

BUT Informatique
1A - Semestre 1
Introduction aux bases de données
(R1.05)

R. Fleurquin

Chapitre 2

Qualité des schémas relationnels et formes normales

L'information stockée « en vrac » ce n'est pas bien !

Statut	Code UAI	Établissement	DépCode	DépNom	Région	Académie	Commune	Sélectivité	Filière1	Filière2	Filière3	Filière4	Filière5	GPS
Public	0221036G	I.U.T de Lannion	22	Côtes-d'armor	Bretagne	Rennes	Lannion	formation sélective BUT	I.U.T de Lannion - BUT - Information communication	Parcours journalisme	BUT - Service	Information communication Parcours journalisme		48.7584, -3.45128
Public	0221036G	I.U.T de Lannion	22	Côtes-d'armor	Bretagne	Rennes	Lannion	formation sélective BUT	I.U.T de Lannion - BUT - Information communication	Parcours communication des organisations	BUT - Service	Information communication Parcours communication des organisations		48.7584, -3.45128
Public	0221036G	I.U.T de Lannion	22	Côtes-d'armor	Bretagne	Rennes	Lannion	formation sélective BUT	I.U.T de Lannion - BUT - Métiers du multimédia et de l'internet	Parcours : Développement web et dispositifs interactifs - Stratégie de communication numérique et design d'expérience (Seconde année possible en apprentissage)	BUT - Production	Métiers du multimédia et de l'internet	Développement web et dispositifs interactifs - Stratégie de communication numérique et design d'expérience	48.7584, -3.45128
Public	0292185T	I.U.T de Brest - Site de Morlaix	29	Finistère	Bretagne	Rennes	Morlaix	formation sélective BUT	I.U.T de Brest - Site de Morlaix - BUT - Gestion administrative et commerciale des organisations	Parcours : Management responsable de projet et des organisations - Management commercial et marketing omnicanal	BUT - Service	Gestion administrative et commerciale des organisations	Management responsable de projet et des organisations - Management commercial et marketing omnicanal	48.5841, -3.83541
Public	0290334F	I.U.T de Brest	29	Finistère	Bretagne	Rennes	Brest	formation sélective BUT	I.U.T de Brest - BUT - Gestion des entreprises et des administrations	Parcours : Gestion comptable, fiscale et financière - Gestion, entrepreneuriat et management d'activités - Contrôle de gestion et pilotage de la performance	BUT - Service	Gestion des entreprises et des administrations	Gestion comptable, fiscale et financière - Gestion, entrepreneuriat et management d'activités - Contrôle de gestion et pilotage de la performance	48.3989, -4.50379
Public	0221036G	I.U.T de Lannion	22	Côtes-d'armor	Bretagne	Rennes	Lannion	formation sélective BUT	I.U.T de Lannion - BUT - Informatique	Parcours : Réalisation d'applications : conception, développement, validation - Administration, gestion et exploitation des données (Seconde année possible en apprentissage)	BUT - Production	Informatique	Réalisation d'applications : conception, développement, validation - Administration, gestion et exploitation des données	48.7584, -3.45128
Public	0560861G	I.U.T de Vannes	56	Morbihan	Bretagne	Rennes	Vannes	formation sélective BUT	I.U.T de Vannes - BUT - Informatique	Parcours : Réalisation d'applications : conception, développement, validation - Administration, gestion et exploitation des données	BUT - Production	Informatique	Réalisation d'applications : conception, développement, validation - Administration, gestion et exploitation des données	47.6442, -2.77691

Une relation mal structurée rend la vie difficile!

- Des difficultés (souvent énormes!) de traitement sur les attributs multivalués
Exemple : trouver tous les BUT info offrant le parcours C...
- Des redondances entraînant une perte de place mémoire et donc des temps de lecture/écriture allongés
Exemple : Morlaix est dans le 29 => Finistère=> Bretagne => Académie de Rennes...
- Des problèmes d'incohérence lors de la mise à jour d'une information
Exemple : mettre les 803 BUT en formation non sélective
Exemple : si on oublie de changer le code UAI de toutes les « apparitions » de l'IUT de Lannion
- Des problèmes lors de la suppression d'une information
Exemple : on supprime l'académie de Rennes, que deviennent les formations de cette académie?
Exemple : on supprime la seule formation à Morlaix, que deviennent les informations sur Morlaix?

