

Part 4

NORMALISATION

Problématique

- Mélanger dans une même relation des informations relatives à plusieurs entités entraîne, en général des redondances d'information qui provoquent les anomalies suivantes:
 - ❑ anomalies **d'insertion**,
 - ❑ anomalies de **mise à jour**,
 - ❑ anomalies de **suppression**.

Exemple

isbn	titre	éditeur	pays
2-212-09283-0	Bases de données – objet et relationnel	Eyrolles	France
2-7117-8645-5	Fondements des bases de données	Vuibert	USA
0-201-70872-8	Databases and Transaction Processing	Addison Wesley	USA
2-212-09069-2	Internet/Intranet et bases de données	Eyrolles	France

- Cette relation qui décrit des livres et leurs éditeurs, contient des redondances provoquant les anomalies suivantes :
 - **insertion** : il n'est pas possible d'insérer un livre sans connaître son éditeur. Attribuer une valeur nulle aux attributs éditeur et pays violerait l'intégrité d'entité : un attribut appartenant à la clé, ne doit pas avoir de valeur nulle.
 - **mise à jour** : si un éditeur change de pays, il faut modifier ce pays pour chacun des livres qu'il a édités.
 - **suppression** : si l'unique livre publié par un éditeur est supprimé, l'information sur cet éditeur est perdue. Attribuer une valeur nulle aux attributs isbn et titre, violerait l'intégrité d'entité.

Normalisation

- La solution à ces problèmes consiste à normaliser la relation en cause.
- **Un schéma relationnel normalisé** doit répondre aux exigences minimales suivantes :
 - **Non redondance**: un attribut n'appartient qu'à une seule relation, donc à une seule table, à moins qu'il n'agisse comme clé étrangère pour assurer l'association avec une autre table ;
 - **Cohérence**: les attributs qui décrivent le même objet appartiennent à la même table et dépendent chacun fonctionnellement et totalement de la clé primaire de la table.

La normalisation

- La normalisation consiste en des **méthodes de décomposition** des relations.
- Cette décomposition repose sur **l'analyse de dépendances** entre attributs qui sont à l'origine de phénomènes de redondance.
 - **Par exemple**, la dépendance entre l'ISBN d'un livre et son titre ou bien entre le nom d'un éditeur et son pays.

Dépendances fonctionnelles

- Un attribut **b** dépend fonctionnellement d'un attribut **a** si **à une valeur de a correspond au plus une valeur de b**. La dépendance fonctionnelle est notée **$a \rightarrow b$** .
- Le **membre droit** de l'écriture s'appelle le **dépendant**, le **membre gauche** s'appelle le **déterminant**.

Exemples

- Supposons les exemples suivants, qui concernent des **pilotes** ayant un **numéro**, un **nom**, une **fonction** (copilote, commandant, instructeur...), et un **nombre d'heures de vol** pour un **jour** particulier:
 - $\text{numPilote}, \text{jour} \rightarrow \text{nbHeuresVol}$ est une DF, car à un couple (numPilote, jour) correspond au plus un nombre d'heures de vol ;
 - $\text{numPilote} \rightarrow \text{nomPilote}, \text{fonction}$ est équivalente aux écritures $\text{numPilote} \rightarrow \text{nomPilote}$ et $\text{numPilote} \rightarrow \text{fonction}$ qui sont deux DF. En conséquence $\text{numPilote} \rightarrow \text{nomPilote}, \text{fonction}$ est une DF ;
 - $\text{nomPilote} \rightarrow \text{fonction}$ est une DF s'il n'y a pas d'homonymes dans la population des pilotes enregistrés dans la base de données. Dans le cas contraire, ce n'est pas une DF, car à un nom de pilote peuvent correspondre plusieurs fonctions ;
 - $\text{fonction} \rightarrow \text{nomPilote}$ n'est pas une DF, car à une fonction donnée correspondent éventuellement plusieurs pilotes.

DF élémentaire

- Une DF $a, b \rightarrow c$ est **élémentaire** si ni $a \rightarrow c$, ni $b \rightarrow c$ ne sont des DF.
- Considérons les exemples suivants :
 - la dépendance fonctionnelle $\text{numPilote}, \text{jour} \rightarrow \text{nbHeuresVol}$ est **élémentaire**, car $\text{numPilote} \rightarrow \text{nbHeuresVol}$ n'est pas une DF (un pilote vole différents jours, donc pour un pilote donné, il existe plusieurs nombres d'heures de vol), pas plus que $\text{jour} \rightarrow \text{nbHeuresVol}$ (à un jour donné, plusieurs vols sont programmés) ;
 - la dépendance fonctionnelle $\text{numPilote}, \text{nomPilote} \rightarrow \text{fonction}$ n'est pas **élémentaire**, car le numéro du pilote suffit pour retrouver sa fonction ($\text{numPilote} \rightarrow \text{fonction}$ est une DF).

