

E6 - Compétence

Base de données



Introduction	3
Présentation du concept	4
Fonctionnement d'une base de données.	5
A quoi servent-elle	6
MCD	7
Modèle E/A	8
Schéma relationnel	10
Les cardinalités	11
Le modèle relationnel	12
3 formes normale d'optimisation de base de données.	12
1FN : première forme normale :	12
2FN : Deuxième forme normale :	13
3FN : troisième forme normale :	14
Avantages et inconvénients	15
Avantage :	15
Inconvénient :	15
Conclusion	15
Bibliographie	16

Introduction

L'histoire des bases de données remonte aux années 1960, avec l'apparition des bases de données réseau et des bases de données hiérarchiques. Dans les années 1980, ce sont les bases de données orientées objet qui a fait leur apparition. Aujourd'hui, les bases de données prédominantes sont les SQL et NoSQL.

Les disques durs, mémoire de masse de grande capacité, ont été inventés en 1956. L'invention du disque dur a permis d'utiliser les ordinateurs pour collecter, classer et stocker de grandes quantités d'informations de façon plus souple et plus performante que la bande magnétique.

Les premières bases de données hiérarchiques sont apparues au début des années 1960. Les informations étaient découpées en deux niveaux de hiérarchie : un niveau contenait les informations qui sont identiques sur plusieurs enregistrements de la base de données. Le découpage a ensuite été étendu pour prendre la forme d'un diagramme en arbre.

En 1965, Charles BACHMAN conçoit l'architecture Ansi/Sparc encore utilisée de nos jours. En 1969, il créa le modèle de données réseau au sein du consortium CODASYL pour des applications informatiques pour lesquelles le modèle hiérarchique ne convient pas. Charles Bachman a reçu le prix Turing en 1973 pour ses « contributions exceptionnelles à la technologie des bases de données ».

En 1968, Richard Pick crée Pick, un système d'exploitation contenant un système de gestion de base de données « multivaluée » (SGBDR MV).

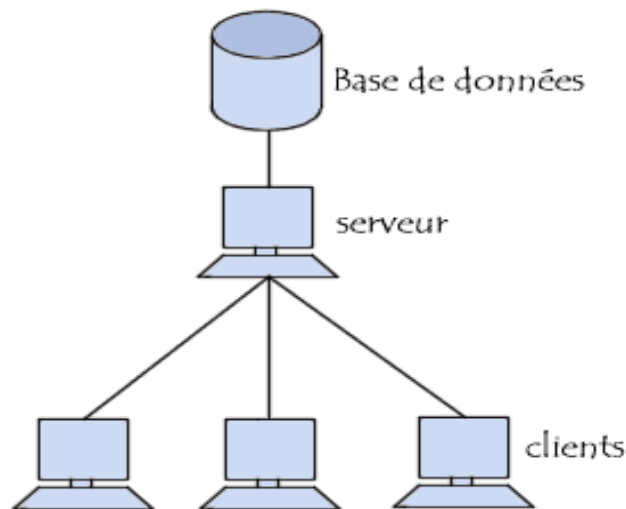
Apparues dans les années 1990, les bases de données objet-relationnel utilisent un modèle de données relationnel tout en permettant le stockage des objets. Dans ces bases de données les associations d'héritage des objets s'ajoutent aux associations entre les entités du modèle relationnel.

Présentation du concept

Une base de données est une entité dans laquelle, nous pouvons stocker des données de façon structurée, organisé et avec le moins de redondance possible.

Ces données doivent pouvoir être utilisées par des programmes, par des utilisateurs différents. Ainsi, la notion de base de données est généralement couplée à celle de réseau, afin de pouvoir mettre en commun ces informations.

On parle généralement de système d'information pour désigner toute la structure regroupant les moyens mis en place pour pouvoir partager des données.



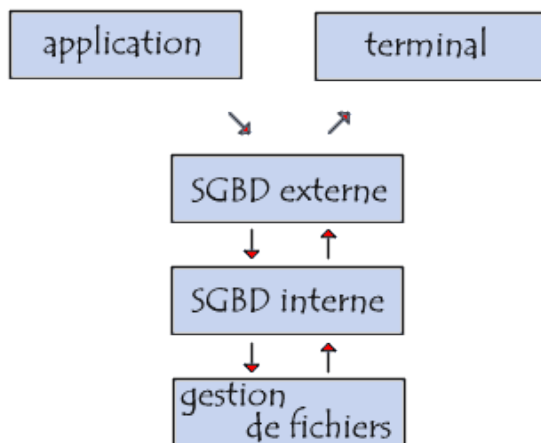
Fonctionnement d'une base de données.

