

Bases de données et systèmes de gestion de bases de données

Master IPH - EC IPH73C - 2013

1

Exemple accessible via une interface Web

- Une base de données consultable en ligne :

The Trans-atlantic slave trade database:

<http://www.slavevoyages.org/tast/index.faces>

Voir notamment « The African Name Database » et le lien avec la vue sur les voyages

Master IPH - EC IPH73C - 2013

2

Généralités

- Fichiers manuels : idée de la boîte de fiches
- Tenir à jour des fichiers manuels : une activité « vieille comme le monde »
- Possibilités nouvelles offertes par l'informatique, mais aussi nouvelles contraintes

Master IPH - EC IPH73C - 2013

3

Définitions

- (Wikipedia) Une **base de données** est un lot d'informations stockées dans un dispositif informatique
- (Une définition plus précise) Une **base de données** est un ensemble de données structurées, fiables, cohérentes, partagées par des utilisateurs ayant des besoins différents.
- Un **système de gestion de bases de données (SGBD)** est un logiciel permettant de gérer et de manipuler une base de données.

Master IPH - EC IPH73C - 2013

4

- Préoccupations majeures :
 - Cohérence et consistance des données
 - Non redondance, car la redondance est source d'incohérences
 - Intégrité des données, pas de données inaccessibles
- Une base de données s'appuie sur un modèle logique :
 - Ce qui est stocké dans la base de données est vrai
 - Ce qui ne s'y trouve pas est faux (hypothèse du monde clos)
- Exemples : *base de données des horaires de la SNCF, catalogue d'une bibliothèque*

Master IPH - EC IPH73C - 2013

5

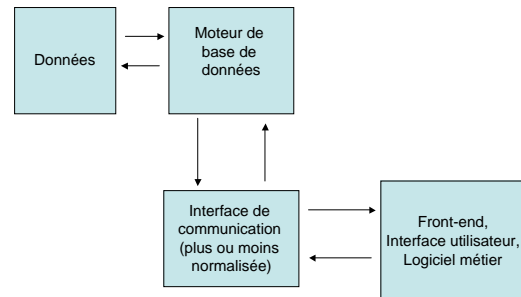
Bases de données : des besoins et des objectifs divers

- Systèmes documentaires : gros volumes de données, informations stables,
- Bases de données d'annuaires : le carnet d'adresse de votre logiciel de mail, l'annuaire LDAP de l'UBO
- D'autres exemples : Apogée, logiciels de comptabilité/gestion, horaires SNCF, etc : utilisent des SGBDR (systèmes de gestion de bases de données relationnelles)

Master IPH - EC IPH73C - 2013

6

- Idées essentielles présentes dans la définition:
 - Stockage des données indépendant des traitements
 - Utilisateurs ayant des besoins différents, se connectant à l'aide d'outils variés
 - Différents outils à disposition des utilisateurs : vues sur la base de données, front-end,



Exemples de SGBDR (ou presque relationnels)

- Des outils comportant un « front end » intégré : Microsoft Access, Filemaker Pro, OpenOffice Base, 4^{ème} Dimension
- Moteurs de bases de données : Oracle, Microsoft SQL Server, Interbase, MySQL, SQLite, etc
- Plusieurs composants assemblés : MySQL + PHP MyAdmin
- Outils d'interrogation sans moteur de base de données: Microsoft Query, logiciels de Statistiques tq Statistica
- Logiciels « métiers » s'appuyant sur un SGBDR : logiciels de traitement d'enquête tels que Sphinx ou Modalisa

SGBDR

- Un système de gestion de bases de données relationnel est un SGBD permettant de manipuler le contenu de bases de données relationnelles.
- Une base de données relationnelles est une BD organisée selon les principes de l'algèbre relationnelle.
- Algèbre relationnelle : développée notamment par Edgar F. Codd entre 1965 et 1975

Modèle relationnel - Algèbre relationnelle

- Un seul type d'objet : les tables ou relations
- Schéma d'une relation, extension de la relation
- Opérations sur les tables : ajout ou suppression de n-uplets, sélection, projection, jointure
- Créer une relation (ou table) : lui donner un nom, et définir ses attributs et leurs types
- L'un des attributs a un rôle particulier : identifiant ou clé, clé primaire

Schéma de la table « Mes Amis »

Mes Amis (Nom, Rue, Ville, CP, Téléphone)

Extension de la table :

Nom	Rue	Ville	CP	Tél
Dupond	Rue du Lac	Brest	29200	43 44 45
Duval	Place de la Gare	Brest	29200	44 41 80
Durand	Rue du Lac	Quimper	29000	42 35 56
Martin	Rue de Brest	Morlaix	29300	67 89 43

Projection sur un ensemble d'attributs : ne conserver que les colonnes correspondant à ces attributs et supprimer les doublons éventuels.

