

Bases de données - Langage SQL

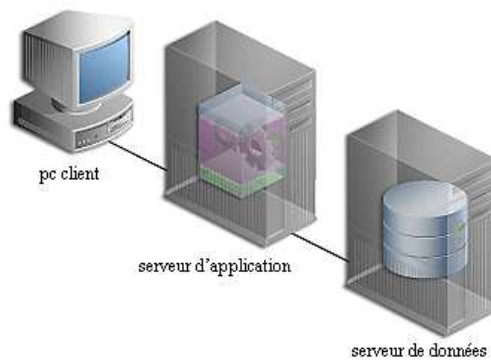
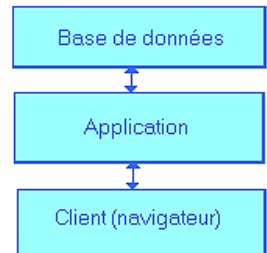
Enjeux : Big Data

Architecture trois tiers

Ce nom provient de l'anglais *tier* signifiant *étage* ou *niveau*. Il s'agit d'un modèle logique d'architecture applicative qui vise à modéliser une application comme un empilement de trois couches logicielles (étages, niveaux, tiers ou strates) dont le rôle est clairement défini :

- la présentation des données (affichage, dialogue avec l'utilisateur) ;
- le traitement des données ;
- l'accès aux données persistantes (données destinées à être conservées dans la durée, voire de manière définitive).

D'après Wikipédia : http://fr.wikipedia.org/wiki/Architecture_trois_tiers



Source image : <http://mnttechnologie.wordpress.com/type/aside/>

Exemples : pronote, banque en ligne, réseaux sociaux...

Introduction aux bases de données

L'exemple suivant (bibliothèque de prêt) ainsi que la base de données associée sont issus de travaux réalisés par M. Laurent Chéno, Inspecteur Général d'Informatique, mis à disposition sous licence CC BY-SA 3.0 FR.

Gestion d'une bibliothèque de prêt grâce à un tableur (structure de données plate) :

index	titre	genre	auteurN	auteurP	nom	prénom	tél	date	retour
1	Salambo	Roman	Flaubert	Gustave	Castel	Claude	0612345112	12-août-10	23-août-10
2	93	Roman	Hugo	Victor	Le Dray	Camille	0422113827	17-sept.-10	03-oct.-10
3	Mme Bovary	Littérature	Flaubert	Gustave	Loutard	Annie	064331282	01-nov.-10	02-déc.-10
1	Salambô	Roman	Flaubert	Gustave	Filatre	Jean	0432168719	02-sept.-10	14-sept.-10
4	Aphorismes	Littérature	Wilde	Oscar	Castel	Marie	0412324494		
5	L'immoraliste	Roman	Gide	André	Biraud	Michèle	0434578612	05-févr.-11	
3	Madame Bovary	Roman	Flaubert	Gustave	Castel	Claude	0412324494	06-janv.-10	
2	Quatrevingt-treize	Roman	Hugo	Victor	Filâtre	Jean	0432168719	20-oct.-10	03-nov.-10
1	Salambo	Littérature	Flaubert	gustave	Ledray	Camille	0422113627	16-sept.-10	27-sept.-10
...

Limites des structures de données *plates* pour le stockage et la recherche d'informations :

- **Intégrité / cohérence des données** (doublons orthographiques, formats variables \Rightarrow problèmes pour les recherches).
- **Efficacité des recherches** (lecture séquentielle de la totalité du tableau \Rightarrow complexité en temps).

Gestion d'une bibliothèque de prêt grâce à une base de données (modèle relationnel) :

Solution : modèle relationnel

Les données sont stockées en fonction des **relations** qu'elles entretiennent entre elles.
À chaque **relation** correspond une **table** de la base de données.

Dans l'exemple ci-dessus, il paraît judicieux de prévoir une **table** « Auteurs » contenant le nom, le prénom ainsi qu'un numéro d'identification unique (id) en cas d'homonymie (ci-contre).

