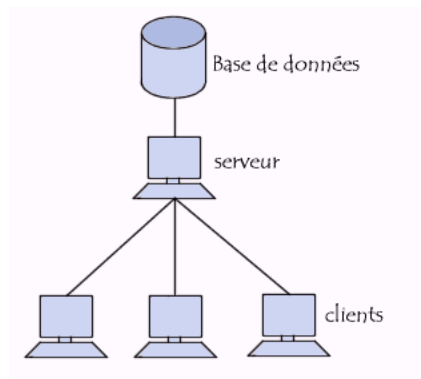


Les bases de données : Merise

INTRODUCTION : Les bases de données ?

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

Ces données doivent pouvoir être utilisées par des programmes, par des utilisateurs différents.



Une base de données permet de mettre des données à la disposition d'utilisateurs pour :

- une consultation
- une saisie
- une mise à jour
- une manipulation

tout en s'assurant des droits accordés à ces derniers.

Les bases de données : Merise

DEROULEMENT DU COURS

I.] MERISE au service des systèmes d'information

- A.) Le système d'information
- B.) MERISE

II.] Modélisation d'une base de données au niveau conceptuel

- A.) Les règles de gestion métiers
- B.) Le dictionnaire des données
- C.) Les dépendances fonctionnelles
- D.) Le Modèle Conceptuel de Données (MCD)

III.] Modélisation d'une base de données au niveau logique et passage au SQL

- A.) Le passage du MCD au MLD et SQL
 - 1.) Les relations
 - 2.) Règles de conversion
 - 3.) Élaboration du MLD et passage au SQL
- B.) Règles de vérification des niveaux de normalisation
- C.) Cas particuliers

IV.] Les extensions apportées par MERISE II

- A.) L'identification relative
- B.) L'héritage et ses limites
- C.) Les contraintes entre associations
- D.) Les CIF (contraintes d'intégrités fonctionnelles) et agrégations

V.] Conclusion

Les bases de données : Merise

I.] MERISE au service des systèmes d'information

A.) Le système d'information

Le **système d'information** ou SI, peut être défini comme étant l'ensemble des moyens humains, matériels et immatériels mis en œuvre afin de gérer l'information au sein d'une unité, une entreprise par exemple.

Il ne faut toutefois pas confondre un **système d'information** avec un **système informatique**. En effet, les systèmes d'information ne sont pas toujours totalement informatisés et existaient déjà avant l'arrivée des nouvelles technologies de l'information et des communications dont l'informatique fait partie intégrante.

Le SI possède 4 fonctions essentielles :

- La **saisie** ou **collecte** de l'information
- La **mémorisation** de l'information à l'aide de fichier ou de base de données
- Le **traitement** de l'information afin de mieux l'exploiter (consultation, organisation, mise à jour, calculs pour obtenir de nouvelles données, ...)
- La **diffusion** de l'information

Autrefois, l'information était stockée sur papier à l'aide de formulaires, de dossiers, ... et il existait des procédures manuelles pour la traiter. Aujourd'hui, les systèmes informatisés, comme les systèmes de gestion de bases de données relationnelles (SGBDR), sont mis au service du système d'information.

B.) MERISE

MERISE est une méthode française née dans les années 70, développée initialement par **Hubert Tardieu**. Elle fut ensuite mise en avant dans les années 80, à la demande du Ministère de l'Industrie qui souhaitait une méthode de conception des SI.

MERISE est donc une méthode d'analyse et de conception des SI basée sur le principe de la séparation des données et des traitements. Elle possède un certain nombre de **modèles** (ou **schémas**) qui sont répartis sur 3 niveaux :

- Le niveau **conceptuel**,
- Le niveau **logique** ou **organisationnel**,
- Le niveau **physique**.

Dans ce cours, nous ne nous intéresserons qu'à certains schémas permettant la conception d'une base de données relationnelle puis sa réalisation sur un SGBDR.

