

BASES DE DONNÉES I

INTRODUCTION AUX SYSTÈMES DE GESTION DE BASE DE DONNÉES

Laurent Kaczmarek

PCSI² 2013-2014
Lycée Louis Le Grand

Lundi 12 mai 2014

**BASES DE
DONNÉES I**
INTRODUCTION
AUX SYSTÈMES
DE GESTION DE
BASE DE
DONNÉES

**LAURENT
KACZMAREK**

INTRODUCTION

PARTAGER DES
DONNÉES

STRUCTURER DES
DONNÉES

LES SYSTÈMES
DE GESTION DE
BASES DE
DONNÉES

ET
CONCRÈTEMENT ?

I

INTRODUCTION

**BASES DE
DONNÉES I**

INTRODUCTION
AUX SYSTÈMES
DE GESTION DE
BASE DE
DONNÉES

**LAURENT
KACZMAREK**

INTRODUCTION

PARTAGER DES
DONNÉES

STRUCTURER DES
DONNÉES

LES SYSTÈMES
DE GESTION DE
BASES DE
DONNÉES

ET
CONCRÈTEMENT ?

PRINCIPALES UTILISATIONS DES ORDINATEURS

- ▶ Les ordinateurs ont été initialement créés *pour calculer*.
- ▶ Mais, depuis l'invention des disques durs en 1956, les ordinateurs permettent aussi de *collecter, classer et stocker* de grandes quantités d'informations.

LES BASES DE DONNÉES

- ▶ Le terme *database* (base de données) est apparu en 1964 pour désigner une collection d'informations partagées par différents utilisateurs d'un même réseau.
- ▶ Les premières utilisations furent militaires.

LES BD SONT INCONTOURNABLES

Les bases de données sont très largement utilisées, de l'échelle la plus petite (accès local, par les employés d'une entreprise par exemple) à la plus grande (accès global, par n'importe quel internaute par exemple).

DEUX POINTS FONDAMENTAUX

- ▶ Le partage à grande échelle de données implique certaines précautions (pour garantir la pérennité des données par exemple).
- ▶ Pour traiter un grand nombre de données, il faut également réfléchir à la meilleure manière de les structurer afin de faciliter les algorithmes de recherche.

PLAN DU COURS

- I. Introduction.
- II. Partager des données.
- III. Structurer des données.
- IV. Les SGDB.
- V. Et concrètement ?

II

PARTAGER DES DONNÉES

QUELQUES EXEMPLES

- ▶ La *Bibliothèque Nationale de France* possède un catalogue impressionnant d'ouvrages enregistrés dans une base de données. Elle peut être consultée par les utilisateurs qui peuvent par exemple effectuer des recherches thématiques : quelles sont les ouvrages disponibles écrits par tel écrivain ? quelles sont les éditions disponibles ? quelle est la date d'acquisition de telle édition ? etc. Les bibliothécaires ont un droit supplémentaire : modifier les données (ajouter des données, en supprimer, les altérer).
- ▶ En achetant sur *Amazon*, un client consulte une base de données constituée des descriptions des biens vendus, des prix, des vendeurs, des stocks et des commentaires des acheteurs. Un achat modifie automatiquement ces données (l'état du stock par exemple). Les acheteurs ont également accès aux données et peuvent modifier certaines données (prix, stocks, etc).

QUELQUES EXEMPLES (SUITE)

- ▶ Les données des stations météorologiques sont enregistrées dans des bases de données puis traitées par les ingénieurs de *Météo-France*.

UN PARTAGE QUI IMPLIQUE DES RÈGLES

- ▶ Les données doivent être facilement accessibles par un grand nombre de clients et il faut garantir une certaine pérennité des données.
- ▶ Afin de protéger les données, il convient de respecter les points suivants :
 - ▶ Les utilisateurs ne manipulent pas directement les fichiers.
 - ▶ Les utilisateurs ne s'occupent pas de la façon dont sont stockées les informations, ni où elles le sont.
- ▶ Les utilisateurs doivent pouvoir collecter, fournir ou modifier des données sans programmer. Ceci impose l'utilisation d'un *langage de description de requêtes*.

