

TNSI - Modélisation et conception d'une base de données

1. Concevoir une base de donnée

1.1 Généralités

Intérêts d'un Système de Gestion de Base de Données (SGBD)

- Assure la persistance des données
- Structure l'information
- Permet de trouver rapidement une information
- Multi-utilisateurs
- Sécurise les données
 - Filtre les données qu'un groupe d'utilisateurs peut voir
 - Centralise la sauvegarde et la mise à jour des données
- Maintient la cohérence des données
 - Non redondance
 - Contrôle de l'intégrité des données (lors de la saisie, mise à jour, suppression)

Type de SGBD

- Bases hiérarchiques (structure arborescente) ou réseau (structure de graphe)
 - navigation entre les données
- Bases relationnelles
 - Données sous forme de table, langage SQL
- Bases déductives
 - Intégration d'ensemble de règles, langage DATALOG
- Bases objet
 - Données sous forme d'objets
- Bases noSQL
 - Pas de structuration des données

1.2 Volumétrie

Exemple d'une banque (BNP Paribas)

Grande masse d'informations $8 * 10^6$ clients

Plusieurs utilisateurs simultanément

- 4 comptes par client, donc $32 * 10^6$ comptes
- 20 écritures par mois par compte, donc $6.4 * 10^8$ écritures par mois
- 2140 agences
- 31.460 collaborateurs
- des milliers d'accès internet

1.3 Analyse

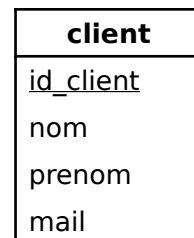
1. Travail préalable à la création de la base de données
2. Déterminer les informations qui sont nécessaires à l'application
 - gestion de la paie, des congés, du stock
 - application web
3. Dictionnaire de données
 - Parmi toutes les informations, on repère les données élémentaires ou propriétés
 - niveau de granularité dépend du contexte (adresse, client,...)
 - Le dictionnaire de données représente l'ensemble des données élémentaires

1.4 Modèle Conceptuel de Données: MCD

- Le MCD est une représentation du système d'informations à l'aide d'entités et d'associations
- C'est le résultat du travail des analystes, il sert de base à la création de la base de données
- Peut être lu et compris par des non informaticiens
- Un MCD est toujours contextuel

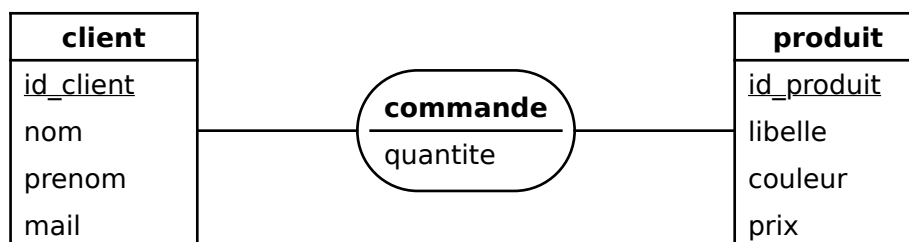
Notion d'identité

- Une entité regroupe les propriétés relatives à un même sujet, qui a du sens
 - Exemple: une voiture, un individu...
- Comporte un identifiant
 - peut être composé par une seule ou plusieurs propriétés
 - est unique: ne peut être le même pour deux entités
- Se représente par un rectangle, l'identifiant est souligné



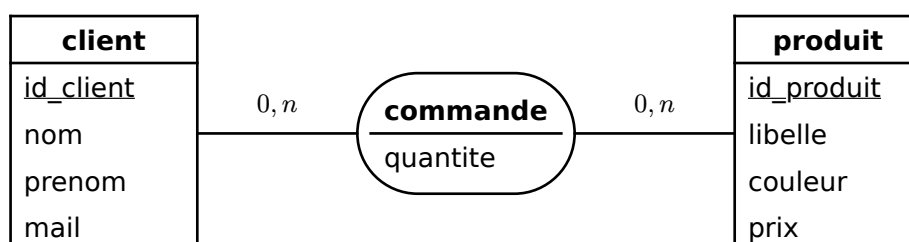
Notion d'association

- Les entités peuvent être liées par des associations
- Une association est une représentation abstraite de la mémorisation d'un lien entre entités
- Elle est représentée par un cercle entre entités
 - Exemples: commande, est_inscrit, travaille_pour, est_marie, habite_dans....