Les bases de données sont stockées sous forme de fichiers ou d'ensemble de fichiers sur un serveur de stockage. Les bases de données traditionnelles (hiérarchiques) sont organisées par champs (fields). Un champ est une seule pièce d'information. Un enregistrement est un ensemble de champs. Un fichier est une collection d'enregistrements.

La faculté de consulter ou de modifier une database (lire ou écrire) est conférée aux divers utilisateurs par un database managé (gestionnaire de base). Les bases de données sont principalement présentes au sein des systèmes avec des machines puissantes.

La gestion de la base de données se fait grâce à un système appelé **SGBD** (système de gestion de bases de données) ou en anglais DBMS (Database management system). Le SGBD est un ensemble de services (applications logicielles) permettant de gérer les bases de données, c'est-à-dire :

- permettre l'accès aux données de façon simple
- autoriser un accès aux informations à de multiples utilisateurs
- manipuler les données présentes dans la base de données (insertion, suppression, modification)



Le SGBD peut se décomposer.

En trois sous-systèmes :

- le système de gestion de fichiers : il permet le stockage des informations sur un support physique.
- le SGBD interne : il gère l'ordonnancement des informations.
- le SGBD externe : il représente l'interface avec l'utilisateur.

Principaux SGBD :

- | | |
|------------------------|---------------------------------------|
| • Interbase | Datant de 1991. Dernière version 2017 |
| • Microsoft SQL server | Datant de 1989. Dernière version 2017 |
| • Microsoft Access | Datant de 1992. Dernière version 2016 |
| • Oracle | Datant de 1972. Dernière version 2016 |
| • MySQL | Datant de 1995. Dernière version 2018 |
| • SQL Server 11 | Datant de 1989. Dernière version 2017 |

A quoi servent-elle

La base de données permet un échange d'information entre l'utilisateur et le serveur. Elle permet d'envoyer uniquement les données qu'un utilisateur a besoin, cela permet à un utilisateur de s'enregistrer sur un site, créer un compte, modifier une information le concernant, ajoute une information, supprimé les données qu'il a plus renseigné sur un site lors de son inscription. Elle reste invisible d'accès aux yeux de l'utilisateur.

En arrière-plan, un informaticien se repose très souvent sur une base de données. Grâce à elle, l'information peut être organisée, interrogée et améliorée en toute fiabilité.

Cet outil permet de faire le tri, de retrouver la bonne info au bon moment et de pouvoir déverser des quantités toujours plus importantes de « data » dans une « armoire toujours bien rangée » est primordial. C'est à tout cela que sert une base de données.

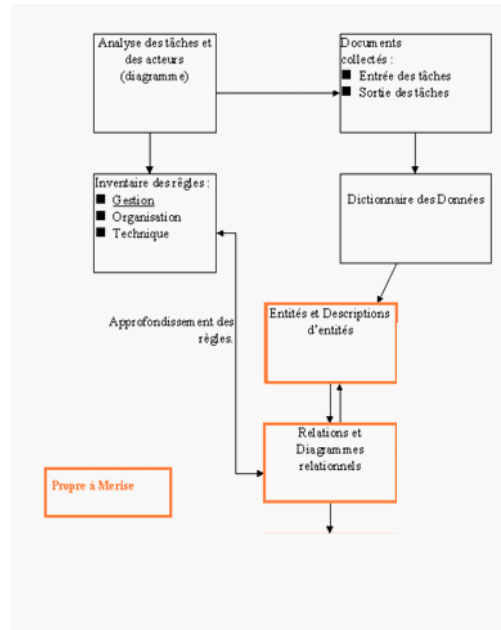
C'est pour cela que derrière le défi technique (volume, synchronisation, autorisations, calculs complexes), les bases de données sont si importantes. Elles sont le support d'une information à laquelle on doit pouvoir accéder en toute confiance. Les données doivent être justes, leurs évolutions (ajout/modification/suppression/affichage) doivent pouvoir être tracées (qui a fait l'action, pourquoi, dans quel contexte) et leur intégrité doit être à toute épreuve.

