

# L1 Sciences Fondamentales, S2 Normalisation

*Jean-François* COUCHOT  
Université de Franche-Comté, UFR-ST



# Plan



Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet





## Introduction à la normalisation

Motivations

Dépendances Fonctionnelles

Armstrong

Décomposition sans perte d'information

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet



# Plan

## Introduction à la normalisation

### Motivations

Dépendances Fonctionnelles

Armstrong

Décomposition sans perte d'information

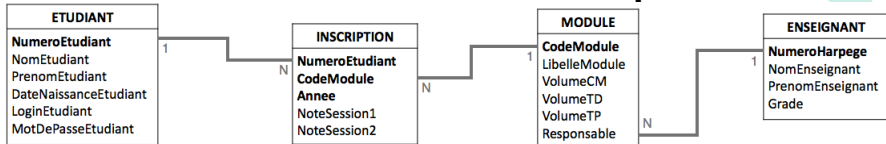
1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet



# Une seule relation universelle : exemple



Pourquoi pas n'utiliser qu'une seule relation ?

UNIVERSELLE(NumeroEtudiant\*, NomEtudiant, PrenomEtudiant, DateNaissanceEtudiant, LoginEtudiant, MotDePasseEtudiant, CodeModule\*, LibelleModule, VolumeCM, VolumeTD, VolumeTP, NumeroHarpege, NomEnseignant, PrenomEnseignant, Grade, Annee\*, NoteSession1, NoteSession2)

NumeroEtudiant*	NomEtudiant	...	CodeModule*	LibelleModule	...	NumeroHarpege	Annee	...
23794	Dornier	...	BD_L1	Bases de données	...	7914	2009	...
23794	Dornier	...	BD_L1	Bases de données	...	7914	2010	...
23794	Dornier	...	PROG_L1	Programmation	...	7358	2009	...
32911	Martin	...	BD_L1	Bases de données	...	7914	2010	...
32911	Martin	...	PROG_L1	Programmation	...	7358	2010	...
...	...	...	...	...	...	...	...	...

Conséquences générales sur les données

- ▶ Intégrité : un libellé de module unique pour un CodeModule ?
- ▶ Redondance : (NumeroEtudiant, NomEtudiant,...) répété de nombreuses x

# Une seule relation universelle : anomalies

NumeroEtudiant*	NomEtudiant	...	CodeModule*	LibelleModule	...	NumeroHarpege	Annee	...
23794	Dornier	...	BD_L1	Bases de données	...	7914	2009	...
23794	Dornier	...	BD_L1	Bases de données	...	7914	2010	...
23794	Dornier	...	PROG_L1	Programmation	...	7358	2009	...
32911	Martin	...	BD_L1	Bases de données	...	7914	2010	...
32911	Martin	...	PROG_L1	Programmation	...	7358	2010	...
...	...	...	...	...	...	...	...	...

## Insertions impossibles sauf si NULL autorisé

- ▶ Etudiant inscrit dans aucun module
- ▶ Enseignant responsable d'aucun module
- ▶ Module sans inscription

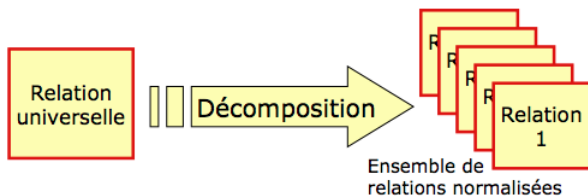
## Suppressions impossibles

- ▶ Etudiant, seul inscrit dans un module : perte des informations sur le module relatives au module...

# Normalisation : démarches possibles

## Approche par décomposition : vue en L1

1. A partir de toutes les données du SI : construction d'une relation universelle
2. Décomposition de cette relation universelle en relations (1NF, 2NF...)  
sans anomalie jusqu'à obtention d'un schéma normalisé



## Approche méthodique de construction : vue en L2

1. Construction d'un schéma relationnel avec méthode (dictionnaire des données, matrice de dépendance fonctionnelle).
2. Validation

# Décomposition : intuition $\rightsquigarrow$ pertes possibles

## Décomposition de $R(Att_1, Att_2, \dots)$

- ▶ Choix des ss-relations :  $R_1(Att_{1,1}, Att_{1,2}, \dots)$ ,  $R_2(Att_{2,1}, Att_{2,2}, \dots)$ , ... tq.  
 $\{Att_{1,1}, Att_{1,2}, \dots\} \cup \{Att_{2,1}, Att_{2,2}, \dots\} \cup \dots = \{Att_1, Att_2, \dots\}$
- ▶ Construct° de  $R_i$  : en projetant  $R$  sur les attributs de  $R_i$

## Exemples de décompositions

NumeroEtudiant	NomEtudiant	DateNaissanceEtudiant
27750	Martin	2-09-1989
32911	Martin	26-02-1998

### Décomposition 1

NumeroEtudiant	NomEtudiant
27750	Martin
32911	Martin

NumeroEtudiant	DateNaissanceEtudiant
27750	2-09-1989
32911	26-02-1998

Jointure selon NumeroEtudiant  $\rightsquigarrow R$   
(pas de perte)

### Décomposition 2

NumeroEtudiant	NomEtudiant
27750	Martin
32911	Martin

NomEtudiant	DateNaissanceEtudiant
Martin	2-09-1989
Martin	26-02-1998

Jointure selon NomEtudiant  $\rightsquigarrow$  perte !

NumeroEtudiant	NomEtudiant	DateNaissanceEtudiant
27750	Martin	2-09-1989
27750	Martin	26-02-1998
32911	Martin	2-09-1989
32911	Martin	26-02-1998





## Démarche générale de normalisation

- ▶ Pour éviter les anomalies
- ▶ Par décomposition d'une relation universelle
- ▶ Sans perte d'information

## Comment décomposer sans perte ?

- ▶ En ne regroupant dans une relation que les attributs "fortement liés"
- ▶ En éliminant le maximum de redondance et ne garder que le stricte nécessaire
- ▶ Notion de dépendance fonctionnelle





## Introduction à la normalisation

Motivations

Dépendances Fonctionnelles

Armstrong

Décomposition sans perte d'information

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet



# Dépendance fonctionnelle : définition et exemple

Définition pour  $X$  et  $Y$  deux sous-ensembles des attributs d'une rel.  $R$

- ▶ Dans  $R$ ,  $Y$  dépend fonctionnellement de  $X$  si \_\_\_\_\_
- ▶ Dit autrement : \_\_\_\_\_
- ▶ Notation :  $X \rightarrow Y$  qui se lit " \_\_\_\_\_ "
- ▶  $X$  à gauche : nommé \_\_\_\_\_
- ▶  $Y$  à droite : nommé \_\_\_\_\_