Auteurs		
id	nom	prénom
1	Flaubert	Gustave
2	Hugo	Victor
3	Gide	André
4	Wilde	Oscar
...

Pour communiquer et effectuer des opérations (recherches, traitements) sur les bases de données, on utilise un **système de gestion de base de données** (abr. **SGBD** et en anglais DBMS pour database management system) : SQLite, MySQL (Oracle), PostgreSQL...

Un **langage spécifique** nommé **SQL** (Structured Query Language) ou langage de requêtes structurées permet d'effectuer des **requêtes** (opérations) sur les bases de données.

Il existe des **interfaces graphiques** permettant d'effectuer ces **requêtes** dans un environnement graphique (permettant également de visualiser les tables et de les modifier) : phpMyAdmin, SQLiteStudio ...

Extrait du fichier initial :

index	titre	genre	auteurN	auteurP	nom	prénom	tél	date	retour
1	Salambo	Roman	Flaubert	Gustave	Castel	Claude	0612345112	12-août-10	23-août-10

Le fichier initial est décomposé en 5 tables : Livres, Auteurs, Emprunts, Emprunteurs et Genre.

Livres			
id	auteur	titre	genre
1	1	Salambô	1
2	2	Quatre-vingt-treize	1
3	3	L'immoraliste	1
4	4	Aphorismes	2
5	1	Madame Bovary	1
...

Auteurs		
id	nom	prénom
1	Flaubert	Gustave
2	Hugo	Victor
3	Gide	André
4	Wilde	Oscar
...

Emprunts			
qui	quoi	date	rendu
1	1	12-août-10	23-août-10
6	1	02-sept.-10	14-sept.-10
2	1	16-sept.-10	27-sept.-10
2	2	17-sept.-10	03-oct.-10
6	2	20-oct.-10	03-nov.-10
3	5	01-nov.-10	03-déc.-10
1	5	06-janv.-11	NULL
5	3	05-févr.-11	NULL
...

Emprunteurs			
id	nom	prénom	tél
1	Castel	Claude	0612345112
2	Le Dray	Camille	0422113827
3	Loutard	Annie	0654331282
4	Castel	Marie	0412324494
5	Biraud	Michèle	0434578612
6	Filâtre	Jean	0432168719
...

Genre	
id	genre
1	Roman
2	Littérature
...	...

La table Livres fait référence aux auteurs via un numéro d'identification qui renvoie à la table Auteurs. De même pour le genre.

Les autres relations sont analysées ci-dessous.

Aucune répétition : cohérence des données

Introduction au modèle relationnel

Vocabulaire

Exemple	Table formelle	Définitions																																																																				
<table border="1"> <thead> <tr> <th colspan="4">Emprunteurs</th> </tr> <tr> <th>id</th> <th>nom</th> <th>prénom</th> <th>tél</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Castel</td> <td>Claude</td> <td>0612345112</td> </tr> <tr> <td>2</td> <td>Le Dray</td> <td>Camille</td> <td>0422113827</td> </tr> <tr> <td>3</td> <td>Loutard</td> <td>Annie</td> <td>0654331282</td> </tr> <tr> <td>4</td> <td>Castel</td> <td>Marie</td> <td>0412324494</td> </tr> <tr> <td>5</td> <td>Biraud</td> <td>Michèle</td> <td>0434578612</td> </tr> <tr> <td>6</td> <td>Filâtre</td> <td>Jean</td> <td>0432168719</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>	Emprunteurs				id	nom	prénom	tél	1	Castel	Claude	0612345112	2	Le Dray	Camille	0422113827	3	Loutard	Annie	0654331282	4	Castel	Marie	0412324494	5	Biraud	Michèle	0434578612	6	Filâtre	Jean	0432168719	<table border="1"> <thead> <tr> <th colspan="4">Table</th> </tr> <tr> <th>A₁</th> <th>A₂</th> <th>A₃</th> <th>A₄</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Table				A ₁	A ₂	A ₃	A ₄																									<p>Attributs = noms des colonnes.</p> <p>n-uplet (tuple) ou enregistrement = une ligne.</p>
Emprunteurs																																																																						
id	nom	prénom	tél																																																																			
1	Castel	Claude	0612345112																																																																			
2	Le Dray	Camille	0422113827																																																																			
3	Loutard	Annie	0654331282																																																																			
4	Castel	Marie	0412324494																																																																			
5	Biraud	Michèle	0434578612																																																																			
6	Filâtre	Jean	0432168719																																																																			
...																																																																			
Table																																																																						
A ₁	A ₂	A ₃	A ₄																																																																			

