propres critères (le prix, le nom, la taille, ...).

Base de données

La notion de fichier en elle même ne permet pas d'apporter une solution à toutes ces questions. D'où la nécessité d'avoir :

- 1 Un modèle adéquat pour nos données (permettant de mettre en place les solutions aux problèmes soulevés dans les slides précédents),
- 2 Un type particulier de programme, mettant effectivement en place les solutions en questions.

Base de données

On sera donc amené à stocker les données dans une **base de données** (en abrégé : BD), les programmes gérant ces bases de données étant appelés des **systèmes de gestion de bases de données** (en abrégé : SGBD, en anglais DBMS pour DataBase Management System).

Le modèle de base de donnée le plus courant actuellement est le **modèle relationnel**.

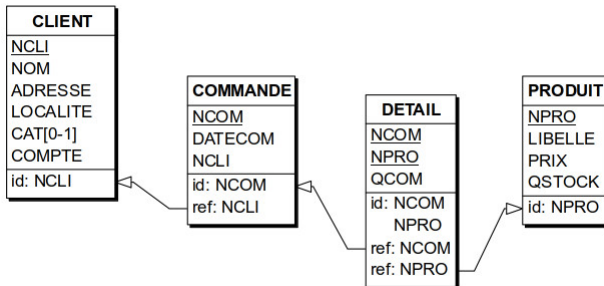
Le modèle relationnel

L'idée intuitive des BD relationnelles est simple : les données sont stockées dans des tables, partitionnés en lignes. Voici par exemple une table Client :

CLIENT					
NCLI	NOM	ADRESSE	LOCALITE	(CAT)	COMPTE
B062	GOFFIN	72, r. de la Gare	Namur	B2	-3200
B112	HANSENNE	23, r. Dumont	Poitiers	C1	1250
B332	MONTI	112, r. Neuve	Genève	B2	0
B512	GILLET	14, r. de l'Eté	Toulouse	B1	-8700
C003	AVRON	8, r. de la Cure	Toulouse	B1	-1700
C123	MERCIER	25, r. Lemaître	Namur	C1	-2300
C400	FERARD	65, r. du Tertre	Poitiers	B2	350
D063	MERCIER	201, bvd du Nord	Toulouse		-2250
F010	TOUSSAINT	5, r. Godefroid	Poitiers	C1	0
F011	PONCELET	17, Clos des Erables	Toulouse	B2	0
F400	JACOB	78, ch. du Moulin	Bruxelles	C2	0
K111	VANBIST	180, r. Florimont	Lille	B1	720
K729	NEUMAN	40, r. Bransart	Toulouse		0
L422	FRANCK	60, r. de Wépion	Namur	C1	0
S127	VANDERKA	3, av. des Roses	Namur	C1	-4580
S712	GUILLAUME	14a, ch. des Roses	Paris	B1	0

Le modèle relationnel

Des liens logiques entre les différentes tables d'une BD relationnelle peuvent être créés (nous en reparlerons dans la suite) :



SGBD relationnels

Quelques SGBD relationnels très courants :

- ◇ Oracle, Microsoft SQL Server (SGBD commerciaux, très courant en entreprise),
- ◇ SQLite (SGBD léger, fort utilisé dans les applications mobiles sur smartphone),
- ◇ MySQL,
- ◇ Microsoft Access,
- ◇ DB2 (IBM).
- ◇ ...

Le langage SQL

SQL est l'acronyme de **Structured Query Language**. Il s'agit d'un langage permettant de manipuler des BD relationnelles.

La plupart des SGBD relationnels actuels (quel que soit le fabricant) reconnaissent le langage SQL.

Le langage SQL

Le langage SQL comprend des commandes permettant notamment de :

- ◇ Créer/Détruire des tables dans la BD,
- ◇ Insérer/Supprimer/Modifier des lignes d'une table donnée,
- ◇ Sélectionner les lignes satisfaisant certains critères d'une table,
- ◇ Interdire/Autoriser l'accès de certaines tables à certaines utilisateur.

◇ ...

Le cours de DON2

Qu'allons-nous voir ensemble cette année ?

Nous utiliserons plusieurs bases de données installées sur plusieurs SGBDR en local, sur un serveur de l'école ou en ligne. Nous apprendrons à les manipuler ensemble, au moyen du langage SQL. Nous étudierons aussi les bases théoriques du modèle relationnel.

BD non relationnelles

Existe-t-il d'autres modèles que le modèle relationnel ?

Réponse : oui ! Mais nous n'en parlerons pas cette année, le modèle relationnel étant le plus couramment utilisé à l'heure actuelle.²

2. Voir le livre de J.-L. Hainaut (Chapitre 10) pour des informations sur les modèles non relationnels.

Le modèle ANSI-SPARC

Avant de poursuivre, donnons quelques explications sur la façon dont sont organisées les données dans un SGBD.

Un SGBD relationnel stocke les données dans des tables. Les données de ces tables sont (évidemment !) aussi stockées dans des fichiers.

Le modèle ANSI-SPARC

Un des gros avantages des BD : elles permettent à l'utilisateur de se focaliser sur l'aspect logique de la manipulation des données.

L'implémentation physique (répartition des données en fichiers, contrôle des accès, ...) est pris en charge pour vous par le SGBD.

Le modèle ANSI-SPARC

Pour une BD donnée, on distingue donc :

- 1 Son **schéma conceptuel** : celui-ci résume la logique du modèle de données (les différentes entités qu'on souhaite représenter, les liens qui existent entre elles, ...) sans se soucier de l'implémentation en machine.
- 2 Son **schéma interne** : les détails de la répartition en fichiers, la façon dont ces fichiers sont organisées, etc.

Le modèle ANSI-SPARC

Exemple : Le schéma conceptuel de la BD Client-Commande-Détails-Produits (introduit précédemment) est constitué de quatre entités : les clients, les commandes, les détails, les produits (chacune de ces entités deviendra une table de notre base de donnée).

La conception du schéma conceptuel d'un domaine d'application donné et sa traduction en un schéma de BD relationnelle n'est pas immédiat (nous en reparlerons l'an prochain).

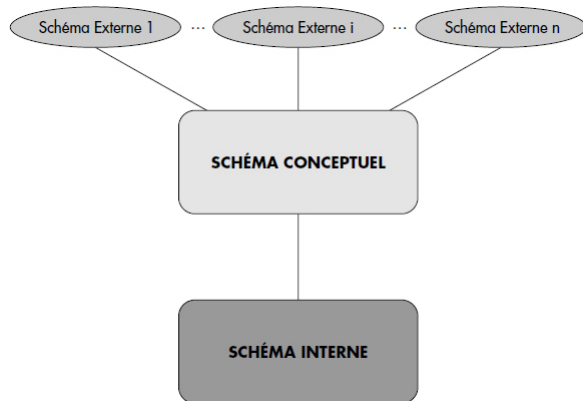
Le modèle ANSI-SPARC

On pourrait aussi définir des **schémas externes** sur les BD.

Il s'agit simplement de vues particulières sur les données, demandées par un utilisateur de la BD.

Le modèle ANSI-SPARC

Tout ceci peut se résumer au moyen du modèle ANSI-SPARC³ :



3. Source de l'image : Bases de données, Georges Gardarin - Eyrolles 2003.