Exemple de dépendance fonctionnelle

NumeroEtudiant	NomEtudiant	DateNaissanceEtudiant
27750	Martin	2-09-1989
32911	Martin	26-02-1998

- ▶ Pour chaque numéro : un seul nom et une seule date de naissance
- ▶ Pour chaque nom : possible d'avoir plusieurs numéros d'étudiant
- ▶ Bilan :  $\{\text{NumeroEtudiant}\} \rightarrow \{\text{NomEtudiant}, \text{DateNaissanceEtudiant}\}$   
Simplifiée en :  $\text{NumeroEtudiant} \rightarrow \text{NomEtudiant}, \text{DateNaissanceEtudiant}$
- ▶ NumeroEtudiant : \_\_\_\_\_
- ▶ NomEtudiant, DateNaissanceEtudiant : \_\_\_\_\_

# Dépendance fonctionnelle totale (DFT)



Définition pour  $X$  et  $Y$  deux sous-ensembles d'attributs d'une rel.  $R$

$Y$  est en *dépendance fonctionnelle totale* (DFT) avec  $X$  si

- ▶ \_\_\_\_\_ et
- ▶ Pour tout \_\_\_\_\_, \_\_\_\_\_  
\_\_\_\_\_
- ▶ Dit autrement : \_\_\_\_\_
- ▶ Parfois, on peut lire "dépendance fonctionnelle minimale"

Exemples de dépendances fonctionnelles totales ou pas

- ▶ KO : NumeroEtudiant, CodeModule  $\rightarrow$  NomEtudiant : DF non totale
- ▶ OK : NumeroEtudiant  $\rightarrow$  NomEtudiant : DFT



# Fermeture des DF : motivation, définition



## Motivation

A partir des dépendances proposées, comment en inférer d'autres ? Et jusqu'à combien ?

## Définition : fermeture d'un ensemble de DF

L'ensemble de toutes les dépendances fonctionnelles qui sont inférées à partir d'un ensemble  $S$  de dépendances fonctionnelles est appelé *fermeture* de  $S$  et est dénoté  $S^+$ .

## Remarque

- ▶  $S^+$  : \_\_\_\_\_
- ▶ Comment construire  $S^+$  ? : à partir des règles d'Armstrong ...





## Introduction à la normalisation

Motivations

Dépendances Fonctionnelles

Armstrong

Décomposition sans perte d'information

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet



# Règles d'inférences d'Armstrong<sup>1</sup>

## Règles pour $X$ , $Y$ et $Z$ sous-ensembles d'attributs de $R$

- ▶ Réflexivité : \_\_\_\_\_
- ▶ Augmentation : \_\_\_\_\_
- ▶ Transitivité : \_\_\_\_\_
- ▶ Additivité : \_\_\_\_\_
- ▶ Fermeture  $S^+$  : \_\_\_\_\_

## Exemple d'application des règles d'inférence d'Armstrong

- ▶ Réflexivité à partir de  $\text{PrenomEtudiant}$ ,  $\text{DateNaissanceEtudiant}$  :  
\_\_\_\_\_
- ▶ Augmentation de  $\text{NoEtudiant} \rightarrow \text{NomEtudiant}$ ,  $\text{PrenomEtudiant}$  :  
\_\_\_\_\_  
\_\_\_\_\_
- ▶ Transitivité à p. de  $\text{CodeModule} \rightarrow \text{NoHarpege}$  et  $\text{NoHarpege} \rightarrow \text{NomEnseignant}$  : \_\_\_\_\_

---

1. Armstrong, W. W. (1974, August). Dependency Structures of Data Base Relationships. In IFIP congress (Vol. 74, pp. 580-583).

# A vous de jouer : DF

## Base de données de contrats d'assurance

- ▶ Un membre peut avoir plusieurs contrats, plusieurs numéros d'enfant.
- ▶ Données d'une relation universelle :

NoMem.	NomMem.	PrenomMem.	NoEnfant	NoCont.	CodeCont.	TypeCont.
1	Dupont	Jean	4	011	AssV1	AssuranceVoiture
1	Dupont	Jean	3	011	AssV1	AssuranceVoiture
2	Dupont	Marie	4	124	AssVi	AssuranceVie
2	Dupont	Marie	3	124	AssVi	AssuranceVie
1	Dupont	Jean	4	123	AssVi	AssuranceVie
1	Dupont	Jean	3	123	AssVi	AssuranceVie
1	Dupont	Jean	4	012	AssV2	AssuranceVoiture
1	Dupont	Jean	3	012	AssV2	AssuranceVoiture

## Ensemble $S$ des dépendances fonctionnelles et fermeture $S^+$ ?

- ▶  $S = \{$  \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_ }  
\_\_\_\_\_ }
- ▶  $S^+ = S \cup \{$  \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_ }  
\_\_\_\_\_ }



# Fermeture $X^+$ p.r. à l'ensemble $S$ des DF

## Fermeture $X^+$

Soit un ensemble d'attributs  $X$  de  $R$  et  $S$  l'ensemble des DF concernant  $R$ . La fermeture de  $X$  pour  $S$  est l'ensemble  $X^+$  des attributs qui dépendent fonctionnellement de  $X$ .

## Algorithme de calcul de la fermeture $X^+$

**entrées** :  $X$  : ensemble d'attributs,  $S$  : ensemble de dépendances fonctionnelles

**sortie** :  $X^+$  : fermeture de  $X$

$X^+ := X$  ; changé := vrai

**TANT QUE** changé **et** il existe une DF non utilisée **FAIRE**

    changé := faux

**POUR CHAQUE** DF :  $X' \rightarrow Y$  de  $S$  **FAIRE**

**SI**  $X' \subseteq X^+$  **ALORS**

$X^+ := X^+ \cup Y$

            marquer la DF comme "utilisée"

            changé := vrai

**FSI**

**FIN POUR**

**FIN TANT QUE**

# Clé candidate d'une relation universelle $R$



## Définition pour $X$ d'un ensemble d'attributs de $R$

$X$  une clé candidate si :

- ▶ Pour tout sous-ensemble d'attributs  $Y$  de  $R$  : \_\_\_\_\_  
\_\_\_\_\_
- ▶ La fermeture  $X^+$  : \_\_\_\_\_
- ▶  $X$  est irréductible : \_\_\_\_\_  
\_\_\_\_\_