Exemple : « Les villes dans lesquelles j'ai des amis »

Ville	CP
Brest	29200
Quimper	29000
Morlaix	29300

Master IPH - EC IPH73C - 2013

13

Sélection selon une condition: ne conserver que les n-uplets satisfaisant la condition.

Exemple : sélectionner « mes amis brestoïis ». La condition est alors: Ville=Brest.

Nom	Rue	Ville	CP	Tél
Dupond	Rue du Lac	Brest	29200	43 44 45
Duval	Place de la Gare	Brest	29200	44 41 80

Master IPH - EC IPH73C - 2013

14

Jointure entre deux tables sur un attribut commun : schéma obtenu par « réunion » des deux schémas. N-uplets obtenus en concaténant les n-uplets de chacune des deux tables pour lesquels l'attribut a la même valeur.

Bâteau (Id-Bat, Nom, Nationalité)

Id-Bat	Nom	Nationalité
1	Formiga	Portugal
2	Jeune Louis	France
3	Xerxes	Espagne

Voyage (Id-Voyage, Date, Bateau)

Id-Voyage	Date	Bateau
V1	1821	1
V2	1825	2
V3	1826	2

Master IPH - EC IPH73C - 2013

15

Jointure sur les attributs Id-Bat et Bateau :

Id-Voyage	Date	Bateau=Id-Bat	Nom	Nationalité
V1	1821	1	Formiga	Portugal
V2	1825	2	Jeune Louis	France
V3	1826	2	Jeune Louis	France

Master IPH - EC IPH73C - 2013

16

Autres opérations : réunion, intersection, etc.

Comportement particulier de toutes ces opérations vis-à-vis des valeurs vides ou manquantes :

Un texte manquant est en général différent du texte vide « »

Une valeur numérique vide n'est pas 0

Par exemple, dans une jointure, que fait-on des lignes de l'une des tables qui n'ont pas de ligne correspondante dans l'autre table ?

jointure interne : elles ne figurent pas dans le résultat

jointure externe : elles y figurent

Master IPH - EC IPH73C - 2013

17

Structured Query Language (SQL)

Un langage conforme aux règles de l'algèbre relationnelle, utilisé par les moteurs des SGBDR, assez bien standardisé : SQL.

Exemples de requêtes :

```
CREATE TABLE truc (...);
INSERT into truc VALUES (...);
SELECT champ1, champ2, ... FROM truc;
JOIN ...
```

Master IPH - EC IPH73C - 2013

18

Concevoir une base de données

- On a circonscrit un domaine du monde réel.
- Les objets du monde réel sont représentés par des données
- Notre problème :

Comment structurer ces données pour les inclure dans une base de données satisfaisant aux critères de consistance, cohérence, non-redondance ?

Quelle structure conceptuelle pour nos données ?

Quelle méthode de modélisation appliquer ?

Master IPH - EC IPH73C - 2013

19

Première méthode : algèbre relationnelle, dépendances fonctionnelles et théorie des formes normales

- On forme une table rassemblant l'ensemble des données manipulées.
- On utilise des règles (formes normales) pour scinder certaines colonnes, décomposer la table initiale en plusieurs tables plus petites, liées par des jointures, introduire des clés dans chacune de ces tables, etc.

- *Dépendances fonctionnelles* : B dépend fonctionnellement de A si la connaissance de A induit celle de B.

- *Formes normales*

Première forme normale : pas d'attribut multivalué

Deuxième forme normale : tout attribut non clé dépend fonctionnellement de la (ou des) clés

Master IPH - EC IPH73C - 2013

20

Deuxième méthode : modèle entités-associations

Peter Chen, Hubert Tardieu (1970, 1980)

- « Schématiser » les informations du monde réel, les structurer, avec un double objectif :
 - Respecter leur sémantique, leur signification dans le monde réel
 - Permettre la traduction de cette schématisation sous forme d'un schéma de base de données.
- Résultat : un MCD ou modèle conceptuel des données

Pour illustrer la démarche : l'exemple que nous traiterons ensuite sous Access : L'enquête agricole de 1852.