Les bases de données : Merise

II.] Modélisation d'une base de données au niveau conceptuel

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. Le MCD est basé sur deux notions principales : les **entités** et les **associations**, d'où sa seconde appellation : le **schéma Entité/Association**.

L'élaboration du MCD passe par les étapes suivantes :

- La mise en place de **règles de gestion** (si celles-ci ne vous sont pas données),
- L'élaboration du **dictionnaire des données**,
- La recherche des **dépendances fonctionnelles** entre ces données,
- L'élaboration du MCD (création des **entités** puis des **associations** puis ajout des **cardinalités**).

A.) Les règles de gestion métiers

Avant de vous lancer dans la création de vos tables (ou même de vos entités et associations pour rester dans un vocabulaire conceptuel), il vous faut recueillir les besoins des futurs utilisateurs de votre application. Et à partir de ces besoins, vous devez être en mesure d'établir les règles de gestion des données à conserver.

Prenons l'exemple d'un développeur qui doit informatiser le SI d'une bibliothèque. On lui fixe les règles de gestion suivantes :

- Pour chaque livre, on doit connaître le titre, l'année de parution, un résumé et le type (roman, poésie, science fiction, ...).
- Un livre peut être rédigé par aucun (dans le cas d'une œuvre anonyme), un ou plusieurs auteurs dont on connaît le nom, le prénom, la date de naissance et le pays d'origine.
- Chaque exemplaire d'un livre est identifié par une référence composée de lettres et de chiffres et ne peut être paru que dans une et une seule édition.
- Un inscrit est identifié par un numéro et on doit mémoriser son nom, prénom, adresse, téléphone et adresse e-mail.
- Un inscrit peut faire zéro, un ou plusieurs emprunts qui concernent chacun un et un seul exemplaire. Pour chaque emprunt, on connaît la date et le délai accordé (en nombre de jours).
- AUTEUR
 - un Auteur a écrit un ou plusieurs livres.
 - Un auteur est originaire d'un seul pays.
- TYPE DE LIVRE
 - Un type de livre peut posséder pas ou plusieurs livres.
- EXEMPLAIRES
 - Un exemplaire correspond à un livre et un seul.
 - Un exemplaire peut être emprunté une ou plusieurs fois.
 - Un exemplaire paraît dans une édition et une seule.

Les bases de données : Merise

- ÉDITEURS
- Un éditeur a paru 0 à plusieurs exemplaires.
- EMPRUNTS
- Un emprunt peut être effectué par un inscrit et un seul.
- Un emprunt concerne un seul exemplaire.
- LES INSCRITS
- Un inscrit a effectué 0 à plusieurs emprunts.

Ces règles vous sont parfois données mais vous pouvez être amené à les établir vous-même dans deux cas :

- Vous êtes à la fois maîtrise d'œuvre (MOE) et maîtrise d'ouvrage (MOA), et vous développez une application pour votre compte et/ou selon vos propres directives.
- **Ce qui arrive le plus souvent** : les futurs utilisateurs de votre projet n'ont pas été en mesure de vous fournir ces règles avec suffisamment de précision ; c'est pourquoi vous devrez les interroger afin d'établir vous-même ces règles. N'oubliez jamais qu'en tant que développeur, vous avez un devoir d'assistance à maîtrise d'ouvrage si cela s'avère nécessaire.

B.) Le dictionnaire des données

C'est une étape intermédiaire qui peut avoir son importance, surtout si vous êtes plusieurs à travailler sur une même base de données, d'un volume conséquent.