## Remarques

- ▶  $X$  est une clé candidate de  $R$  si \_\_\_\_\_  
\_\_\_\_\_
- ▶ Attention : \_\_\_\_\_



# Calcul de $X^+$ sur l'exemple (1)



## Dépendances fonctionnelles de la relation universelle

- ▶ DF1 : NoEtud  $\rightarrow$  NomEtud, PrenomEtud, DateNaissEtud, LoginEtud, MDPEtud
- ▶ DF2 : LoginEtud  $\rightarrow$  NoEtud, NomEtud, PrenomEtud, DateNaissEtud, MDPEtud
- ▶ DF3 : NoHarpege  $\rightarrow$  NomEns, PrenomEns, Grade
- ▶ DF4 : CodeModule  $\rightarrow$  LibelleModule, VolCM, VolTD, VolTP, NoHarpege
- ▶ DF5 : NoEtud, CodeModule, Annee  $\rightarrow$  NoteSession1, NoteSession2

Montrons que  $X = \{\text{NoEtud}, \text{CodeModule}, \text{Annee}\}$  est une clé candidate de la relation.

## Initialisation

$X^+ = \{\text{NoEtud}, \text{CodeModule}, \text{Annee}\}$ ; changé := vrai



# Calcul de $X^+$ sur l'exemple (2)



1ère Itération :  $X^+ = \{\text{NoEtud}, \text{CodeModule}, \text{Annee}\}$  ; changé := vrai

1. DF1 : NoEtud  $\rightarrow$  NomEtud, PrenomEtud, DateNaissEtud, LoginEtud, MDPEtud
  - ▶ NoEtud \_\_\_\_\_ : déterminés de DF1 ajoutés dans  $X^+$  ; marquage de DF1 comme utilisée ; changé à vrai
  - ▶  $X^+ = \{ \text{NoEtud}, \text{CodeModule}, \text{Annee}, \text{_____} \}$
2. DF2 : LoginEtud  $\rightarrow$  NoEtud, NomEtud, PrenomEtud, DateNaissEtud, MDPEtud
  - ▶ LoginEtud \_\_\_\_\_ : déterminés de DF2 ajoutés dans  $X^+$ , \_\_\_\_\_ ; marquage de DF2 comme utilisée ; changé à vrai



# Calcul de $X^+$ sur l'exemple (3)

1ère Itération :  $X^+ = \{ \text{NoEtud, CodeModule, Annee, NomEtud, PrenomEtud, DateNaissEtud, LoginEtud, MDPEtud} \}$ ; changé := vrai

3. DF3 : NoHarpege  $\rightarrow$  NomEns, PrenomEns, Grade
  - ▶ NoHarpege \_\_\_\_\_ : DF3 non utilisée à cette itération ;  $X^+$  inchangée
4. DF4 : CodeModule  $\rightarrow$  LibelleModule, VoICM, VoITD, VoITP, NoHarpege
  - ▶ CodeModule \_\_\_\_\_ : déterminés de DF4 ajoutés dans  $X^+$  ; marque de DF4 comme utilisée ;
  - ▶  $X^+ = \{ \text{NoEtud, CodeModule, Annee, NomEtud, PrenomEtud, DateNaissEtud, LoginEtud, MDPEtud, _____, _____} \}$
5. DF5 : NoEtud, CodeModule, Annee  $\rightarrow$  NoteSession1, NoteSession2
  - ▶  $\{ \text{NoEtud, CodeModule, Annee} \}$  \_\_\_\_\_, déterminés de DF5 ajoutés dans  $X^+$  ; marquage de DF5 comme utilisée ;
  - ▶  $X^+ = \{ \text{NoEtud, CodeModule, Annee, NomEtud, PrenomEtud, DateNaissEtud, LoginEtud, MDPEtud, LibelleModule, VoICM, VoITD, VoITP, NoHarpege, _____} \}$

# Calcul de $X^+$ sur l'exemple (4)

2<sup>ème</sup> itération :  $X^+ = \{ \text{NoEtud, CodeModule, Annee, NomEtud, PrenomEtud, DateNaissEtud, LoginEtud, MDPEtud, LibelleModule, VoICM, VoITD, VoITP, NoHarpege, NoteSession1, NoteSession2} \}$  ;

1. DF3 : NoHarpege  $\rightarrow$  NomEns, PrenomEns, Grade

- ▶ NoHarpege \_\_\_\_\_ : déterminés de DF3 ajoutés dans  $X^+$  ; DF3 marquée comme utilisée ; changé à vrai
- ▶  $X^+ = \{ \text{NoEtud, CodeModule, Annee, NomEtud, PrenomEtud, DateNaissEtud, LoginEtud, MDPEtud, LibelleModule, VoICM, VoITD, VoITP, NoHarpege, NoteSession1, NoteSession2, _____} \}$

Fin de la 2<sup>ème</sup> itération.

Plus de DF non utilisée. Fin de l'algorithme.

$X = \{ \text{NoEtud, CodeModule, Annee} \}$  est une clé candidate car

- ▶  $X^+ = R$
- ▶  $X$  est irréductible : retirer un des attributs de  $X \rightsquigarrow X^+ \neq R$

# A vous de jouer : clés candidates

## Données

- ▶  $R(\text{NoMem}, \text{NomMem}, \text{PrenomMem}, \text{NoEnfant}, \text{NoCont}, \text{CodeCont}, \text{TypeCont})$
- ▶  $\text{DF1} : \text{NoMem} \rightarrow \text{NomMem}, \text{PrenomMem}$  ;  $\text{DF2} : \text{NoCont} \rightarrow \text{CodeCont}$
- ▶  $\text{DF3} : \text{CodeCont} \rightarrow \text{TypeCont}$   $\text{DF4} : \text{NoCont} \rightarrow \text{TypeCont}$

## Clés candidates d'une telle relation ?

- ▶ \_\_\_\_\_  
\_\_\_\_\_
- ▶ \_\_\_\_\_  
\_\_\_\_\_
- ▶ \_\_\_\_\_  
\_\_\_\_\_
- ▶ \_\_\_\_\_  
\_\_\_\_\_
- ▶ \_\_\_\_\_  
\_\_\_\_\_
- ▶ \_\_\_\_\_  
\_\_\_\_\_

# Attributs primaires ou non



## Définition pour une relation $R$

- ▶ Attributs primaires : \_\_\_\_\_  
\_\_\_\_\_
- ▶ Attributs non primaires : \_\_\_\_\_  
\_\_\_\_\_