Théorie de la normalisation = Formes Normales ET Normalisation

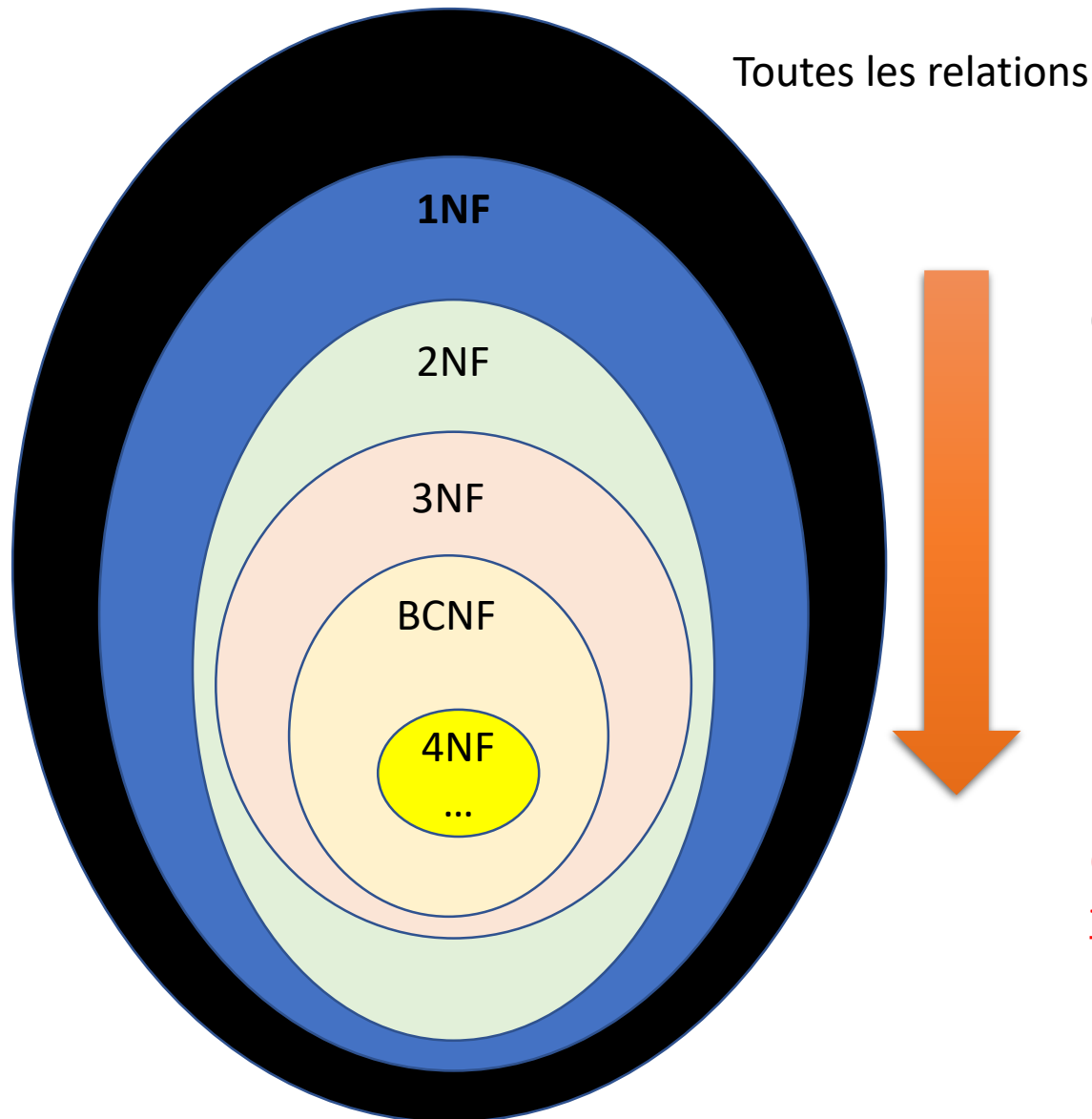


1. Déterminer le
niveau de qualité



2. Obtenir un
niveau de qualité

Les 8 formes normales : évaluer la redondance



Qualité croissante
(diminution des redondances)