Le dictionnaire des données est un document qui regroupe toutes les données que vous aurez à conserver dans votre base (et qui figureront donc dans le MCD). Pour chaque donnée, il indique :

- Le **code mnémonique** : il s'agit d'un libellé désignant une donnée (par exemple «*titre_l*» pour le titre d'un livre)
- La **désignation** : il s'agit d'une mention décrivant ce à quoi la donnée correspond (par exemple «*titre du livre*»)
- Le **type de donnée** :
 - **A** ou **Alphabétique** : lorsque la donnée est uniquement composée de caractères alphabétiques (de 'A' à 'Z' et de 'a' à 'z')
 - **N** ou **Numérique** : lorsque la donnée est composée uniquement de nombres (entiers ou réels)
 - **AN** ou **Alphanumérique** : lorsque la donnée peut être composée à la fois de caractères alphabétiques et numériques
 - **Date** : lorsque la donnée est une date (au format AAAA-MM-JJ)
 - **Booléen** : Vrai ou Faux
- La **taille** : elle s'exprime en nombre de caractères ou de chiffres. Dans le cas d'une date au format AAAA-MM-MM, on compte également le nombre de caractères, soit 10 caractères. Pour ce qui est du type booléen, nul besoin de préciser la taille (ceci dépend de l'implémentation du SGBDR).
- Et parfois des **remarques** ou **observations** complémentaires (par exemple si une donnée est strictement supérieure à 0, etc).

Reprenons l'exemple de notre bibliothèque et du système de gestion des emprunts que nous sommes chargés d'informatiser. Après l'étude des règles de gestion, nous pouvons établir le dictionnaire des données suivant :

Les bases de données : Merise

Code mnémorique	Désignation	Type	Taille	Remarque
id_inscrit	Identifiant numérique, numéro de l'inscrit	N		
nom_inscrit	Nom d'un inscrit	A	30	
prenom_inscrit	Prénom d'un inscrit	A	30	
Ville_inscrit	Ville de l'inscrit	A	50	
cp_inscrit	code postal de l'inscrit	AN	10	
rue_inscrit	Rue de l'inscrit	AN	50	
date_naissance_inscrit	date de naissance de l'inscrit	Date	10	Au format AAAA-JJ-MM
tel_inscrit	Numéro fixe de l'inscrit	AN	15	
tel_port_inscrit	Mobile de l'inscrit	AN	15	
email_inscrit	Email de l'inscrit	AN	50	
id_livre	Identifiant numérique d'un livre	N		
titre_livre	titre d'un livre	AN	50	
annee_livre	Année d'écriture du livre	N	4	
resume_livre	Résumé du livre	AN	1000	
id_type_livre	identifiant numérique d'un type de livre	N		
libelle_type_livre	Libellé d'un type de livre	AN	30	
id_exemplaire	identifiant numérique de l'exemplaire	N		
id_edition	Identifiant numérique d'un édition	N		
nom_edition	Nom de l'édition du livre	AN	30	
id_auteur	Identifiant numérique de l'auteur	N		
nom_auteur	Nom de l'auteur	A	30	
prenom_auteur	Prénom de l'auteur	A	30	
date_naissance_auteur	Date de naissance de l'auteur	Date		Au format aaaa-jj-mm
id_pays	identifiant numérique du pays	N		
nom_pays	Nom du pays	A	50	

Les bases de données : Merise

id_emprunt	identifiant numérique de l'emprunt	N		
date_emprunt	Date de l'emprunt	Date		Au format aaaa-jj-mm
delai_emprunt	Délai autorisé de l'emprunt	N	3	S'exprime en nombre de jours
	Etat si l'emprunt est terminé ou non	B		0 : emprunt en cours 1 : emprunt terminé

Remarques :

- Les données qui figurent dans le MCD (et donc dans le dictionnaire des données) doivent être, dans la plupart des cas, **élémentaires** :
 - Elles ne doivent pas être **calculées** : les données calculées doivent être obtenues, par le calcul, à partir de données élémentaires qui, elles, sont conservées en base. Cependant, il existe quelques cas où il s'avère pertinent de conserver, pour des raisons d'optimisation, une donnée calculée, le montant d'une commande par exemple. On ne conservera cependant pas les données calculées intermédiaires sauf en cas d'obligation légale (c'est le cas pour un montant HT par exemple, où les composantes peuvent d'ailleurs avoir un prix variable dans le temps). En effet, cela évite de refaire les calculs plusieurs fois pour un résultat qui restera fixe.
 - Elles ne doivent pas être **composées** : les données composées doivent être obtenues par la concaténation de données élémentaires conservées en base. Par exemple une adresse est obtenue à partir d'une rue, d'une ville et d'un code postal : ce sont ces trois dernières données qui sont conservées et donc qui figureront dans le MCD (et dans le dictionnaire des données).
- Lorsque l'on n'effectue jamais de calcul sur une donnée numérique, celle-ci doit être de type AN (c'est le cas par exemple pour un numéro de téléphone).