Cet aspect de confiance est fondamental et c'est pour cela qu'on ne construit pas une base de données sur un coup de tête, pour avoir un outil qui soit fiable et ainsi qui permet un bon fonctionnement, construire une base de données ne s'improvise pas. C'est à ce prix que l'on peut déléguer l'information aux machines et se baser dessus pour les valoriser.

MCD

Il s'agit de l'élaboration du modèle conceptuel des données (MCD) qui est une représentation graphique et structurée des informations mémorisées par un SI (Système informatique). Le MCD est basé sur deux notions principales, d'un côté les entités et de l'autre les associations, d'où sa seconde appellation : le schéma Entité/Association.

L'élaboration du MCD passe par les différentes étapes :



- La mise en place de règles de gestion. Le but étant de décrire par une phrase comment sont associés les deux concepts. Exemple est-ce qu'une personne est obligatoirement chef de projet au moins une fois ?
- L'élaboration du dictionnaire des données. c'est une collection de données de référence nécessaire à la conception. Il décrit des données d'importantes que les clients, les nomenclatures de produits et de services, les annuaires, etc.
- L'élaboration du MCD (création des entités puis des associations puis ajout des cardinalités). Il nous est donc possible d'élaborer le MCD complet à partir des données présentes dans le dictionnaire des données :

- Souvent, pour un même ensemble de règles de gestion, plusieurs solutions sont possibles au niveau conceptuel.
- Pour que le MCD soit valide, toute entité doit être reliée à au moins une association.
- Les entités et les propriétés peuvent être historiées. (cela permet de préciser que l'on archivera toutes les modifications sur une entité ou une propriété donnée). Cela doit également répondre à une règle de gestion.

Modèle E/A

Le modèle entité-association a été inventé par Peter Chen en 1975. Il est destiné à clarifier l'organisation des données dans les bases de données relationnelles.

Dans le modèle relationnel, la relation désigne l'ensemble des informations d'une table, tandis que l'association, du modèle entité-association, désigne le lien logique qui existe entre deux tables contenant des informations connexes.

Une entité a une existence autonome et elle est identifiable sans ambiguïté.

Exemples :

- L'étudiant « Valentin »
- Le journal « La Presse »
- Le cinéma « ABC »

Nom de la classe d'Entités
Attribut 1
Attribut 2
Attribut 3

Entité :

Dans un processus de modélisation on ne s'intéresse pas à chaque entité séparément mais à un type d'entité. Classe d'entité : c'est un regroupement d'entités de même nature.

Exemples : employés représente l'ensemble des employés.

Propriété :

Information ayant des valeurs qualifiant une occurrence d'une entité.

Une propriété ne doit dépendre que de la seule entité à laquelle elle appartient.

Nom de la classe d'Entités

Un attribut peut-être:

- Atomique ou élémentaire : comporte un seul champ (Prénom, Nom, ...).
- Composé : peut être divisé en sous parties d'attributs élémentaires ayant des significations indépendantes (Date de naissance (Jour, Mois, Année)).
- Mono-valué : une seule valeur pour une entité donnée (Nom, Adresse, Date de naissance).
- Multi-valués : plusieurs valeurs pour une entité donnée (Prénoms, auteurs, diplômes)



Un identifiant unique d'une entité est un sous-ensemble d'attribut, dont la valeur est unique pour chaque entité.

Chaque entité liée joue un rôle déterminé conformément aux règles de gestion dictées par le système d'information.

Type relation	Types-entités	Occurrence de type-relation
Posséder	{Personnes, Voitures}	Jérôme possède une AX sport
Habiter à	{Personnes, Villes}	Jean habite à paris

Propriété d'une association

Une association peut avoir une propriété si celle-ci dépend de toutes les occurrences d'entités qu'elle relie.