Lien : <http://acrh.revues.org/index3696.html>

Master IPH - EC IPH73C - 2013

21

Informations et données

- Dans le monde réel : des informations, contenues dans des rubriques de documents.

Voir, pour notre exemple :

<http://acrh.revues.org/3697?file=1>

- Donnée : forme « normalisée » d'une information présente dans une rubrique de document.

Master IPH - EC IPH73C - 2013

22

« Qualités » d'une donnée

- Donnée élémentaire ou donnée composée : nom, prénom sont des données élémentaires, identité (sous la forme nom + prénom) est une donnée composée.
- Donnée de base ou donnée calculée: une donnée calculée peut être retrouvée à partir d'autres données, à l'aide d'un « calcul ».
- Ne pas confondre donnée calculée et donnée numérique

Master IPH - EC IPH73C - 2013

23

Types de données

- Type « texte de longueur fixe » ou alphanumérique
- Type « nombre » (entier, nombre réel, ...)
- Type « texte de longueur libre » ou « memo »
- Type booléen (vrai/faux, oui/non)
- Type « date » ou « date / heure »
- Type « objet externe » ou binaire (par exemple un fichier, une image, etc)

Master IPH - EC IPH73C - 2013

24

Dictionnaire des données

- Recensement des données manipulées dans le domaine étudié : attribuer à chaque donnée un nom, un type, une description -> constituer un dictionnaire des données.

Sur notre exemple : utiliser les 2 fichiers descriptif-detaille.xls et Enquete-Agricole-Cereales.xls

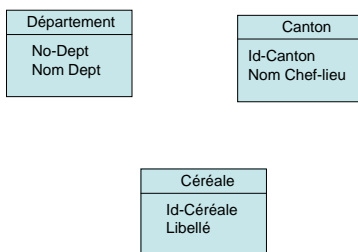
Données et entités

Certaines données se regroupent naturellement car elles correspondent à un même objet du monde réel : ce sont des attributs d'une même entité.

Sur notre exemple :

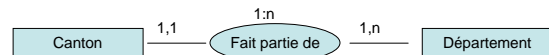
- Des départements
- Des arrondissements ou cantons
- Des céréales
- (peut-être) des productions

Représentation des entités

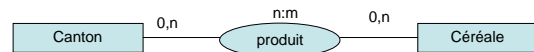


Associations entre entités

Un canton fait partie d'un département :

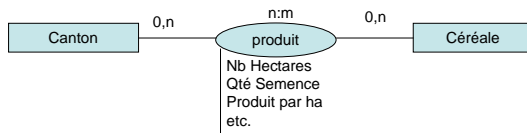


Un canton produit des céréales :



Attributs d'associations

Un canton produit des céréales :



- Une **entité** est une collection homogène d'occurrences d'entité.
- Une **occurrence d'entité** est un exemplaire différenciable et identifiable d'une chose, d'une personne ou d'un concept abstrait pour lequel on doit gérer des informations significatives
- Les liens entre les entités, qui s'expriment souvent par des verbes, sont modélisés par des **associations** entre les entités.

Cardinalités des associations

- 1,1 - 1,1 ou 1,1 - 0,1 : un à un
- 1,1 – 1,n : hiérarchique totale
- 1,1 – 0,n ou 0,1 – 1,n ou 0,1 – 0,n : hiérarchique partielle
- 1,n – 1,n ou 0,n – 1,n ou 0,n – 0,n : plusieurs à plusieurs

Validation du MCD

- Le MCD représente-t-il correctement le monde réel ?
- Toutes les données du dictionnaire des données figurent-elles dans le MCD ?
- Chaque donnée apparaît-elle bien *une seule fois* dans le MCD ?

Traduction du MCD dans le modèle relationnel

- A chaque entité correspond une table
- Une association « plusieurs à plusieurs » ou une association faisant intervenir plus de 2 entités se traduit par une table ayant pour attributs :
 - Les clés des entités participantes
 - Les attributs éventuels de l'association
- Une association hiérarchique ou « un à un » peut souvent se traduire par l'ajout de la clé de l'entité du côté « n » à la table correspondant à l'entité du côté « un ».

Vocabulaire

Algèbre relationnelle	SGBD (Access, OOBbase, etc)	Modèle entités-associations
Relation ou table	Table	Entité – certaines associations
N-uplet	Enregistrement	Occurrence d'entité
Jointure	Relation	Certaines associations
	Type de relation	Cardinalités