

Bases de données-Bases de données relationnelles

Les bases de données sont une des grandes réussites de l'informatique du vingtième siècle au même titre qu'Internet ou le Web

De quoi s'agit-il ?

Nous avons vu l'année dernière que lorsque les données deviennent volumineuses, nous étions obligés d'utiliser des **fichiers externes** par exemple le fichier .csv concernant les accidents de la route de plusieurs mégaoctets

Voici par exemple les premières lignes de la table des accidents de la route en France en 2018 (Au moins 130 000 lignes) La première ligne contient les descripteurs de la table

Num_Acc	place	catu	grav	sexe	trajet	secu	locp	actp	etatp	an_nais	num_veh
201800000001	1	1	3	1	0	11	0	0	0	1928	B01
201800000001	1	1	1	1	5	11	0	0	0	1960	A01
201800000002	1	1	1	1	0	11	0	0	0	1947	A01
201800000002	1	3	4	1	0	2	2	3	1	1959	A01
201800000003	1	1	3	1	5	21	0	0	0	1987	A01
201800000003	1	1	1	1	0	3	0	0	0	1977	C01
201800000004	1	1	3	1	5	2	0	0	0	2013	B01
201800000004	1	1	1	1	5	11	0	0	0	1982	A01
201800000005	1	1	4	1	5	21	0	0	0	2001	A01
201800000005	1	1	1	1	5	11	0	0	0	1946	B01
201800000006	1	1	1	1	0	3	0	0	0	1984	A01

ainsi par exemple an_nais désigne l'année naissance de la personne accidentée.

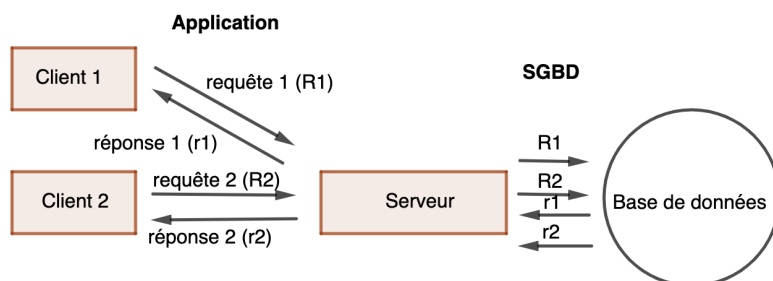
On avait alors par programmation extrait des informations d'une copie de ce fichier, **chacun de son côté**

1. Les données sont statiques au sens où elles ne sont pas mis à jour régulièrement
2. Plusieurs personnes ne peuvent pas consulter les mêmes données en même temps

Or à l'heure actuelle il est courant que , des données volumineuses et **actualisées** soient consultées en ligne par plusieurs personnes, par exemple des réservations de billets de train ou de locations de vacances ou encore des cartes (OpenstreetMap)

On peut observer d'ailleurs cette "concurrence" dans la consultation des données grâce à des messages affichées sur la page Web comme "il ne reste que 5 billets pour cette destination" ou encore "plusieurs personnes consultent cette destination"

Schématiquement le principe est le suivant : plusieurs **clients** peuvent exprimer des requêtes en ligne à un **serveur** qui est en relation avec une **base de données** gérée par un logiciel appelé **système de gestion de bases de données** (SGBD)



Définition 1. Une *base de données* est un *ensemble structuré* de données stockées sur un support persistant

Un système de gestion de base de données (SGBD) est un logiciel qui gère les données de la base de données

1 Bases de données relationnelle

Au début des années 1970 à partir de travaux sur l'algèbre et la logique sont apparus les bases de données relationnelles

De manière intuitive une base de données relationnelles est un ensemble de tableaux à deux dimensions appelés tables ou relations

Définition 2. Une *base de données relationnelle* est un *ensemble structuré de tables ou relations* stockées sur un support persistant

Un système de gestion de base de données relationnelle (SGBDR) est un logiciel qui gère les données de la base de données relationnelle

La base de données relationnelle, qui va nous servir d'exemple en cours, s'appelle Mondial Data Base est hébergée et consultable ici <http://www.semwebtech.org/sqlfrontend/>

Un autre exemple de bases de données relationnelle, celle qu'on va utiliser en TP est une base de données de Films consultable ici <http://deptfod.cnam.fr/bd/tp/requetes/> (Merci à Philippe Rigaux et au CNAM de mettre cette base en ligne)

Voici un extrait de la relation ou table

country(NAME, CODE, CAPITAL, PROVINCE, AREA, POPULATION)

qui décrit un pays par son nom, sa capitale, la province de sa capitale, sa superficie et le nombre d'habitants

NAME	CODE	CAPITAL	PROVINCE	AREA	POPULATION
China	CN	Beijing	Beijing	9596960	1427647786
Pakistan	PK	Islamabad	FCT Islamabad	803940	207776954
India	IND	Delhi	Delhi	3287590	1210854977
Indonesia	RI	Jakarta	DKI Jakarta	1919440	252124458
United States	USA	Washington	District of Columbia	9372610	318857056
Brazil	BR	Brasília	Distrito Federal	8511965	202768562