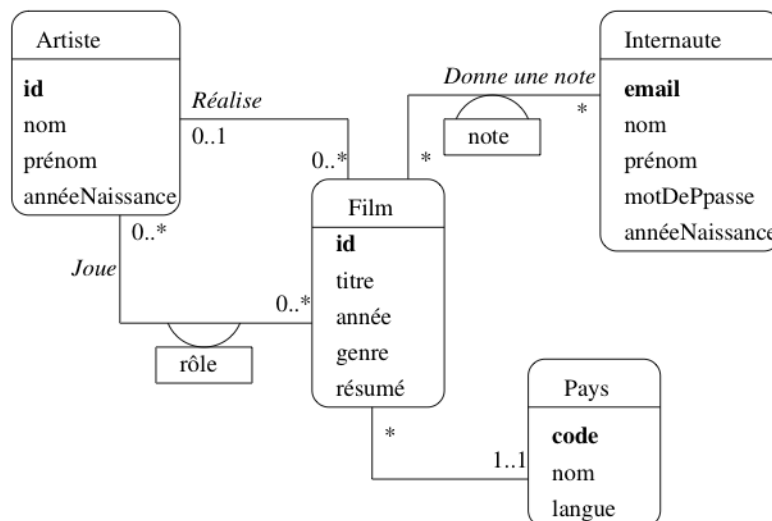


Schéma relationnel

Dans le schéma relationnel, les données sont représentées par des tables, de la façon dont les informations sont stockées dans la machine. Les tables constituent donc la *structure* logique du modèle relationnel. Au niveau physique, le système est libre d'utiliser n'importe quelle technique de stockage (fichiers séquentiels, indexage, adressage dispersé, séries de pointeurs, compression...) dès lors qu'il est possible de relier ces structures à des tables au niveau logique. Les tables ne représentent donc qu'une abstraction de l'enregistrement physique des données en mémoire.

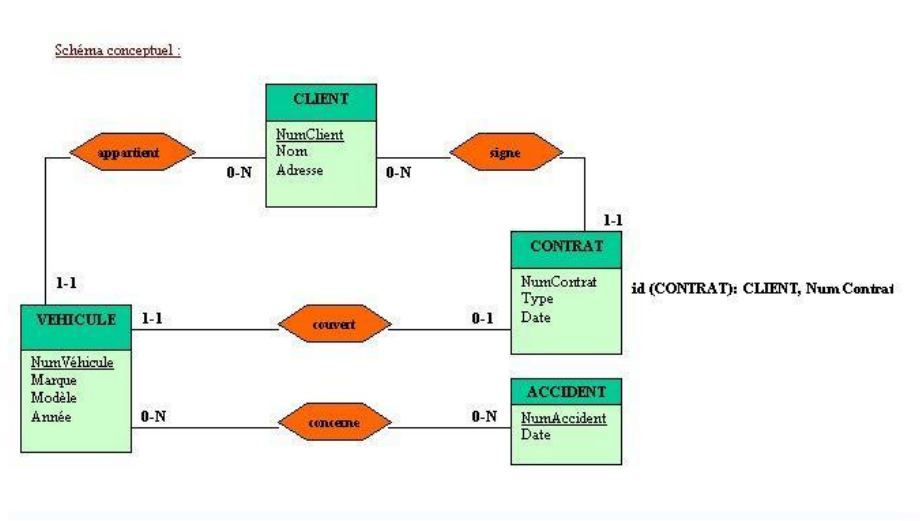
Les objectifs du modèle relationnel sont :

- Proposer des schémas de données faciles à utiliser.
- Améliorer l'indépendance logique et physique.
- Mettre à la disposition des utilisateurs des langages de haut niveau ;
- Optimiser les accès à la base de données.
- Améliorer l'intégrité et la confidentialité.
- Fournir une approche méthodologique dans la construction des schémas.

On peut définir le modèle relationnel de la manière suivante :

- Les données sont organisées sous forme de tables à deux dimensions, encore appelées relations.
- Les données sont manipulées par des opérateurs de l'algèbre relationnelle ;
- L'état cohérent de la base est défini par un ensemble de contraintes d'intégrité.

Au modèle relationnel, est associée à la théorie de la normalisation des relations qui permet de se débarrasser des incohérences au moment de la conception d'une base de donnée relationnelle.



Les cardinalités

Définition : les cardinalités relient un type-association et un type-entité précise le nombre de fois minimal et maximal d'interventions d'une entité du type-entité dans une association de type-association.

Exemple de cardinalité : une personne peut être l'auteur de 0 à n livre, mais un livre ne peut être écrit que par une personne.