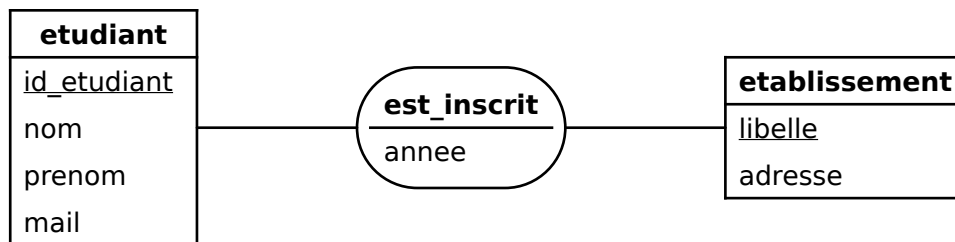


Cardinalités

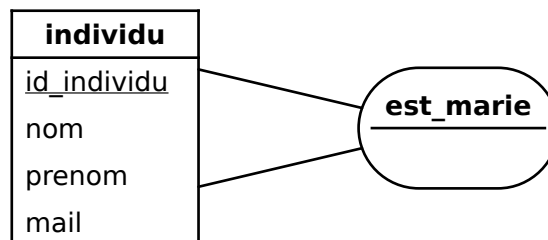
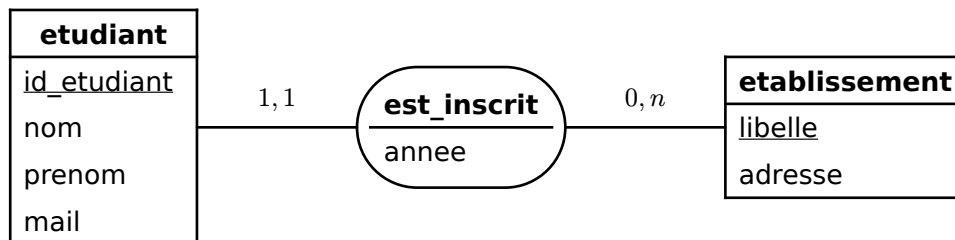
- Les cardinalités précisent le nombre de fois que l'entité peut intervenir dans une association.
- La valeur minimale est 0 ou 1, la valeur maximale est 1 ou n
- L'association peut-être hiérarchique (maximum 1 d'un côté, n de l'autre) ou maillée (maximum n des deux côtés), entre une ou plusieurs entités



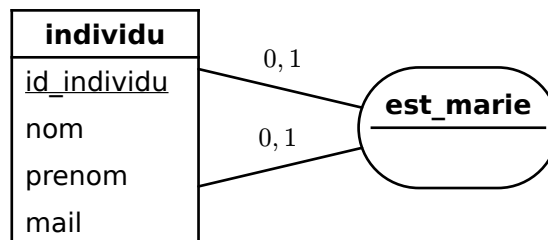
Autres exemples



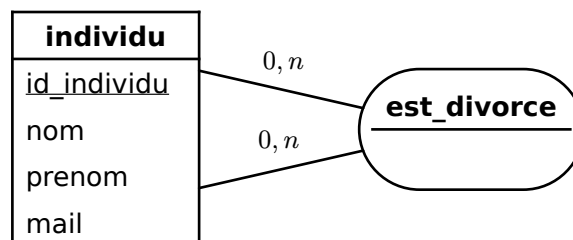
Un (min=1, max=1) étudiant peut être inscrit dans aucune ou plusieurs (min = 0 max = n) universités.