DF directe

- Une DF $a \rightarrow c$ est **directe** si elle n'est pas déduite par transitivité, c'est-à-dire s'il n'existe pas de DF $a \rightarrow b$ et $b \rightarrow c$.
- Considérons l'exemple d'un **avion** avec les attributs suivants : (**immat** : numéro d'immatriculation d'un avion ; **typeAvion** : type de l'avion; **nomConst** : nom du constructeur de l'avion).
 - La dépendance **immat** \rightarrow **nomConst** **n'est pas une DF directe**.
 - La dépendance **immat** \rightarrow **typeAvion** est une DF directe.
 - La dépendance **typeAvion** \rightarrow **nomConst** est une DF directe.

Propriétés de DF

- Les propriétés suivantes sont appelées «**axiomes d'Armstrong**».
 - **Réflexivité** : si b est un sous-ensemble de a alors $a \rightarrow b$ est une DF.
 - **Augmentation** : si $a \rightarrow b$ est une DF, alors $a, c \rightarrow b, c$ est une DF.
 - **Transitivité** : si $a \rightarrow b$ et $b \rightarrow c$ sont des DF, alors $a \rightarrow c$ est une DF.
 - **Union** : si $a \rightarrow b$ et $a \rightarrow c$ sont des DF, alors $a \rightarrow b, c$ est une DF.
 - **Pseudo-transitivité** : si $a \rightarrow b$ et $b, c \rightarrow d$ sont des DF, alors $a, c \rightarrow d$ est une DF.
 - **Décomposition** : si $a \rightarrow b, c$ est une DF, alors $a \rightarrow b$ et $a \rightarrow c$ sont des DF.

Normalisation

- Les **classifications** formelles utilisées pour décrire le **niveau** de normalisation d'une base de données relationnelle sont appelées les **formes normales (FN)**
- Il existe **huit** formes normales. Les **quatre** premières sont les plus pratiques et sont à connaître:
 - la 1ère forme normale (1FN),
 - la 2ème forme normale (2FN),
 - la 3ème forme normale (3FN),
 - la forme normale de Boyce-Codd qui est la plus aboutie (FNBC).

1FN

- Une relation est en **1FN** si tous ses attributs ont une valeur atomique.

pas en 1FN

isbn	auteur
2-212-09283-0	Gardarin
2-7117-8645-5	Abiteboul, Hull, Vianu

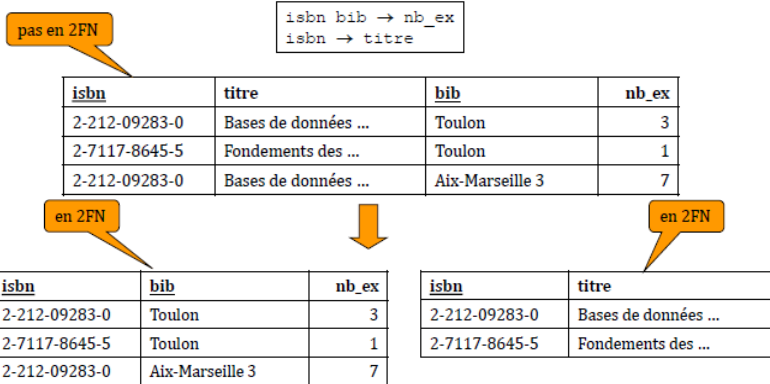
en 1FN

isbn	auteur
2-212-09283-0	Gardarin
2-7117-8645-5	Abiteboul
2-7117-8645-5	Hull
2-7117-8645-5	Vianu

- Une fois on passe à la 1FN il faut designer la clé primaire !

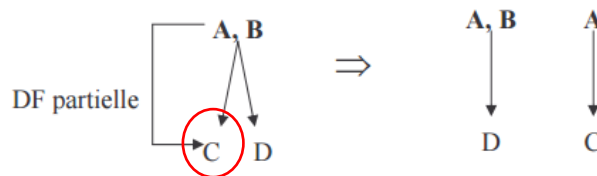
2FN

- Une relation est en 2FN si elle est en 1FN et si **chaque attribut non clé dépend** totalement et non partiellement **de la clé primaire** (c-à-d que la dépendance est élémentaire).