Une cardinalité minimale est toujours 0 ou 1 et une cardinalité maximale est toujours 1 ou n .

Ainsi, si une cardinalité maximale est connue et vaut 2, 3 ou plus, alors nous considérons qu'elle est indéterminée et vaut n . En effet, si nous connaissons n au moment de la conception, il se peut que cette valeur évolue au cours du temps. Il vaut donc mieux considérer n comme inconnue dès le départ. De la même manière, on ne modélise pas des cardinalités minimales qui valent plus de 1, car ces valeurs sont également susceptibles d'évoluer. Enfin, une cardinalité maximale de 0 n'a pas de sens, car elle rendrait le *type association* inutile.

Les seules cardinalités admises sont donc :

0,1 : une occurrence du type entité peut exister tout en n'étant impliquée dans aucune association et peut être impliquée dans au maximum une association.

0, n : une occurrence du type entité peut exister tout en n'étant impliquée dans aucune association et peut être impliquée, sans limitation, dans plusieurs associations.

1,1 : une occurrence du type entité ne peut exister que si elle est impliquée dans exactement (au moins et au plus) une association.

1, n : une occurrence du type entité ne peut exister que si elle est impliquée dans au moins une association.

Le modèle relationnel

L'idée centrale du modèle relationnel est de décrire un ensemble de données comme une collection. Il faut alors décrire les toutes contraintes sur les valeurs et les combinaisons de valeurs possibles.

C'est un modèle logique de la base de données, à savoir un ensemble de relations, une part variable de relation, de telle sorte que toutes les propositions soient comblé. Une demande d'informations de la base de données (une requête de base de données) est également un prédicat.

Un grand avantage des données construites à partir de ce modèle relationnel est que l'utilisateur peut, dans sa retranscription physique en base de données, y accéder sans savoir où sont physiquement les données ni comment elles sont stockées. C'est un grand avantage par rapport au modèle hiérarchique implémenté dans les bases de données hiérarchique.

3 formes normale d'optimisation de base de données.

1FN : première forme normale :

Pour être en première forme normale (1FN), il faut que chaque attribut soit atomique. En d'autres termes, aucun attribut ne doit être multivalué (liste de valeurs) ou composé (si on le décompose, on obtient des informations supplémentaires.).

Dans cette table : si une personne a plusieurs adresses e-mail, alors elles sont notées dans l'attribut e-mail, séparées par des virgules.

L'attribut nom contient le nom, mais aussi le prénom et la civilité de la personne.

Pour passer en 1FN en modifiant la table ainsi :

id	pseudo	civilite	prenom	nom	email	email_2
1	Ed' la poignée	M.	Édouard	Bracame	edouard.bracame@joe-bar-team.org	ed@jbt.org
2	Joe l'arsouille	M.	Jean	Manchzeck	jean.manchzeck@joe-bar-team.org	

Il faut que chaque attribut soit atomique c'est-à-dire qu'aucun attribut ne soit multivalué ou composé. (25 rue du général de Gaulle 61 200 Argentan => on doit le diviser en 4 attribut numero_rue = 25 / nom_rue = générale de Gaulle / code_postal = 61 200 / ville = Argentan)

2FN : Deuxième forme normale :

La deuxième forme normale (2FN) ne concerne que les tables avec une **clé primaire composite** (composé avec plus d'un attribut).

Pour être en deuxième forme normale (2FN), il faut déjà être en 1FN et en plus respecter la règle suivante : aucun attribut ne faisant pas partie de la clé primaire ne doit dépendre que d'une partie la clé primaire.

Prenez la table résultat donnant une note artistique à des skateurs dans leur réalisation de figures :

skateur	figure	note
Richie	Ollie	8
Richie	Flip	10
Richie	<u>Hard-Flip</u>	9
Nyjah	Flip	7
Nyjah	<u>Hard-Flip</u>	8

<u>skateur</u>	<u>figure</u>	<u>difficulte</u>	note
Richie	Ollie	2	8
Richie	Flip	5	10
Richie	<u>Hard-Flip</u>	10	9
Nyjah	Flip	5	7
Nyjah	<u>Hard-Flip</u>	10	8

Nom	<u>Difficulte</u>
Ollie	2
Flip	5
<u>Hard-Flip</u>	10

Ajout de l'information difficulté.