## Attributs primaires ou non de la base de Scolarité

- ▶ Clés candidates : \_\_\_\_\_  
\_\_\_\_\_
- ▶ Attributs primaires : \_\_\_\_\_
- ▶ Attributs non primaires : \_\_\_\_\_







## Introduction à la normalisation

Motivations

Dépendances Fonctionnelles

Armstrong

Décomposition sans perte d'information

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet



# Décomposition sans perte d'information

Définition pour la relation  $R$  sur un ensemble d'attributs  $T$

- ▶ Soit  $B_1, B_2$  deux ensembles non vides d'attributs tels que  $B_1 \cup B_2 = T$  et
- ▶  $R_1 = [B_1]R$  et  $R_2 = [B_2]R$  forment une décomposition de  $R$
- ▶ Si la jointure naturelle \_\_\_\_\_ alors la décomposition est dite \_\_\_\_\_

Exemples de 2 décomposition de  $R = \{ \text{NoEtud, NomEtud, DateNaissEtud} \}$

NoEtud	NomEtud	DateNaissEtud
27750	Martin	2-09-1989
32911	Martin	26-02-1998

$B_1 = \{ \text{NoEtud, NomEtud} \}$

$B_1 = \{ \text{NoEtud, NomEtud} \}$

$B_2 = \{ \text{NoEtud, DateNaissEtud} \}$

$B_2 = \{ \text{NomEtud, DateNaissEtud} \}$

NoEtud	NomEtud
27750	Martin
32911	Martin

NoEtud	NomEtud
27750	Martin
32911	Martin

NoEtud	DateNaissEtud
27750	2-09-1989
32911	26-02-1998

NomEtud	DateNaissEtud
Martin	2-09-1989
Martin	26-02-1998

Jointure selon NoEtud : égale à  $R$

Jointure selon NomEtud : diff. de  $R$

↪ \_\_\_\_\_

↪ \_\_\_\_\_

# Décomposition sans perte d'information

Théorème de Heath pour la relation  $R$  sur un ensemble d'attributs  $T$

Soit  $X$ ,  $Y$  et  $Z$  une \_\_\_\_\_ de  $T$  avec \_\_\_\_\_. La décomposition de  $R$  en \_\_\_\_\_ et \_\_\_\_\_ est \_\_\_\_\_.

Preuve en posant  $B_1 = X \cup Y$  et  $B_2 = X \cup Z$

Soit  $R_1 = [B_1]R$  et  $R_2 = [B_2]R$ ; on a  $B_1 \cap B_2 = X$ ; montrons que  $R_1[X]R_2 = R$ .

- ▶ \_\_\_\_\_ Soit un  $n$ -uplet  $r \in R$ .
- ▶ Le projeté  $[X \cup Y]r$  appartient à  $[X \cup Y]R = R_1$ . De même  $[X \cup Z]r$  appartient à  $[X \cup Z]R = R_2$ .
- ▶ Les  $n$ -uplets  $[X \cup Y]r$  et  $[X \cup Z]r$  sont égaux sur  $X$ .
- ▶ Le  $n$ -uplet  $[X \cup Y]r[X][X \cup Z]r$  appartient donc à  $R_1[X]R_2$ .
- ▶ \_\_\_\_\_ Soit un  $n$ -uplet  $r \in R_1[X]R_2$ .
- ▶ Donc  $[X \cup Y]r \in R_1 = [X \cup Y]R$ ; il existe donc  $s_1 \in R$  tq.  $[X \cup Y]s_1 = [X \cup Y]r$ .
- ▶ Similairement  $[X \cup Z]r \in R_2 = [X \cup Z]R$ ; il existe donc  $s_2 \in R$  tq.  $[X \cup Z]s_2 = [X \cup Z]r$ .
- ▶ Donc  $[X]r = [X]s_1 = [X]s_2$ .
- ▶ Comme  $X \rightarrow Y$ ,  $[Y]s_1 = [Y]s_2$  et donc  $[X \cup Y \cup Z]s_2 = r \in R$ .

# Plan



Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

1NF

2NF

3NF

Boyce-Codd Normal Form (3.5 NF)

4NF et 5NF

Un exemple complet



# Les premières formes normales

## Procédure de normalisation

- ▶ Réalisée par réductions successives réversibles d'un ensemble de relations en une forme plus "satisfaisante"
- ▶ Construite à partir du concept de formes normales

## Forme normale : qu'est-ce et inclusions ?

- ▶ Une relation est dans une *forme normale* particulière si elle satisfait un ensemble de conditions prédéfinies.

$$1NF \supset 2NF \supset 3NF \supset BCNF \supset 4NF \supset 5NF$$

- ▶ 1NF, 2NF, 3NF :
  - ▶ introduites par Codd<sup>2</sup>
  - ▶ permettent la décomposition d'un ensemble de relations \_\_\_\_\_  
\_\_\_\_\_ et en utilisant \_\_\_\_\_

---

2. Further Normalization of the Data Base Relational Model. Research Report / RJ / IBM / San Jose, California RJ909 (1971)



Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

1NF

2NF

3NF

Boyce-Codd Normal Form (3.5 NF)

4NF et 5NF

Un exemple complet



# 1NF : atomicité des attributs



## Définition

- ▶ Une relation est 1NF si \_\_\_\_\_
- ▶ Un schéma relationnel est 1NF si \_\_\_\_\_

## Exemple : une relation ETUDIANT pas 1NF

NoEtu*	NomEtu	PrenomEtu	DateNaissEtu	Login	MDPEtu
23794	Dornier	Arnaud	18-08-1998	adornier	adornier
32911	Martin	Maxime	26-02-1998	mmartin	mmartin
33818	Schatt	Bastien, Charles	06-05-1998	bschatt	bschatt

Prénom pouvant contenir 2 valeurs : pas atomique, donc pas en 1NF.

## Décomposition d'une relation $R$ pour se conformer à 1NF

Décomposition \_\_\_\_\_



# A vous de jouer : 1NF ?

## A traiter

- ▶ *UNIVERSELLE*(NoMem, NomMem, PrenomMem, NoEnfant, NoCont, CodeCont, TypeCont)

NoMem.	NomMem.	PrenomMem.	NoEnfant	NoCont.	CodeCont.	TypeCont.
1	Dupont	Jean	4	011	AssV1	AssuranceVoiture
1	Dupont	Jean	3	011	AssV1	AssuranceVoiture
2	Dupont	Marie	4	124	AssVi	AssuranceVie
2	Dupont	Marie	3	124	AssVi	AssuranceVie
1	Dupont	Jean	4	123	AssVi	AssuranceVie
1	Dupont	Jean	3	123	AssVi	AssuranceVie
1	Dupont	Jean	4	012	AssV2	AssuranceVoiture
1	Dupont	Jean	3	012	AssV2	AssuranceVoiture

- ▶ Est-ce 1NF ?