Perte de contraintes et de performance
(jointures multiples)



Compromis : on vise des relations au minimum 3NF et c'est toujours possible sans perte : i) ni d'information ii) ni de df

L'outil théorique dont a besoin *la dépendance fonctionnelle (df)!*



$R(A_1, A_2, \dots, A_n)$ un schéma de la relation R et X et Y des sous-ensembles de $\{A_1, A_2, \dots, A_n\}$.
On dit que Y dépend (fonctionnellement) de X ssi :

Pour toute valeur de X il n'existe qu'une seule valeur possible de Y dans les tuples de R (dit autrement il existe une fonction f tel que $Y=f(X)$).

On note cette df : $X \rightarrow Y$

A	B
1	a
2	a
3	b
1	a

$A \rightarrow B$ est vraie!

Mais $B \rightarrow A$ est fausse!

Une df $X \rightarrow Y$ est de plus **élémentaire** ssi

i) Y est réduit à un seul attribut

ii) $Y \not\subset X$ (non trivialité)

iii) Aucun sous ensemble X' de X ne vérifie $X' \rightarrow Y$ (minimalité)

- $A \rightarrow A$, B non élémentaire : i) fausse
- $A \rightarrow A$ non élémentaire : ii) fausse
- $A, C \rightarrow B$ non élémentaire : iii) fausse
- $A, B \rightarrow A, B$ viole i), ii) et iii)
- $A \rightarrow B$ est élémentaire

Un petit dernier pour la route!

Commandes

n°Client	n°Produit	nomClient	villeClient	nomProduit	prixProduit	date
1	1	Paul	Vannes	perceuse	120	12/01
1	4	Paul	Vannes	scie	30	14/02
2	1	Mac	Nantes	perceuse	120	14/02
3	3	Mic	Nantes	marteau	8	15/03
3	1	Mic	Nantes	perceuse	120	15/03
3	2	Mic	Nantes	tournevis	8	15/03
4	3	Paul	Brest	marteau	30	2/01

Quelques df élémentaires

n°Produit → nomProduit

n°Client → villeClient

n°Client, n°Produit → date

Quelques df non « élémentaires »

n°Client, nomClient → villeClient (iii)

villeClient → villeClient (ii)

n°Client → nomClient, villeClient (i)

Fausse df!

nomClient → ville

prixProduit → nomProduit

date, prixProduit → nomProduit

Utilité des df élémentaires

Toutes les df d'une relation peuvent se retrouver par *axiomatique d'Armstrong* en partant uniquement du sous-ensemble de celles qui sont élémentaires.

Exemple : de $A \rightarrow B$ et $A \rightarrow C$ qui sont élémentaires on peut retrouver $A \rightarrow B, C$ ou $A, B \rightarrow C$ ou $A, D \rightarrow C, D$, etc.

Les df élémentaires **conservent donc toute l'information utile!**

Les df élémentaires sont l'outil théorique de base pour étudier le niveau de redondance des relations !

Première forme normale : 1NF

Atomique tu seras!

Un attribut est **atomique** si vis-à-vis du domaine il ne contient qu'une seule valeur informative pour un tuple donné, et donc s'il ne regroupe pas dans une même case un ensemble de plusieurs valeurs « unitairement intéressantes ».



Une relation est un 1NF si tous ses attributs sont atomiques.

Une relation qui n'est pas 1NF limite les traitements et augmentent considérablement les coûts de ces traitements.

num	nom	prénom
1	Jules	Pierre
2	Jules	Paul

1NF

num	nom	prénoms
1	Jules	Pierre, Marc
2	Jules	Paul, Léo

Pas 1NF

num	nom	adresse
1	Jules	4 rue Julia, 56000 Vannes
2	Ridard	7 avenue Jean, 56400 Ploeren

1NF?

Deuxième forme normale : 2NF

Ne jamais faire les choses à moitié !

Une relation est en 2NF si elle vérifie :

- i. Elle est en 1NF
- ii. Tout attribut n'appartenant à aucune clé candidate ne dépend pas d'une partie seulement d'une des clés candidates.