2FN

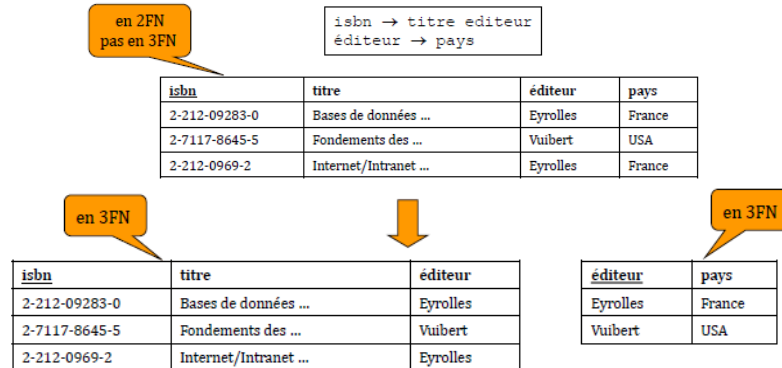
- Comment normaliser en 2FN:
 - Isoler la DF partielle dans une nouvelle relation
 - Eliminer l'attribut cible (le dépendant) de DF de la relation initiale



- Remarque:** Une relation en 1FN dont la clef est atomique (mono-attribut) est forcément en 2FN

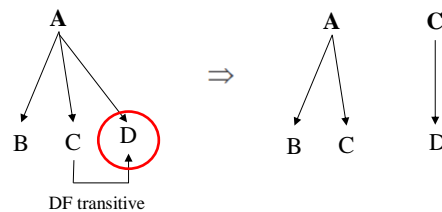
3FN

- Une relation est en 3FN si elle respecte la 2FN et si les **dépendances fonctionnelles** entre la clé primaire et les autres attributs sont **directes**.



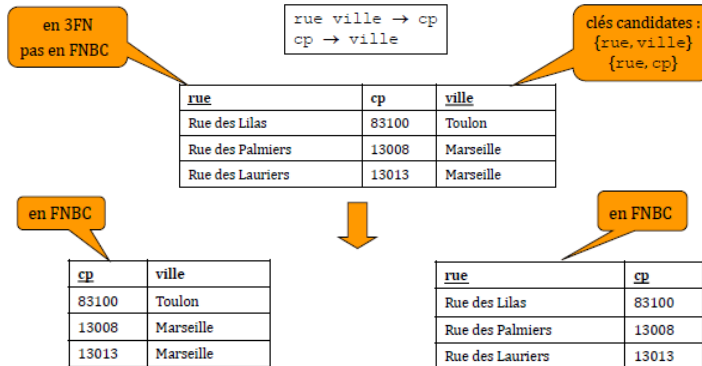
3FN

- Comment normaliser en 3FN:
 - Isoler la DF transitive dans une nouvelle relation
 - Eliminer la cible de DF transitive de la relation initiale



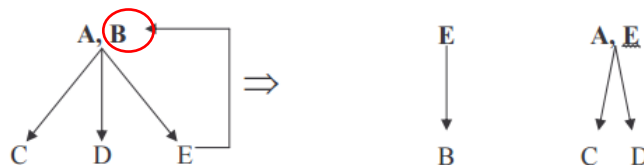
FNBC

- Une relation est en forme normale de Boyce-Codd si elle est en 3FN et que **le seul déterminant** (membre gauche des DF) existant dans la relation **est la clé primaire**.



FNBC

- Comment normaliser en FNBC:
 - Isoler la DF problématique dans une nouvelle relation
 - Éliminer la cible de DF problématique et la remplacer par sa source dans la relation initiale.



Exemple

- Considérons l'exemple des assurances de véhicules. Décomposer cette relation de manière à obtenir un schéma normalisé.

Ventes

num	nom	{Vehicules}					
		nimm	type	ncons	nomCons	dateAchat	prixAchat
1	Soutou	6748-XW-31	320 i	1	BMW	12-05-1999	62 500
		734-AJH-31	1200 GSF	3	SUZUKI	19-07-2000	46680
		358-ALZ-31	320 Ci	1	BMW	25-01-2001	213 970
2	Bidal	955-NEH-75	Scenic	2	RENAULT	16-01-1995	105 000
		52-AIM-31	Beetle	4	VOLKSWAGEN	16-01-2002	109 500

1FN

Ventes1

num	nom	nimm	type	ncons	nomCons	dateAchat	prixAchat
1	Soutou	6748-XW-31	320 i	1	BMW	12-05-1999	62 500
1	Soutou	734-AJH-31	1200 GSF	3	SUZUKI	19-07-2000	46680
1	Soutou	358-ALZ-31	320 Ci	1	BMW	25-01-2001	213 970
2	Bidal	955-NEH-75	Scenic	2	RENAULT	16-01-1995	105 000
2	Bidal	52-AIM-31	Beetle	4	VOLKSWAGEN	16-01-2002	109 500

- En partant de l'hypothèse qu'un assuré puisse vendre puis racheter sa voiture, la clé de cette relation sera la concaténation des attributs **num**, **nimm** et **dateAchat**.