\_\_\_\_\_







Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

1NF

2NF

3NF

Boyce-Codd Normal Form (3.5 NF)

4NF et 5NF

Un exemple complet



# 2NF : att. non prim. en DFT avec chq. clé candidate

## Définition

- ▶ Une relation  $R$  est 2NF si
  - ▶ elle est \_\_\_\_\_ et
  - ▶ chaque attribut non primaire : \_\_\_\_\_
- ▶ Un schéma relationnel est 2NF \_\_\_\_\_

## Décomposition d'une relation $R$ pour se conformer à 2NF

- ▶ \_\_\_\_\_ :
  - ▶ \_\_\_\_\_ décomposée en
  - ▶ \_\_\_\_\_
- ▶ Attention : \_\_\_\_\_
- ▶ Pratiquement : \_\_\_\_\_



# Exemple scolarité : mise en 2NF (1)

## Relation Universelle et $S^+$

$R(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu}, \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NoteSession1}, \text{NoteSession2}, \text{NomEns}, \text{PrenomEns}, \text{Grade})$

DF1 :  $\text{NoEtu} \rightarrow \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu}$

DF2 :  $\text{LoginEtu} \rightarrow \text{NoEtu}, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{MDPEtu}$

DF3 :  $\text{NoHarpege} \rightarrow \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF4 :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}$

DF4b :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade}$  (additivité)

DF5 :  $\text{NoEtu}, \text{CodeModule}, \text{Annee} \rightarrow \text{NoteSession1}, \text{NoteSession2}$

\_\_\_\_\_  $\rightsquigarrow R$  décomposée en



\_\_\_\_\_



$R_2(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NoteSession1}, \text{NoteSession2}, \text{NomEns}, \text{PrenomEns}, \text{Grade})$

## Exemple scolarité : mise en 2NF (2)

Relations  $R_1$ ,  $R_2$  et  $S^+$

$R_1(\text{NoEtu}^*, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu})$

$R_2(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NoteSession1}, \text{NoteSession2}, \text{NomEns}, \text{PrenomEns}, \text{Grade})$

DF1 :  $\text{NoEtu} \rightarrow \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu}$

DF2 :  $\text{LoginEtu} \rightarrow \text{NoEtu}, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{MDPEtu}$

DF3 :  $\text{NoHarpege} \rightarrow \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF4 :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}$

DF4b :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF5 :  $\text{NoEtu}, \text{CodeModule}, \text{Annee} \rightarrow \text{NoteSession1}, \text{NoteSession2}$

\_\_\_\_\_  $\rightsquigarrow R_2$  décomposée en



\_\_\_\_\_



\_\_\_\_\_  $R_4(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{NoteSession1}, \text{NoteSession2})$



# Exemple scolarité : mise en 2NF (3)

Relations  $R_1$ ,  $R_3$ ,  $R_4$  et  $S^+$

$R_1(\text{NoEtu}^*, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu})$

$R_3(\text{CodeModule}^*, \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade})$

$R_4(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{NoteSession1}, \text{NoteSession2})$

DF1 :  $\text{NoEtu} \rightarrow \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu}$

DF2 :  $\text{LoginEtu} \rightarrow \text{NoEtu}, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{MDPEtu}$

DF3 :  $\text{NoHarpege} \rightarrow \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF4 :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}$

DF4b :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF5 :  $\text{NoEtu}, \text{CodeModule}, \text{Annee} \rightarrow \text{NoteSession1}, \text{NoteSession2}$

## Remarques

- ▶  $R_1$  et  $R_3$  : \_\_\_\_\_
- ▶  $R_4$  : \_\_\_\_\_

# A vous de jouer : mise en 2NF

## Données

$R(\text{NoMem}^*, \text{NoEnfant}^*, \text{NoCont}^*, \text{NomMem}, \text{PrenomMem}, \text{CodeCont}, \text{TypeCont})$

DF1 :  $\text{NoMem} \rightarrow \text{NomMem}, \text{PrenomMem}$  ;

DF2 :  $\text{NoCont} \rightarrow \text{CodeCont}$  ;

DF3 :  $\text{CodeCont} \rightarrow \text{TypeCont}$  ;

DF4 :  $\text{NoCont} \rightarrow \text{TypeCont}$  ;

DF5 :  $\text{NoCont} \rightarrow \text{CodeCont}, \text{TypeCont}$

## 2NF ?

- ▶ 1 seule clé candidate :  $X = \{\text{NoMem}, \text{NoEnfant}, \text{NoCont}\}$
- ▶ Att. non prim. : \_\_\_\_\_
  - ▶  $\text{NomMem}, \text{PrenomMem}$  : \_\_\_\_\_
  - ▶  $\text{CodeCont}$  : \_\_\_\_\_
  - ▶  $\text{TypeCont}$  : \_\_\_\_\_

# A vous de jouer : mise en 2NF

## Données

$R(\text{NoMem}^*, \text{NoEnfant}^*, \text{NoCont}^*, \text{NomMem}, \text{PrenomMem}, \text{CodeCont}, \text{TypeCont})$

DF1 :  $\text{NoMem} \rightarrow \text{NomMem}, \text{PrenomMem}$  ;

DF2 :  $\text{NoCont} \rightarrow \text{CodeCont}$  ;

DF3 :  $\text{CodeCont} \rightarrow \text{TypeCont}$  ;

DF4 :  $\text{NoCont} \rightarrow \text{TypeCont}$  ;

DF5 :  $\text{NoCont} \rightarrow \text{CodeCont}, \text{TypeCont}$

## Mise en 2NF

- ▶ Décomposition de  $R$  selon DF1

- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_

# Plan



Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

1NF

2NF

3NF

Boyce-Codd Normal Form (3.5 NF)

4NF et 5NF

Un exemple complet





# 3NF : att. non prim. en dep. directe des clés candidates

## Définition

- ▶ Une relation  $R$  est 3NF si
  - ▶ elle est \_\_\_\_\_ et
  - ▶ chaque attribut non primaire : ne dépend \_\_\_\_\_ que \_\_\_\_\_  
\_\_\_\_\_
- ▶ Un schéma relationnel est 3NF \_\_\_\_\_

## Décomposition d'une relation $R$ pour se conformer à 3NF

- ▶ \_\_\_\_\_ :
- ▶ \_\_\_\_\_ décomposée en ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ Pratiquement : \_\_\_\_\_