III

STRUCTURER DES DONNÉES

LES LIMITES DES STRUCTURES PLATES

- Considérons le problème suivant : créer un catalogue des œuvres exposées dans les musées de la RMN.
- Les œuvres et les musées ont des attributs propres : nom et auteur pour les œuvres, nom et ville pour le musée.
- On peut représenter le catalogue par une liste de musées contenant chacune une liste d'œuvres :

```
catalogue=[  
    ["Louvre","paris",  
     [ ["La Joconde","De Vinci"],  
       ["Gilles","Watteau"],...]],  
    ["Palais des beaux-arts","Lille",  
     [ ["Les vieilles","Goya"],...]]  
    ...  
]
```

LES LIMITES DES STRUCTURES PLATES (SUITE)

- ▶ Il est facile de trouver la liste des œuvres d'un musée donné au moyen d'un parcours en temps linéaire de la liste **catalogue**.
- ▶ La détermination de la liste de toutes les œuvres exposées dans les musées de la RMN d'un peintre donnée est un peu plus longue à obtenir car nécessite un parcours linéaire de **catalogue** au sein duquel il faut parcourir linéaire la liste des œuvres de chacun des musées (ce que l'on peut réaliser au moyen de deux boucles for imbriquées).
- ▶ La situation serait encore plus délicate dans le cas de la recherche des œuvres portant le même nom.
- ▶ La manière dont est structuré **catalogue** induit une disymétrie dans le traitement des données : l'appartenance d'une œuvre à un musée est privilégiée au lien entre l'œuvre et son auteur.

LES LIMITES DES STRUCTURES PLATES (FIN)

- ▶ On aurait pu bien sûr structurer inversement : une liste d'auteurs contenant une liste d'œuvre avec un attribut musée. Dans ce cas, les difficultés sont inversées, il est plus difficile d'obtenir la liste des œuvres d'un musée que la liste des œuvres d'un artiste.
- ▶ Les bases de données permettent de simplifier et d'améliorer l'exploration de données en les structurant de manière à ne privilégier aucun lien.

MAIS ALORS, COMMENT STRUCTURER LES DONNÉES ?

- ▶ Depuis les années soixante, plusieurs modèles ont été créés. Ils correspondent à des manières différentes de *structurer des données*.

IV

LES SYSTÈMES DE GESTION DE BASES DE DONNÉES

CAHIER DES CHARGES D'UN SGBD

- ▶ La gestion et l'accès à une base de données sont assurés par un ensemble de programmes qui constituent un Système de gestion de Bases de Données (SGBD).
- ▶ Un SGBD doit permettre l'ajout, la modification et la recherche de données.
- ▶ Un SGBD doit masquer la représentation physique des données et assurer la cohérence et la protection des données dans le contexte d'un partage des données entre un grand nombre d'utilisateurs.

EXEMPLES DE SGBD

Les principaux acteurs dans le secteur des SGBD commerciaux sont Oracle Corporation (*Oracle*), IBM (*DB2*) et Mycrossoft (*SQL-Server*).

TROIS NIVEAUX ET UN MODÈLE

La réponse apportée à cette double contrainte liée au partage et à la structure des données caractérise un SGBD.

- ▶ Les SGBD sont conçus selon l'*architecture Ansi/Sparc* développée par Bachman vers 1965. Elle repose sur une division en trois niveaux : *physique* (là où sont stockées les données), *logique* (la manière dont elles sont structurées) et *externe* (l'interface avec l'utilisateur).
- ▶ Un SGBD est caractérisé par le modèle de description des données qu'il supporte.

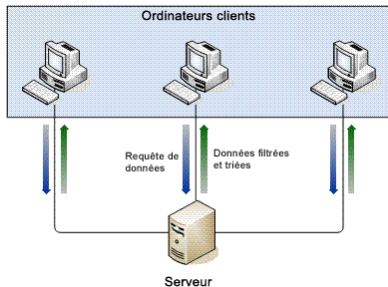
RAPIDE HISTORIQUE DES MODÈLES

- ▶ *le modèle hiérarchique*, vers 1960.
- ▶ *le modèle relationnel* créé par Edgar F.Codd, ingénieur chez IBM, en 1970.
- ▶ *le modèle objet-relationnel* au début des années 90.