Autre définition : i) 1NF ii) toutes les df de source une clé sont élémentaires.

Astuce : une relation 1NF qui n'a que des clés candidates mono-attribut est toujours 2NF!

<u>num</u>	nom	prénom
1	Jules	Pierre
2	Jules	Paul
3	Jules	Paul

2NF

$\text{num} \rightarrow \text{nom}$

$\text{num} \rightarrow \text{prénom}$

<u>n°Prof</u>	<u>N°UE</u>	nomProf
1	2	Fleurquin
1	1	Fleurquin
2	1	Ridard
1	3	Fleurquin

1NF mais Pas 2NF

$\text{n°Prof} \rightarrow \text{nomProf}$

<u>numCom</u>	<u>numProd</u>	<u>numCli</u>	nomProd
1	4	1	stylo
2	3	3	gomme
3	2	2	feutre
4	2	3	feutre

1NF mais Pas 2NF

$\text{numProd} \rightarrow \text{nomProd}$

Troisième forme normale : 3NF

Les amis de mes amis ne sont pas mes amis!

Une relation est en 3NF si elle vérifie :

- i. elle est en 2NF
- ii. si tout attribut n'appartenant à aucune clé candidate ne dépend que des clés candidates.

Autre définition : i) 1NF ii) toutes les df élémentaires vers des attributs n'appartenant pas à une clé, sont issues d'une clé.

Astuce : une relation 2NF qui n'a qu'un seul attribut qui n'est pas membre d'une clé est toujours 3NF.

<u>num</u>	nom	prénom
1	Jules	Pierre
2	Jules	Paul
3	Heff	Pierre

3NF car 2NF (clé mono attribut) et les seules dfe sont
 $\text{num} \rightarrow \text{nom}$
 $\text{num} \rightarrow \text{prénom}$

<u>n°Etudiant</u>	<u>n°UE</u>	note
1	1	18
1	2	12
4	1	9
5	1	13

3NF car 2NF ($\text{n°Etudiant}, \text{n°UE} \rightarrow \text{note}$)
et un seul attribut hors clé)

<u>numEns</u>	nomEmp	grade	salaire
1	Fleurquin	EXC1	5000€
2	Portejoie	CN12	3000€
3	Adam	HC1	4000€
4	Le Sommer	CN12	3000€

NON 3NF bien que 2NF (clé mono-attribut)
mais il y a dfe : $\text{grade} \rightarrow \text{salaire}$

Forme normale de Boyce-Codd : BCNF

un chti oublié dans la def de la 3NF... les cibles clés!

Une relation est en BCNF si elle vérifie :

- elle est en 1NF
- Toutes les dépendances fonctionnelles élémentaires (dfe) sont du type $K \rightarrow X$. Avec K une clé et X un attribut quelconque (membre d'une clé ou non).

C'est la définition la plus simple à retenir!

Astuce : une relation 3NF qui n'a pas de dfe ciblant des attributs membre d'une clé est BCNF.

<u>num</u>	nom	prénom
1	Jules	Pierre
2	Jules	Paul

BCNF car 1NF et
les seules dfe sont
 $\text{num} \rightarrow \text{nom}$
 $\text{num} \rightarrow \text{prénom}$