# Exemple Sclolarité : mise en 3NF (1)

A partir du schéma 2NF et de  $S^+$

$R_1(\text{NoEtu}^*, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu})$

$R_3(\text{CodeModule}^*, \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade})$

$R_4(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{NoteSession1}, \text{NoteSession2})$

DF1 :  $\text{NoEtu} \rightarrow \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu}$

DF2 :  $\text{LoginEtu} \rightarrow \text{NoEtu}, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{MDPEtu}$

DF3 :  $\text{NoHarpege} \rightarrow \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF4 :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}$

DF4b :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF5 :  $\text{NoEtu}, \text{CodeModule}, \text{Annee} \rightarrow \text{NoteSession1}, \text{NoteSession2}$

\_\_\_\_\_  $\rightsquigarrow R_3$  décomposée en



\_\_\_\_\_



\_\_\_\_\_

\_\_\_\_\_



## Exemple Sclolarité : mise en 3NF (2)

A partir de  $R_1, R_4, R_5, R_6$  et  $S^+$

$R_1(\text{NoEtu}^*, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu})$

$R_4(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{NoteSession1}, \text{NoteSession2})$

$R_5(\text{NoHarpege}^*, \text{NomEns}, \text{PrenomEns}, \text{Grade})$

$R_6(\text{CodeModule}^*, \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege})$

DF1 :  $\text{NoEtu} \rightarrow \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu}$

DF2 :  $\text{LoginEtu} \rightarrow \text{NoEtu}, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{MDPEtu}$

DF3 :  $\text{NoHarpege} \rightarrow \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF4 :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}$

DF4b :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF5 :  $\text{NoEtu}, \text{CodeModule}, \text{Annee} \rightarrow \text{NoteSession1}, \text{NoteSession2}$

---

NomEtu, PrenomEtu, DateNaissEtu, MDPEtu (non primaires) : \_\_\_\_\_

---

# A vous de jouer : mise en 3NF (1)

Données 2NF et  $S^+$

$R_1(\text{NoMem}^*, \text{NomMem}, \text{PrenomMem})$

$R_3(\text{NoCont}^*, \text{CodeCont}, \text{TypeCont})$

$R_4(\text{NoMem}^*, \text{NoEnfant}^*, \text{NoCont}^*)$

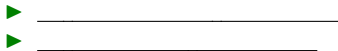
DF1 :  $\text{NoMem} \rightarrow \text{NomMem}, \text{PrenomMem}$  ;

DF2 :  $\text{NoCont} \rightarrow \text{CodeCont}$

DF3 :  $\text{CodeCont} \rightarrow \text{TypeCont}$

DF4 :  $\text{NoCont} \rightarrow \text{TypeCont}$

DF5 :  $\text{NoCont} \rightarrow \text{CodeCont}, \text{TypeCont}$



# A vous de jouer : mise en 3NF (2)



## Données 2NF et $S^+$

$R_1(\text{NoMem}^*, \text{NomMem}, \text{PrenomMem})$

$R_4(\text{NoMem}^*, \text{NoEnfant}^*, \text{NoCont}^*)$

$R_5(\text{CodeCont}^*, \text{TypeCont})$

$R_6(\text{NoCont}^*, \text{CodeCont})$

DF1 :  $\text{NoMem} \rightarrow \text{NomMem}, \text{PrenomMem}$  ;

DF2 :  $\text{NoCont} \rightarrow \text{CodeCont}$

DF3 :  $\text{CodeCont} \rightarrow \text{TypeCont}$

DF4 :  $\text{NoCont} \rightarrow \text{TypeCont}$

DF5 :  $\text{NoCont} \rightarrow \text{CodeCont}, \text{TypeCont}$

---



# Plan



Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

1NF

2NF

3NF

Boyce-Codd Normal Form (3.5 NF)

4NF et 5NF

Un exemple complet



# BCNF : déterminant de chaque DF est super clé

## Définition

- ▶ Une relation  $R$  est BCNF si
  - ▶ elle est \_\_\_\_\_ et pour chaque dépendance fonctionnelle  $Z \rightarrow Y$  de  $R$ , au moins une des conditions suivantes est vraie :
    - ▶  $Z \rightarrow Y$  est une \_\_\_\_\_ ( $Y \subseteq Z$ )
    - ▶ \_\_\_\_\_
  - ▶ Un schéma relationnel est BCNF \_\_\_\_\_

## Décomposition d'une relation $R$ pour se conformer à BCNF

- \_\_\_\_\_ :
- ▶ \_\_\_\_\_
  - ▶ \_\_\_\_\_,
- décomposée en  $\Rightarrow$
- ▶ \_\_\_\_\_
  - ▶ \_\_\_\_\_



# Exemple Sclolarité : mise en BCNF

A partir de  $R_1, R_4, R_5, R_6$  et  $S^+$  3NF

$R_1(\text{NoEtu}^*, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu})$

$R_4(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{NoteSession1}, \text{NoteSession2})$

$R_5(\text{NoHarpege}^*, \text{NomEns}, \text{PrenomEns}, \text{Grade})$

$R_6(\text{CodeModule}^*, \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege})$

DF1 :  $\text{NoEtu} \rightarrow \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu}$

DF2 :  $\text{LoginEtu} \rightarrow \text{NoEtu}, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{MDPEtu}$

DF3 :  $\text{NoHarpege} \rightarrow \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF4 :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}$

DF4b :  $\text{CodeModule} \rightarrow \text{LibelleModule}, \text{VolCM}, \text{VolTD}, \text{VolTP}, \text{NoHarpege}, \text{NomEns}, \text{PrenomEns}, \text{Grade}$

DF5 :  $\text{NoEtu}, \text{CodeModule}, \text{Annee} \rightarrow \text{NoteSession1}, \text{NoteSession2}$

Pour chaque  $R$  : \_\_\_\_\_

▶ dans  $R_1$

▶ \_\_\_\_\_

▶ \_\_\_\_\_

▶ dans  $R_4$ , \_\_\_\_\_

▶ dans  $R_5$ , \_\_\_\_\_

▶ dans  $R_6$ , \_\_\_\_\_



# A vous de jouer : mise en BCNF

## Données 3NF et $S^+$

$R_1(\text{NoMem}^*, \text{NomMem}, \text{PrenomMem})$

$R_4(\text{NoMem}^*, \text{NoEnfant}^*, \text{NoCont}^*)$

$R_5(\text{CodeCont}^*, \text{TypeCont})$

$R_6(\text{NoCont}^*, \text{CodeCont})$

DF1 :  $\text{NoMem} \rightarrow \text{NomMem}, \text{PrenomMem}$  ;