C.) Les dépendances fonctionnelles

Soit deux propriétés (ou données) P1 et P2. On dit que P1 et P2 sont reliées par une **dépendance fonctionnelle** (DF) si et seulement si une **occurrence** (ou valeur) de P1 permet de connaître une et une seule occurrence de P2.

Cette dépendance est représentée comme ceci :

P1 → P2

On dit que P1 est la **source** de la DF et que P2 en est le **but**.

Par ailleurs, plusieurs données peuvent être source comme plusieurs données peuvent être but d'une DF. Exemples :

P1,P2 → P3

P1 → P2,P3

P1, P2 → P3,P4,P5

...

Les bases de données : Merise

En reprenant les données du dictionnaire précédent, on peut établir les DF suivantes :

id_edition → nom_edition

id_emprunt → date_emprunt, delais_emprunt, id_inscrit, id_exemplaire

id_inscrit → nom_inscrit, prenom_inscrit, rue_inscrit, ville_inscrit, cp_inscrit, tel_inscrit, tel_port_inscrit, email_inscrit, date_naissance_inscrit

id_exemplaire → id_livre, id_edition

id_livre → titre_livre, annee_livre, resume_livre, id_type_livre

id_type_livre → libelle_type_livre

id_pays → nom_pays

id_auteur → nom_auteur, prenom_auteur, date_naissance_auteur, id_pays

On peut déduire les conclusions suivantes de ces DF :

- À partir d'un numéro d'emprunt, on obtient une date d'emprunt, un délai, l'identifiant de l'inscrit ayant effectué l'emprunt, la référence de l'exemplaire emprunté.
- À partir d'une référence d'exemplaire, on obtient l'identifiant du livre correspondant.
- À partir d'un numéro de livre, on obtient son titre, son année de parution, un résumé, l'identifiant du type correspondant, son numéro d'édition.
- ...

Remarque :

Une DF doit être :

- élémentaire : C'est l'intégralité de la source qui doit déterminer le but d'une DF. Par exemple si $P1 \rightarrow P3$ alors $P1, P2 \rightarrow P3$ n'est pas élémentaire.
- directe : La DF ne doit pas être obtenue par transitivité. Par exemple, si $P1 \rightarrow P2$ et $P2 \rightarrow P3$ alors $P1 \rightarrow P3$ a été obtenue par transitivité et n'est donc pas directe.

Conclusion :

Les DF qui existent entre les données sont parfois évidentes et ne nécessitent pas toujours une modélisation mais celle-ci peut s'avérer utile car elle permet, entre autres, de distinguer les futures entités du MCD et leur identifiants.

Les bases de données : Merise

D.) Le Modèle Conceptuel de Données (MCD)

1.) Les entités

Chaque entité est unique et est décrite par un ensemble de propriétés encore appelées attributs ou caractéristiques. Une des propriétés de l'entité est l'identifiant. Cette propriété doit posséder des occurrences uniques et doit être source des dépendances fonctionnelles avec toutes les autres propriétés de l'entité. Bien souvent, on utilise une donnée de type entier qui s'incrémente pour chaque occurrence, ou encore un code unique spécifique du contexte.

Le formalisme d'une entité est le suivant :

Nom de l'entité
<u>identifiant</u> propriété1 propriété2 ...

Ainsi, si on reprend notre dictionnaire de données précédent, on schématise par exemple une entité «**Auteur**» comme ceci :

Auteur
<u>id_a</u> nom_a prenom_a date_naissance_a

À partir de cette entité, on peut retrouver la règle de gestion suivante : un auteur est identifié par un numéro unique (id_a) et est caractérisé par un nom, un prénom et une date de naissance.