La difficulté d'une figure ne dépend que de la figure et non du skateur. L'attribut difficulté ne dépend que d'une partie de la clé primaire (figure) et non de la clé primaire complète (skateur, figure) : nous ne respectons pas la 2FN.

Skateur	#figure_nom	note
Richie	Ollie	8
Richie	Flip	10
Nyjah	Hard-Flip	9
Nyjah	Hard-flip	8

La solution consiste donc à isoler les attributs concernés dans des tables dédiées. Je vais créer ici une table figure et déplacer la colonne difficulté dans cette nouvelle table :

3FN : troisième forme normale :

La troisième forme normale (3FN) ressemble à la deuxième (2FN) mais concerne la dépendance entre attributs non-clé.

Pour être en troisième forme normale (3FN), il faut déjà être en 2FN et en plus respecter la règle suivante : aucun attribut ne faisant pas partie de la clé primaire ne doit dépendre d'une partie des autres attributs ne faisant pas non plus partie de la clé primaire.

id	civilite	nom	prenom	sexe
1	Madame	Germain	Sophie	F
2	Monsieur	Hilbert	David	M
3	Madame	Noether	Emmy	F
4	Madame	Mirzakhani	Maryam	F

Cette table ne respecte pas la 3FN car l'attribut sexe peut être déduit de l'attribut non-clé primaire civilite.

Afin de résoudre ce problème, comme pour la 2FN, il convient de créer une nouvelle table civilite et d'y déplacer l'attribut sexe.

id	civilite	nom	prenom
1	Madame	Germain	Sophie
2	Monsieur	Hilbert	David
3	Madame	Noether	Emmy
4	Madame	Mirzakhani	Maryam

Avantages et inconvénients

Avantage :

- Rapidité de mise en œuvre.
- Facilité de maintenance, reprise ou clonage.
- Langage graphique permettant un apprentissage rapide.
- L'indépendance entre données et traitements.
- La duplication des données est réduite.
- L'ordre dans le stockage de données.
- L'utilisation simultanée des données par différents utilisateurs.

Inconvénient :

- Les systèmes de base de données sont complexes, difficiles à concevoir et prennent beaucoup de temps.
- Coûts de démarrage matériels et logiciels substantiels.
- Les dommages à la base de données affectent pratiquement tous les programmes d'applications.
- Coûts de conversion importants lors du transfert d'un système basé sur des fichiers vers un système de base de données.
- Formation initiale requise pour tous les programmeurs et utilisateurs.

Conclusion

En conclusion, les bases de données sont devenues indispensables pour stocker, émettre, partager et transmettre des informations. Elles sont devenues dans l'inconscient des utilisateurs indispensables à leurs usages quotidiens que ce soit sur Facebook, Youtube, Amazon et d'autres... Le monde actuel ne peut plus se passer de ce mode de fonctionnement. Elles ont su évoluer pour devenir indispensable ainsi les SGBD ont plus évolué s'adaptent au nouveau besoin des centres de données brassant de plus en plus d'informations par jour.

Bibliographie

<https://www.commentcamarche.com/contents/104-bases-de-donnees-introduction>

<https://www.culture-informatique.net/comment-marche-base-de-donnees/>

https://fr.wikipedia.org/wiki/Base_de_donn%C3%A9es

<https://www.lebigdata.fr/base-de-donnees#ftoc-heading-3>

https://fr.wikipedia.org/wiki/Mod%C3%A8le_entit%C3%A9-association

[https://fr.wikipedia.org/wiki/Forme_normale_\(bases_de_donn%C3%A9es_relationnelles\)](https://fr.wikipedia.org/wiki/Forme_normale_(bases_de_donn%C3%A9es_relationnelles))

https://blog.developpez.com/cinephil/p10397/bases-de-donnees/regle_de_gestion_bien_ecrite_g_modelisat

https://fr.wikipedia.org/wiki/Dictionnaire_des_donn%C3%A9es

<https://www.developpez.net/forums/d638833/general-developpement/alm/modelisation/schema/regles-cardinalites-relation-entre-tables/>

<https://laurent-audibert.developpez.com/Cours-BD/?page=conception-des-bases-de-donnees-modele-a>