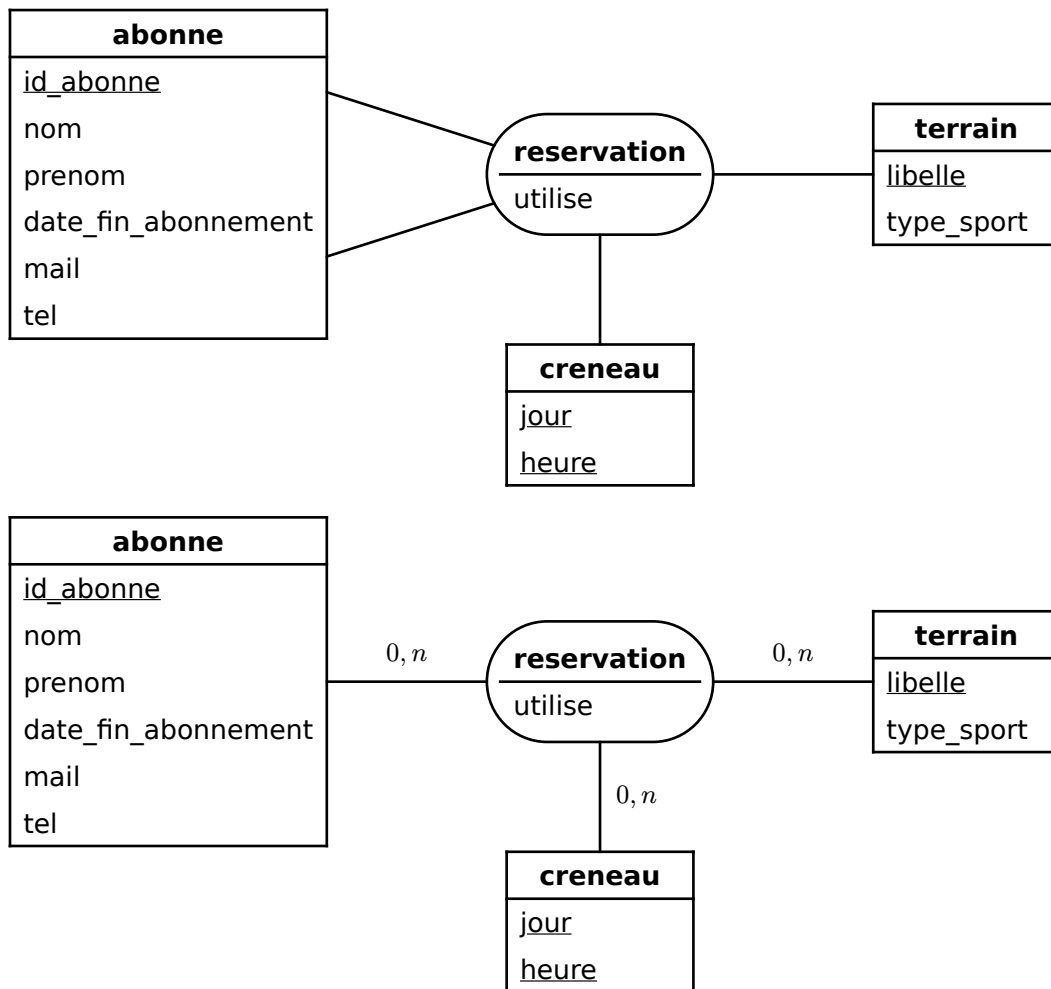


Un individu peut-être marié à au plus une personne. Il peut ne pas être marié. min=0, max=0 des deux côtés.



Un individu peut-être divorcé à plusieurs personnes. Il peut ne pas être divorcé. min=0, max=n des deux côtés.





1.5 Modèle Relationnel

- La traduction concrète du MCD dépend du modèle de la base
- Dans le cas d'une base de données relationnelle, on obtient un modèle relationnel

Principe

- Données sous forme de tables
- Chaque table et chaque colonne (ou attribut) porte un nom
- Chaque attribut est typé
- Chaque ligne représente un enregistrement
- Pas de lien physique entre les tables

1.6 Passage du MCD au Modèle Relationnel

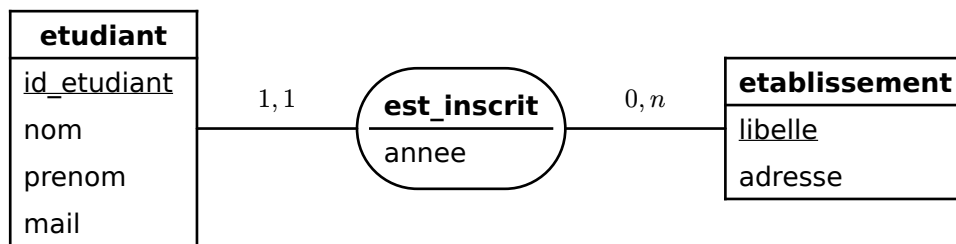
Principe

- Chaque entité devient une table
- Chaque propriété d'une entité devient un attribut
- L'identifiant d'une entité devient la clé primaire de la table (Primary Key)

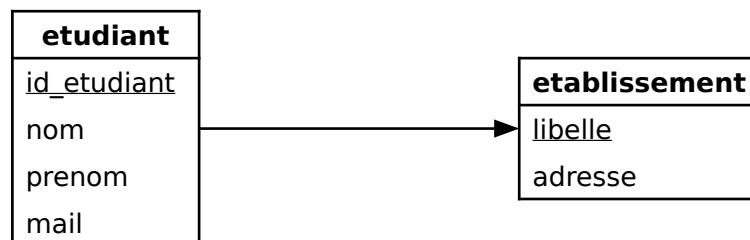
Lien Hiérarchique

Une association (0-n)-(0-1) est un lien hiérarchique.