Le modèle relationnel est très largement utilisé de nos jours et figure au programme.

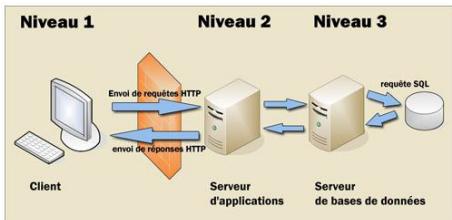
V ET CONCRÈTEMENT ?

ARCHITECTURE CLIENT/SERVEUR



- ▶ Les clients accèdent directement à la base de données (ce sont le plus souvent des programmes que des opérateurs humains, cf. les sites de vente en ligne).
- ▶ Les communications passent le plus souvent par l'intermédiaire d'un réseau (par exemple Internet).
- ▶ L'inconvénient de cette architecture est le lien direct des utilisateurs aux données (bas niveau de sécurité).

ARCHITECTURE TROIS-TIERS



- ▶ La terminologie est un anglicisme, il faut comprendre *architecture à trois étages* (3-tier in english) :
 - ▶ Un client équipé d'une interface chargée de la présentation.
 - ▶ Un serveur d'application (appelé middleware) qui fournit la ressource, mais en faisant appel à un autre serveur.
 - ▶ Un serveur de données qui fournit au middleware les données requises pour répondre au client.
- ▶ Cette architecture renforce la sécurité des données en supprimant le lien avec le client. Le serveur applicatif a un rôle de traducteur, de gendarme et de filtre.

GESTION ET INTERROGATION D'UNE BD

- ▶ La création, l'administration et l'interrogation d'une base de données de taille modeste peut se faire via une application.
- ▶ Il existe de nombreuses applications *MySQL*, *PhpMyAdmin*, etc.
- ▶ Afin d'illustrer ce cours, nous utiliserons *Sqlliteman*, application gratuite.
- ▶ Ouvrons par exemple une base de données disponible sur le site www.data.gouv.fr, la liste des prénoms donnés à Paris de 2004 à 2013.
- ▶ Les données sont structurées en quatre colonnes : *prénom*, *nombre*, *sexe* et *année*.
- ▶ L'utilisateur peut modifier et/ou interroger la base en programmant en *SQL* dans une des fenêtres.

MEET SQLITEMAN

The screenshot shows the Sqliteman application window titled "PrenomsAParis - Sqliteman". The interface includes a "Schema" tab, a "Pragmas" tab, and a "Base de données" sidebar. The sidebar shows a tree structure with "main" containing "Tables (1)" (highlighted as "prénoms parisiens"), "Vues (0)", and "Catalogue système (1)".

The main query editor displays the following SQL query:

```
1 SELECT SUM(nombre) FROM prénums parisiens WHERE prenom='Louise';
```

Below the query editor, the execution status is shown: "Durée: 0.075 secondes" and "Col: 42 Ligne: 1/1".

The results are displayed in a table view, showing the following data:

	prenum	nombre	sexe	annee
1	Liz	5	F	2012
2	Lohan	9	M	2012
3	Lou	8	F	2012
4	Louise	306	F	2012
5	Luc	28	M	2012
6	Lucie	107	F	2012
7	Lucille	8	F	2012

At the bottom, a status bar indicates: "Requête OK" and "Ligne(s) retournée(s): 256 (Plus de lignes peuvent être récupérées. Faites défiler les résultats pour obtenir plus de lignes et/ou lisez la documentation)".

**BASES DE
DONNÉES I**
INTRODUCTION
AUX SYSTÈMES
DE GESTION DE
BASE DE
DONNÉES

**LAURENT
KACZMAREK**

INTRODUCTION

PARTAGER DES
DONNÉES

STRUCTURER DES
DONNÉES

LES SYSTÈMES
DE GESTION DE
BASES DE
DONNÉES

ET
CONCRÈTEMENT ?

Dans la pratique, il est beaucoup plus facile de modifier la base de données en utilisant les différents boutons, on n'utilise le langage *SQL* que pour formuler des requêtes.