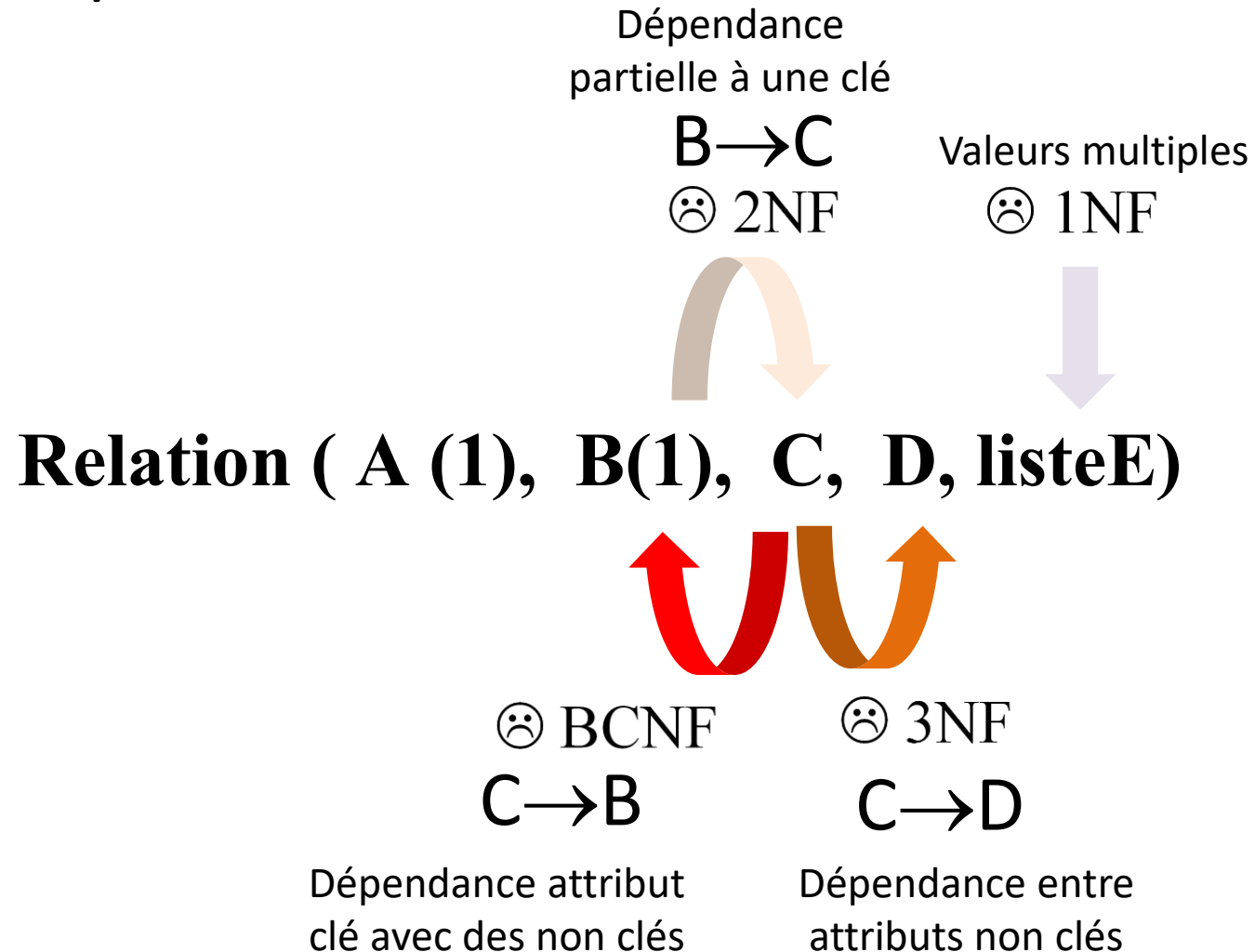
<u>n°Etudiant</u>	<u>n°UE</u>	note
1	1	18
1	2	12
4	1	9
5	1	12

BCNF car 1NF la seule dfe est
 $\text{n°Etudiant}, \text{n°UE} \rightarrow \text{note}$

<u>numPers</u>	<u>pays</u>	nom	région
1	France	Fleurquin	Bretagne
1	Italie	Oquendo	Lombardie
3	France	Pouit	Occitanie
4	France	Charles	Ile de France

3NF mais pas BCNF car dfe : $\text{région} \rightarrow \text{pays}$

On récapitule!



Oui et ensuite?

La théorie de la normalisation prouve qu'il est toujours possible de transformer une relation qui ne l'est pas en 3NF et cela sans

- i) perte d'information
- ii) sans perte de df.

Les formes normales d'ordre supérieur (BCNF, 4NF, etc.) ne peuvent plus garantir à tous les coups la non perte de df.

Bref, on vise la 3NF!

On ne va pas apprendre l'algorithme de normalisation mais utiliser un processus de modélisation qui, le plus souvent, conduit sans aucun effort à des schémas de bases de données au minimum 3NF : *l'analyse du discours via UML*.