Notion de clé primaire (Primary Key)

Dans les 4 tables Livres, Auteurs, Emprunteurs et Genre (mais pas dans la table Emprunts) apparaît un *attribut* noté *id* permettant d'identifier un et un seul *enregistrement* de la table. Cet attribut est la *clé primaire* qui permet une indexation des données (comme l'index d'une liste), ce qui rend les procédures d'interrogation de la table (*requêtes*) plus efficaces.

Dans la table Emprunts, la *clé primaire* est constituée des 3 *attributs* qui, quoi, date.

De façon simplifiée, une *clé* (pour une relation *r*) est un **ensemble *K* d'attributs** qui donne accès à un unique enregistrement : $\forall t, u \in r, t[K] = u[K] \Rightarrow t = u$ (i.e. si les enregistrements *t* et *u* sont identiques pour les attributs du sous ensemble *K* des attributs de la table alors ils sont identiques pour tous les autres attributs de la table).

Une *clé primaire* est une **clé de taille minimale** (souvent notée **id** ou **PK** pour **primary key**).

Une *clé secondaire* est une **clé autre que la clé primaire**.

Notion de clé étrangère (Foreign Key)

La table Emprunts possède un attribut quoi (noté Emprunts.quoi) qui fait le lien avec l'attribut id de la table Livres (noté Livres.id). Cet attribut constitue une *clé étrangère*. De même l'attribut Emprunts.qui est une *clé étrangère* faisant le lien avec l'attribut Emprunteurs.id. **Une clé étrangère fait référence à une clé primaire d'une autre table.**

Livres			
id	auteur	titre	genre
1	1	Salambô	1
2	2	Quatre-vingt-treize	1
3	3	L'immoraliste	1
4	4	Aphorismes	2
5	1	Madame Bovary	1
...

Emprunteurs			
id	nom	prénom	tél
1	Castel	Claude	0612345112
2	Le Dray	Camille	0422113827
3	Loutard	Annie	0654331282
4	Castel	Marie	0412324494
5	Biraud	Michèle	0434578612
6	Filâtre	Jean	0432168719
...

Emprunts			
qui	quoi	date	rendu
1	1	12-août-10	23-août-10
6	1	02-sept.-10	14-sept.-10
2	1	16-sept.-10	27-sept.-10
2	2	17-sept.-10	03-oct.-10
6	2	20-oct.-10	03-nov.-10
3	5	01-nov.-10	03-déc.-10
1	5	06-janv.-11	NULL
5	3	05-févr.-11	NULL
...

Avantages :

- Insertion d'une ligne dans la table Emprunts impossible avec une valeur de l'attribut quoi n'existant pas dans la table des Livres.
- Suppression d'une ligne de la table des Livres impossible si au moins une ligne de la table des Emprunts a une valeur de l'attribut quoi correspondant à la ligne à supprimer.

Graphe des clés étrangères pour les 5 tables (clés primaires en rouge) :

Emprunts			
qui	quoi	date	rendu

Emprunteurs			
id	nom	prénom	tél

Livres			
id	auteur	titre	genre

Auteurs		
id	nom	prénom

Genre	
id	genre

Rq : un tel graphe permet de déterminer l'ordre logique de création des tables : Emprunteurs, Genre et Auteurs (qui ne contiennent aucune clé étrangère) puis Livres et enfin Emprunts.