DF2 :  $\text{NoCont} \rightarrow \text{CodeCont}$

DF3 :  $\text{CodeCont} \rightarrow \text{TypeCont}$

DF4 :  $\text{NoCont} \rightarrow \text{TypeCont}$

DF5 :  $\text{NoCont} \rightarrow \text{CodeCont}, \text{TypeCont}$

---



# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Autres dépendances  $\rightsquigarrow$  autres formes normales

4NF

5NF

Un exemple complet



# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Autres dépendances  $\rightsquigarrow$  autres formes normales

4NF

5NF

Un exemple complet



# Autres dépendances



Formes normales 1NF, 2NF, 3NF et BCNF



Forme normale 4NF



Forme 5NF





Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Autres dépendances  $\rightsquigarrow$  autres formes normales

4NF

5NF

Un exemple complet



# Dépendance multivaluée $X \twoheadrightarrow Y$ d'une relation $R(T)$

Intuition pour  $X$ ,  $Y$  et  $Z$  une partition des attributs de  $R$

- ▶ Pour chaque valeur prise par  $X$  : \_\_\_\_\_  
\_\_\_\_\_

Définition<sup>3</sup> pour  $X$ ,  $Y$  et  $Z$  une partition de  $T$

- ▶  $(x, y, z)$  et  $(x, y', z')$  : \_\_\_\_\_
- ▶  $R$  vérifie la *dépendance multivaluée*  $X \twoheadrightarrow Y$  \_\_\_\_\_  
\_\_\_\_\_
- ▶ On dit \_\_\_\_\_

---

3. Fagin, R. (1977). Multivalued dependencies and a new normal form for relational databases. ACM Transactions on Database Systems (TODS), 2(3), 262-278.

# Dép. multivaluées : exemple avec la relation BAGUE

NoBague*	Taille*	Matiere*
1	40	or
1	41	or
1	42	or
2	40	or
2	42	argent
2	42	or
2	40	argent

## Dépendances fonctionnelles

- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_

## Dépendances multivaluées

▶ NoBague → Taille ?

- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_

# Dépendance multivaluée : décomposit<sup>o</sup> ss. perte

## Premier théorème de R. Fagin

- ▶ Dans la relation  $R(X, Y, Z)$ , \_\_\_\_\_ si et seulement si \_\_\_\_\_.
- ▶ Preuve : admis

## Conséquences

On peut décomposer  $R(X, Y, Z)$  \_\_\_\_\_ en deux projections \_\_\_\_\_ si \_\_\_\_\_.

Relation BAGUE(NoBague,Taille,Matiere) et NoBague  $\rightarrow$  Taille

NoBague*	Taille*	Matiere*
1	40	or
1	41	or
1	42	or
2	40	or
2	42	argent
2	42	or
2	40	argent

$\rightsquigarrow$

NoBague*	Taille*
1	40
1	41
1	42
2	40
2	42

NoBague*	Matiere*
1	or
2	or
2	argent



# Dépendance multivaluée : propriétés

## Dép. fonctionnelle implique dép. multivaluée

- ▶ Si \_\_\_\_\_, alors \_\_\_\_\_
- ▶ Preuve :  $X \rightarrow Y$  implique si deux tuples distincts  $(x, y, z)$  et  $(x, y, z')$  dans  $R$ , alors  $(x, y, z')$  et  $(x, y, z)$  dans  $R$
- ▶ Mais possible d'avoir \_\_\_\_\_ sans \_\_\_\_\_  
\_\_\_\_\_

## Pour une relation $R(T)$ , $X, Y, Z$ partition de $T$

- ▶ Complémentation : si \_\_\_\_\_, alors \_\_\_\_\_
- ▶ Preuve : si \_\_\_\_\_, d'après le 1er théorème de Fagin
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_

## Exemple de complémentation avec BAGUE(NoBague,Taille,Matiere)

Comme \_\_\_\_\_ alors \_\_\_\_\_

# 4NF

## Définition d'une dépendance multivaluée triviale

Une dépendance multivaluée  $X \twoheadrightarrow Y$  est *triviale* si  $X$  et  $Y$  (disjoints) forment une partition de  $T$

## Définition

Une relation  $R$  est *4NF* si

- ▶ elle est \_\_\_\_\_ et
- ▶ pour chaque dépendances multivaluée non triviale  $X \twoheadrightarrow Y$  de  $R$  :  
\_\_\_\_\_.

## Décomposition pour se conformer à 4NF

\_\_\_\_\_ :  
▶ \_\_\_\_\_  
▶ \_\_\_\_\_

décomposée en  $\Rightarrow$

▶ \_\_\_\_\_  
▶ \_\_\_\_\_



# Exemple Sclolarité : mise en 4NF



## Dépendances multivaluées ?

- ▶  $R_1(\text{NoEtu}^*, \text{NomEtu}, \text{PrenomEtu}, \text{DateNaissEtu}, \text{LoginEtu}, \text{MDPEtu})$
- ▶  $R_4(\text{NoEtu}^*, \text{CodeModule}^*, \text{Annee}^*, \text{NoteSession1}, \text{NoteSession2})$
- ▶  $R_5(\text{NoHarpege}^*, \text{NomEns}, \text{PrenomEns}, \text{Grade})$
- ▶  $R_6(\text{CodeModule}^*, \text{LibelleModule}, \text{VolumeCM}, \text{VolumeTD}, \text{VolumeTP}, \text{NoHarpege})$

Existe-t-il une dépendance multivaluée entre NoEtu, CodeModule et Annee ?

- ▶ Non : l'inscription d'un·e étudiant·e à un module une année est

\_\_\_\_\_

- ▶ \_\_\_\_\_



# A vous de jouer : mise en 4NF

Restriction à la relation  $R_4(\text{NoMem}^*, \text{NoEnfant}^*, \text{NoCont}^*)$

NoMem*	NoEnfant*	NoCont*
1	4	011
1	3	011
1	4	123
1	3	123
1	4	012
1	3	012
2	4	124
2	3	124

► Existence de dépendances multivaluées ?

► \_\_\_\_\_

►  $R_4$  est-elle 4NF ?