2FN

- La question à se poser: « **De quoi dépend élémentairement chaque attribut de la relation ?** ».

Ventes1

num	nom	nimm	type	ncons	nomCons	dateAchat	prixAchat
1	Soutou	6748-XW-31	320 i	1	BMW	12-05-1999	62 500
1	Soutou	734-AJH-31	1200 GSF	3	SUZUKI	19-07-2000	46680
1	Soutou	358-ALZ-31	320 Ci	1	BMW	25-01-2001	213 970
2	Bidal	955-NEH-75	Scenic	2	RENAULT	16-01-1995	105 000
2	Bidal	52-AIM-31	Beetle	4	VOLKSWAGEN	16-01-2002	109 500

- Considérons tous les attributs :
 - num, nimm, dateAchat → prixAchat (prixAchat dépend de la clé) ;
 - nimm → ncons, type, nomCons;
 - num → nom.
- Les relations déduites de ces DF sont par définition en 2FN

2FN

Ventes2

num#	nimm#	dateAchat	prixAchat
1	6748-XW-31	12-05-1999	62 500
1	734-AJH-31	19-07-2000	46680
1	358-ALZ-31	25-01-2001	213 970
2	955-NEH-75	16-01-1995	105 000
2	52-AIM-31	16-01-2002	109 500

Assures2

num	nom
1	Soutou
2	Bidal

Vehicules2

nimm	type	ncons	nomCons
6748-XW-31	320 i	1	BMW
734-AJH-31	1200 GSF	3	SUZUKI
358-ALZ-31	320 Ci	1	BMW
955-NEH-75	Scenic	2	RENAULT
52-AIM-31	Beetle	4	VOLKSWAGEN

3FN

- Il faut à présent répondre à la question : « **De quoi dépend directement chaque attribut de la relation ?** ».
- Cherchons les DF indirectes. On trouve la DF **nimm** → **nomCons** qui est indirecte car issue d'une transitivité.
 - **nimm** → **ncons**
 - **ncons** → **nomCons**

3FN

Ventes3

num#	nimm#	dateAchat	prixAchat
1	6748-XW-31	12-05-1999	62 500
1	734-AJH-31	19-07-2000	46680
1	358-ALZ-31	25-01-2001	213 970
2	955-NEH-75	16-01-1995	105 000
2	52-AIM-31	16-01-2002	109 500

Assures3

num	nom
1	Soutou
2	Bidal

Vehicules3

nimm	type	ncons#
6748-XW-31	320 i	1
734-AJH-31	1200 GSF	3
358-ALZ-31	320 Ci	1
955-NEH-75	Scenic	2
52-AIM-31	Beetle	4

Constructeurs3

ncons	nomCons
1	BMW
3	SUZUKI
2	RENAULT
4	VOLKSWAGEN

FNBC

- Les relations sont déjà en FNBC !

Exercice 7

- On considère la relation suivante comme la base de données initiale d'une agence immobilière. Décomposer cette relation de manière à obtenir un schéma normalisé.

Num-Client	Nom-Client	NumApp	AdrApp	DateDLoc	DateFLoc	Montant	NumProp	NomProp
CR76	Jamal Khiyati	PG4	12, rue de la Gare	01.07.2003	31.08.2005	3500	CX40	Farid Assri
		PG16	7, av. République	01.09.2005	01.09.2010	4500	CX93	Ayoub Badr
CR56	Ahmed Zitouni	PG4	12, rue de la Gare	01.09.2002	10.06.2003	3500	CX40	Farid Assri
		PG36	3, Grande Rue	10.10.2003	01.12.2004	3800	CX93	Ayoub Badr
		PG16	7, av. République	01.01.2005	10.08.2005	4500	CX93	Ayoub Badr

- **Remarque :** On suppose qu'une personne peut louer plusieurs appartements en même temps mais elle ne peut pas relouer le même appartement une fois qu'elle le quitte.