Elle se traduit par la migration de la clé primaire côté n vers une clé étrangère (Foreign Key) côté 1



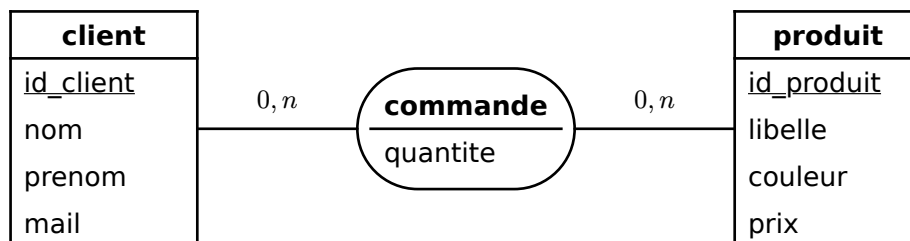
On l'indique avec une flèche **de la clé étrangère vers la clé primaire**.



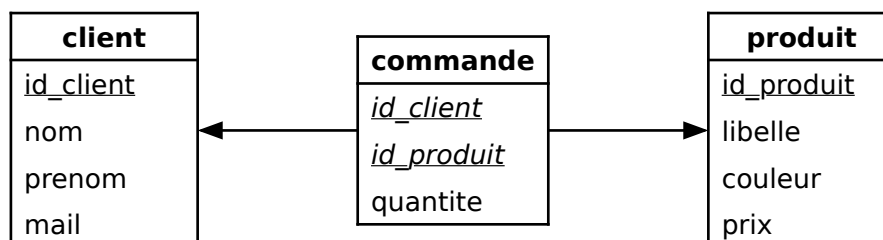
Lien Maillé

Une association maillée (0-n)-(0-n) donne lieu à la création d'une nouvelle table dont la clé primaire est l'union des clés primaires des entités qu'elle relie.

Ces attributs sont aussi des clés étrangères.

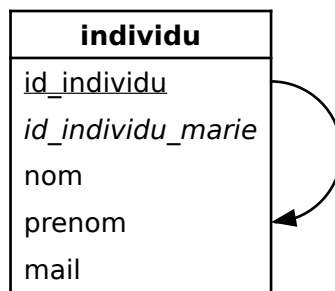


On l'indique donc avec deux flèches, de la table **relation** vers les tables contenant les clés primaires.