Une entité peut n'avoir aucune, une ou plusieurs occurrences. Pour illustrer ce terme d'«occurrence» qui a déjà été utilisé plusieurs fois, voici un exemple de **table d'occurrences** de l'entité **Auteur** :

id_a	nom_a	prenom_a	date_naissance_a
1	Hugo	Victor	1802-02-26
2	Rimbaud	Arthur	1854-10-20
3	de Maupassant	Guy	1850-08-05

Les bases de données : Merise

Cette table est composée de trois occurrences de l'entité **Auteur**.

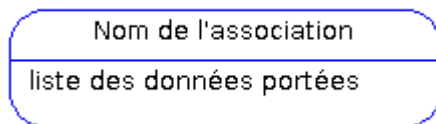
Remarques :

- Les occurrences sont parfois appelés **tuples**. Par ailleurs, la table d'occurrence peut être comparée à l'**instance** d'une **relation** (implantation relationnelle d'une entité ou association) à un moment donné. Nous reviendrons sur cette notion de relation dans la partie III.
- Au niveau conceptuel, on devrait plutôt parler d'**entités-types**, les entités étant en fait des instances d'entités-types. Par soucis de simplicité, on gardera les termes d'entités et associations tout au long du cours.

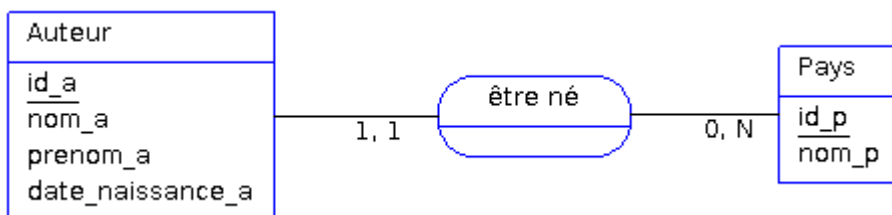
2.) Les associations

Une association définit un lien sémantique entre une ou plusieurs entités. En effet, la définition de liens entre entités permet de traduire une partie des règles de gestion qui n'ont pas été satisfaites par la simple définition des entités.

Le formalisme d'une association est le suivant :



Généralement le nom de l'association est un verbe définissant le lien entre les entités qui sont reliées par cette dernière. Par exemple :



Ici l'association «être né» traduit les deux règles de gestion suivantes :

- Un auteur est né dans **un et un seul** pays,
- Dans un pays, sont nés **aucun, un ou plusieurs** auteurs.

Vous remarquerez, que cette association est caractérisée par ces annotations **1,1** et **0,N** qui nous ont permis de définir les règles de gestions précédentes. Ces annotations sont appelées les **cardinalités**.

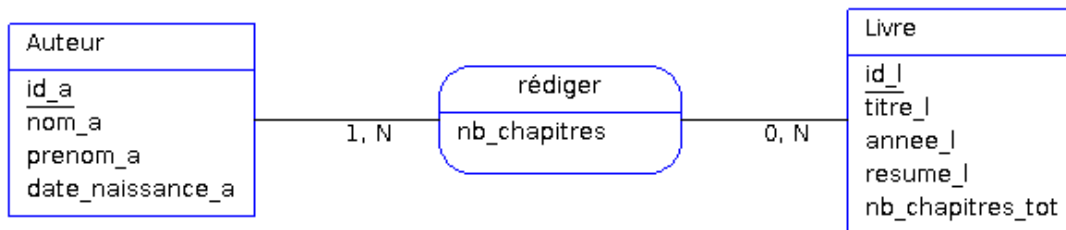
Les bases de données : Merise

Une cardinalité est définie comme ceci :

minimum, maximum

Les cardinalités les plus répandues sont les suivantes : **0,N** ; **1,N** ; **0,1** ; **1,1**. On peut toutefois tomber sur des règles de gestion imposant des cardinalités avec des valeurs particulières, mais cela reste assez exceptionnel et la présence de ces cardinalités imposera l'implantation de traitements supplémentaires.