Voici un extrait de la relation

economy(COUNTRY, GDP, AGRICULTURE, SERVICE, INDUSTRY, INFLATION, UNEMPLOYMENT)

qui décrit l'économie du pays référencé par son code par un certain nombre de caractéristiques

COUNTRY	GDP	AGRICULTURE	SERVICE	INDUSTRY	INFLATION	UNEMPLOYMENT
F	2739000	1.9	79.4	18.7	0.9	10.2
H	130600	3.4	68.7	28	1.9	10.5
I	2068000	2	73.5	24.4	1.2	12.4
EW	24280	3.9	66.2	30	3.4	10.9
LT	46710	3.7	68	28.3	1.2	12.4
PL	513900	4	62.7	33.3	1	10.3
SLO	46820	2.8	68.3	28.9	1.8	13.1
BG	53700	6.7	63	30.3	1.5	11.6
IRL	220900	1.6	70.4	28	0.6	13.5
ET	262000	14.5	48	37.5	9	13.4
SA	718500	2	35.5	62.5	3.7	10.5

Le vocabulaire est important dans le domaine des bases de données aussi on va préciser les termes employés :

1. Un **n-uplet** (une ligne d'un tableau) est une fonction avec des paramètres qu'on appelle ici **attributs**
2. Une **relation** est définie comme l'ensemble des nuplets (lignes)
3. Chaque attribut a un **type** ainsi qu'un **domaine**, ainsi l'attribut NAME est de type string, l'attribut POPULATION est de type int et l'attribut UNEMPLOYMENT de type float
4. Le **domaine** de l'attribut POPULATION est l'intervalle $[[0; 2\ 000\ 000\ 000]]$ et le domaine de l'attribut UNEMPLOYMENT est l'intervalle $[0,100]$ (pourcentage)
5. Un attribut (ou plusieurs) attributs qui identifie (ent) de **manière unique** un nuplet est appelé **clé primaire**, par exemple CODE, ainsi que NAME, CODE est une clé primaire pour la relation country (NAME, CODE, CAPITAL, PROVINCE, AREA, POPULATION)
6. Le **schéma relationnel** d'une base de données est l'ensemble des relations d'une base, voici par exemple ci-dessous le schéma relationnel de la base de données **Films** sur laquelle on va travailler en TP

Schéma de la base

Film (**idFilm**, titre, année, genre, résumé, *idRéalisateur*, *codePays*)

Pays (**code**, nom, langue)

Artiste (**idArtiste**, nom, prénom, annéeNaiss)

Rôle (**idFilm**, **idActeur**, nomRôle)

Internaute (**email**, nom, prénom, région)

Notation (**email**, **idFilm**, note)

7. On observe en gras les clés primaires des relations
Par exemple, **idFilm** est une clé primaire de la relation **Film**
8. Si on veut relier un film à un réalisateur, qui est aussi un artiste on va passer par l'attribut *idRéalisateur* de la relation **Film**
Ainsi un film qui aura un attribut *idRéalisateur* égal à la clé primaire **idArtiste** d'un artiste, est un film réalisé par cet artiste
On dit alors que *idRéalisateur* est une **clé étrangère** de la relation **idFilm**

2 Exercices : Base de données relationnelle

Ex 1

Voici une description de la table ou relation **Population** de la base de données relationnelle Mondial Database

Population: information about the population of the countries.

country: the country code

population-growth: population growth rate (per annum)

infant_mortality: infant mortality (per thousand)

1. Quels sont les attributs ?
2. Quels sont leur types ? les domaines ?
3. Quelle est la clé primaire ?

Ex 2

Une bibliothèque a une base de données composées de trois tables **Auteurs**, **Emprunteurs** et **Emprunts**

1. La relation **Emprunteurs** est définie par **Emprunteurs**(*idEmp*, *Nom*, *Prénom*, *DateNaissance*). Pourquoi ni (*Nom*, *Prénom*), ni (*Nom*, *Prénom*, *DateNaissance*) peuvent constituer des clés primaires ?
2. Quels sont les attributs possibles de la relation **Auteurs** ?
3. Quels sont les attributs possibles de la relation **Emprunts** ?

Ex 3

Voici le schéma relationnel d'une base de données

Voyageur (**idVoyageur**, nom, prénom, ville, région)

Séjour (**idSéjour**, *idVoyageur*, *codeLogement*, début, fin)

Logement (**code**, nom, capacité, type, lieu)

Activité (**codeLogement**, **codeActivité**, description)

1. Combien y-a-t-il de relations ?
2. Citer toutes les clés primaires
3. Citer toutes les clés secondaires et expliquer leur utilité

Ex 4

Voici le schéma relationnel d'une base de données

Contact (**idContact**, nom, prénom, email)

Message (**idMessage**, contenu, dateEnvoi, *idEmetteur*, *idPrédecesseur*)

Envoi (**idMessage**, **idContact**,)

1. Combien y-a-t-il de relations ?
2. Quels sont les attributs de la relation Contact ?
3. Quel est le type de l'attribut contenu de la relation Message ?
4. Quelle est la clé primaire de la relation Envoi ?
5. Quelle est la nature de l'attribut *IdEmetteur* ? A quoi sert il ?