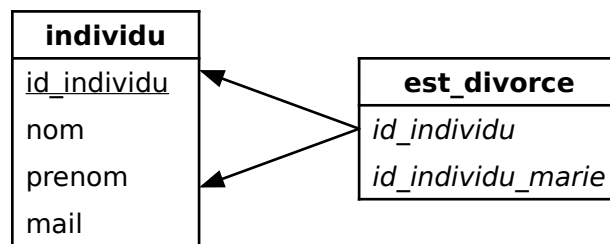


Autres exemples

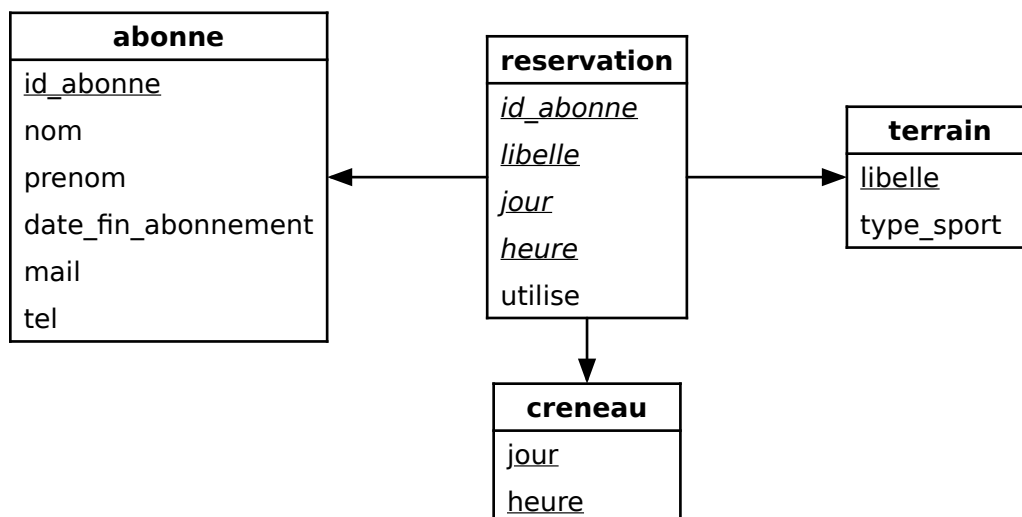
Est marié



Est divorcé



Réservation



2. Conception à partir d'un csv

Retour sur l'exemple des catégories socio-professionnelles

Dans le cas où on récupère des données réelles, le problème se pose différemment.

Le dictionnaire des données est déduit du csv récupéré.

En revanche, les données réelles peuvent poser d'autres problèmes: doublons, incohérences, données non complétées ou peu exploitables.

En général, il faut faire un travail de « nettoyage » pour pouvoir créer les tables.

2.1 Construire une structure optimisée

Regrouper les données en tables

- Mettre dans une même table les données relatives à un même sujet
- Créer de nouvelles tables pour éviter la redondance des données
 - Limite les incohérences lors des mises à jour
 - Facilite la construction des requêtes et améliore la pertinence des résultats

Établir les relations entre tables

- Définir les clés primaires
 - Uniques et non NULL
- Définir les clés étrangères
 - Référencent les clés primaires

Définir des colonnes pertinentes

- Facilité d'interrogation des colonnes
- Données cohérentes au sein d'une colonne
- Ne pas conserver des données qui peuvent être calculées

Cas des catégories socioprofessionnelles du Nord.

Code	Rég	Dépt	Libellé géo	Coordonnées	Date	var	Populat	Sexe	Tranche	Catégorie Socio-Professionnelle	catégorie
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Prof. intermédiaires	28	Femmes	15+	Professions Intermédiaires	Prof. Intermédiaires
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Employés	64	Total	15+	Employés	Employés
59001	31	59	Abancourt	50.2368696873.3.20	2012	55+ - Agriculteurs exploitants	0	Total	55+	Agriculteurs Exploitants	Agriculteurs
59001	31	59	Abancourt	50.2368696873.3.20	2012	15-24 - Employés	16	Total	15-24	Employés	Employés
59001	31	59	Abancourt	50.2368696873.3.20	2012	25-54 - Autres	12	Total	25-54	Autres	Autres
59001	31	59	Abancourt	50.2368696873.3.20	2012	25-54 -	176	Total	25-54	Total	Total
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ -	196	Hommes	15+	Total	Total
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Artisans, Comm., Chefs entr.	44	Total	15+	Artisans, Commerçants, Chefs d'entreprises	Chefs d'entreprises
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Autres	32	Total	15+	Autres	Autres
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Ouvriers	8	Femmes	15+	Ouvriers	Ouvriers
59001	31	59	Abancourt	50.2368696873.3.20	2012	25-54 - Artisans, Comm., Chefs entr.	28	Total	25-54	Artisans, Commerçants, Chefs d'entreprises	Chefs d'entreprises
59001	31	59	Abancourt	50.2368696873.3.20	2012	25-54 - Ouvriers	28	Total	25-54	Ouvriers	Ouvriers
59001	31	59	Abancourt	50.2368696873.3.20	2012	55+ - Autres	0	Total	55+	Autres	Autres
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Artisans, Comm., Chefs entr.	24	Hommes	15+	Artisans, Commerçants, Chefs d'entreprises	Chefs d'entreprises
59001	31	59	Abancourt	50.2368696873.3.20	2012	Population en 2012 (princ)	445	Total	Population	Total	Total
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Cadres, Prof. intel. sup.	16	Total	15+	Cadres, Professions Intellectuelles Supérieures	Cadres, PIS
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Retraités	48	Femmes	15+	Retraités	Retraités
59001	31	59	Abancourt	50.2368696873.3.20	2012	55+ - Prof. intermédiaires	8	Total	55+	Professions Intermédiaires	Prof. Intermédiaires
59001	31	59	Abancourt	50.2368696873.3.20	2012	15+ - Employés	20	Hommes	15+	Employés	Employés

2.2 Construire une structure pertinente

Regrouper les données en tables

- Tables = données relatives à un même sujet
 - Données sur les villes et sur les effectifs des catégories