L'identifiant d'une association ayant des cardinalités 0,N/1,N de part et d'autre, est obtenu par la concaténation des entités qui participent à l'association. Imaginons l'association suivante :



Ici un auteur rédige au moins un ou plusieurs livres et pour chaque livre, on connaît le nombre de chapitres rédigés par l'auteur (on connaît aussi le nombre total de chapitres pour chaque livre).

L'association «rédiger» peut donc être identifiée par la concaténation des propriétés id_a et id_l. Ainsi, le couple **id_a, id_l** doit être unique pour chaque occurrence de l'association. On peut également définir la dépendance fonctionnelle suivante :

$id_a, id_l \rightarrow nb_chapitres$

On dit que nb_chapitres (nombre de chapitres rédigés par un auteur, pour un livre) est une donnée portée par l'association «rédiger». Cette association est donc une association porteuse de données.

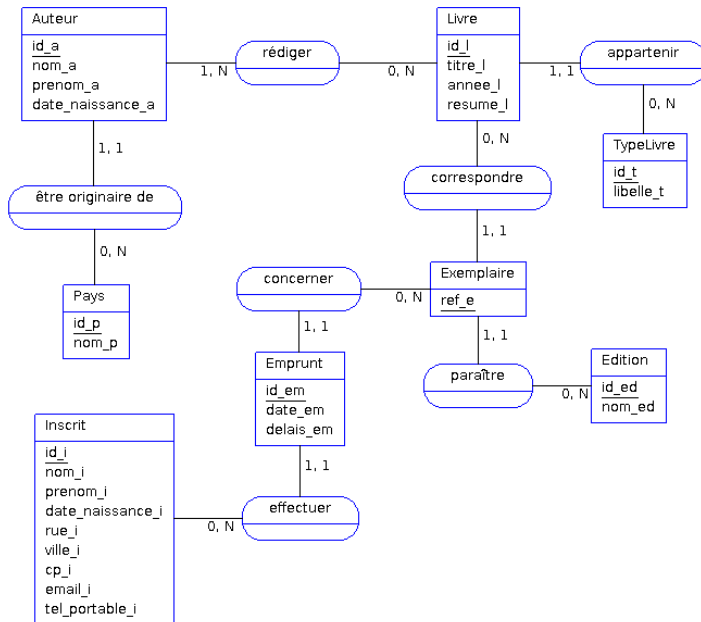
Pour une association ayant au moins une cardinalité de type 0,1 ou 1,1 considérons dans un premier temps que cette dernière ne peut être porteuse de données et qu'elle est identifiée par l'identifiant de l'entité porteuse de la cardinalité 0,1 ou 1,1.

Nous reviendrons plus en détail sur la notion d'identification d'une association lors du passage au modèle logique.

3.) Élaboration du MCD

Avec toutes ces connaissances, il nous est donc possible d'élaborer le MCD complet à partir des données présentes dans le dictionnaire des données :

Les bases de données : Merise



Remarques :

- Souvent, pour un même ensemble de règles de gestion, plusieurs solutions sont possibles au niveau conceptuel. Par exemple, rien ne nous obligerait ici à créer une entité **Type**. Une simple donnée portée par l'entité **Livre** aurait pu convenir également.
- Pour que le MCD soit sémantiquement valide, toute entité doit être reliée à au moins une association.
- Les entités et les propriétés peuvent être historisées. Dans ce cas on met un **(H)** à la fin du nom de l'entité ou de la propriété que l'on souhaite historiser (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.
- Il existe des outils de modélisation payants et d'autres gratuits pour MERISE (powerAMC, OpenModelSphere, AnalyseSI, JMerise, etc).
- On aurait pu, dans ce cas précis, conserver également une date de rentrée des livres, calculée à partir de la date de location et de la durée de celle-ci. C'est un exemple de donnée calculée dont la conservation peut s'avérer pertinente (notamment pour faciliter l'envoi de rappels).