Éviter la redondance des colonnes

- var = Tranche + catégorie
- catégorie = abréviation Catégorie socioprofessionnelle

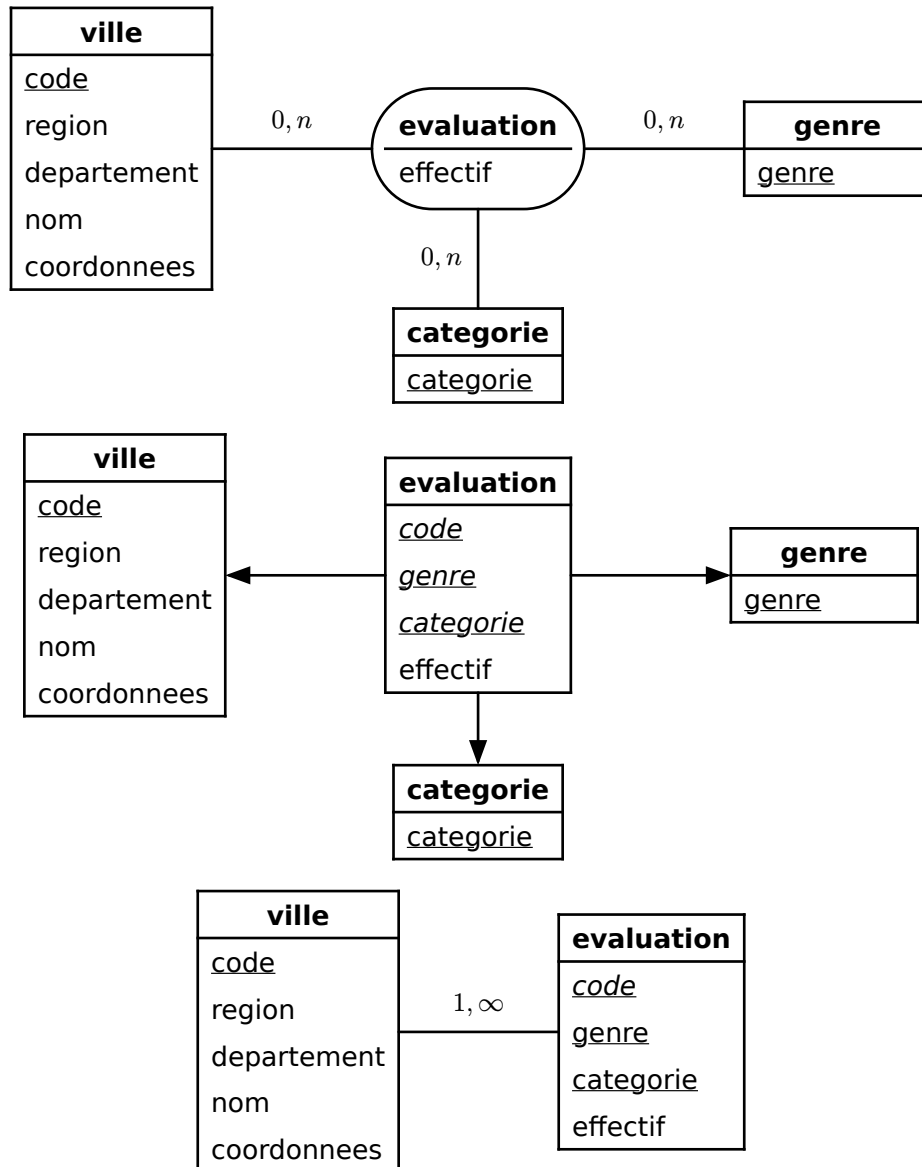
Éviter la redondance des valeurs

- Les colonnes `departement` et `region` ne contiennent qu'une seule valeur
- À conserver uniquement si volonté d'étendre à d'autres données

Données cohérentes

- Les colonnes Sexe, Tranche, ... contiennent aussi des totaux !
 - Les totaux peuvent être calculés à l'aide de fonctions et d'agrégats
- Décomposer coordonnees en deux REAL

2.3 MCD possible



3. Bilan

- Base de données relationnelle: ensemble de tables, aucun lien physique entre les tables
- Contraintes d'intégrité :
 - Clé primaire
 - Clé étrangère (contrainte référentielle)
 - Contrainte de domaine (CHECK prix > 0, CHECK reponse IN (< 0 >, < N >)...)
- Produit Cartésien vs JOIN: JOIN plus efficace
- JOIN indépendant des contraintes référentielles