► \_\_\_\_\_

Mise en 4NF de  $R_4(\text{NoMem}^*, \text{NoEnfant}^*, \text{NoCont}^*)$  selon  $\text{NoMem} \rightarrow \text{NoEnfant}$

NoMem*	NoEnfant*
1	3
1	4
2	3
2	4

NoMem*	NoCont*
1	011
1	123
1	012
2	124



Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

**4NF et 5NF**

Autres dépendances  $\rightsquigarrow$  autres formes normales

4NF

**5NF**

Un exemple complet



# Motivation pour la 5NF-1

Une relation BAGUE différente du T.60

NoBague*	Taille*	Matiere*
1	41	or
2	41	or
2	42	or
1	42	or
2	41	argent

► Dép. multiv. ? \_\_\_\_\_

► \_\_\_\_\_

► \_\_\_\_\_ :

NoBague*	Taille*
1	41
2	41
2	42
1	42

NoBague*	Matiere*
1	or
2	or
2	argent

\_\_\_\_\_ :

NoBague*	Taille*	Matiere*
1	41	or
2	41	argent
2	41	or
2	42	or
2	42	argent
1	42	or

\_\_\_\_\_

\_\_\_\_\_

# Motivation pour la 5NF-2

Une relation BAGUE différente du T.60

NoBague*	Taille*	Matiere*
1	41	or
2	41	or
2	42	or
1	42	or
2	41	argent



\_\_\_\_\_

NoBague*	Taille*
1	41
2	41
2	42
1	42

NoBague*	Matiere*
1	or
2	or
2	argent

Taille*	Matiere*
41	or
42	or
41	argent



\_\_\_\_\_ :

NoBague*	Taille*	Matiere*
1	41	or
2	41	argent
2	41	or
2	42	or
2	42	argent
1	42	or

Taille*	Matiere*
41	or
42	or
41	argent



\_\_\_\_\_



BAGUE : \_\_\_\_\_

# Dépendances de jointure d'une relation $R(T)$

## Rappel sur la jointure naturelle entre les relations $R_1$ et $R_2$

- ▶ C'est l'ensemble de \_\_\_\_\_ dans  $R_1$  et  $R_2$   
dont \_\_\_\_\_
- ▶ Notation : \_\_\_\_\_

## Définition d'une dépendance de jointure d'une relation $R(T)$

- ▶  $\{X_1, X_2, \dots, X_n\}$  : sous-ensembles d'attributs de  $R$  \_\_\_\_\_  
\_\_\_\_\_ et \_\_\_\_\_ ( $\cup_{i=1}^n X_i = T$ )
- ▶  $R$  possède une dépendance de jointure pour  $\{X_1, X_2, \dots, X_n\}$  si  $R$  est  
égale à \_\_\_\_\_  
\_\_\_\_\_
- ▶ Notation de la dépendance de jointure : \_\_\_\_\_
- ▶  $*\{X_1, X_2, \dots, X_n\}$  est une dépendance de jointure \_\_\_\_\_ si \_\_\_\_\_  
\_\_\_\_\_



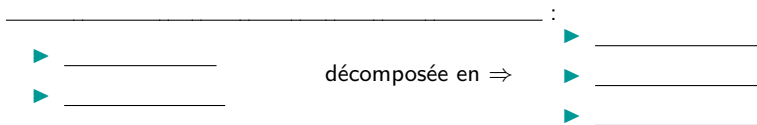


## Définition due à R. Fagin<sup>4</sup>

Une relation  $R$  est 5NF si

- ▶ elle est 4NF et
- ▶ pour toute dépendance de jointure  $*\{X_1, \dots, X_N\}$  non triviale,  $X_i$  est une super-clé de la relation  $R$

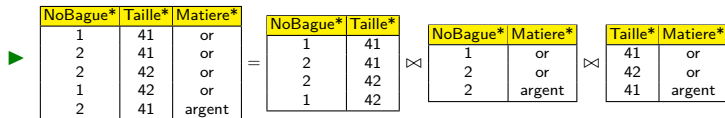
## Décomposition pour se conformer à 5NF



4. Fagin, R. (1979, May). Normal forms and relational database operators. In Proceedings of the 1979 ACM SIGMOD international conference on Management of data (pp. 153-160).

# Application à l'exemple BAGUE

## La relation BAGUE, selon T.63



► Dép. de jointure

► Pour chaque relation projetée : \_\_\_\_\_



# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet

Introduction au biathlon

Premier modèle par IA

Données

Normalisation

Manipulation des données



# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

Un exemple complet

Introduction au biathlon

Premier modèle par IA

Données

Normalisation

Manipulation des données



# Biathlon en quelques mots<sup>5</sup>

- ▶ du latin bi- (deux), et du grec athlon (combat, lutte) : sport qui combine ski de fond (endurance) et tir à la carabine (calme, adresse)
- ▶ disciplines individuelles :
  - ▶ en contre la montre : “sprint”, “individuel” dont les distances, le nombre de tirs et les pénalités sont définies par discipline et par genre
  - ▶ en poursuite : “poursuite” et “mass start” dont le nombre de tirs est toujours 4 (2 debouts et 2 couchés), mais où les départs sont différés dans le premier cas (selon les classements d’un sprint) où en groupe dans le second cas ; les distances ne sont pas les mêmes qu’en sprint
- ▶ disciplines par équipe :
  - ▶ relais par genre : chacun · e des 4 biathlètes effectue 3 tours, entrecoupés d’un tir debout et 1 tir couché. Les distances varient selon le genre
  - ▶ relais mixte : principe identique, mais où chaque équipe est composée de 2 hommes et 2 femmes qui courent dans cet ordre. Les distances varient selon le genre.

---

5. <https://fr.wikipedia.org/wiki/Biathlon>

# Biathlon en images



## en images

- ▶ Relais mixte d'Östersund, le 25/11/2023
- ▶ Sprint homme Annecy-Le Grand-Bornand, le 15/12/2022



# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

## Un exemple complet

Introduction au biathlon

Premier modèle par IA

Données

Normalisation

Manipulation des données



# Modélio–1 de chatgpt, 20/04/24 : discutable !



You

construit une base de données complète de biathlon



**ChatGPT**

La construction d'une base de données complète de biathlon implique la création de tables pour stocker diverses informations telles que les athlètes, les compétitions, les résultats, les pistes de biathlon, les équipes, etc. Voici une suggestion de schéma de base de données pour une base de données de biathlon :

## 1. Table "Athlètes":

- ID (clé primaire)
- Nom
- Prénom
- Date de naissance
- Nationalité
- Genre
- Pays

## 2. Table "Compétitions":

- ID (clé primaire)
- Nom de la compétition
- Lieu
- Date
- Type de compétition (par exemple, sprint, poursuite, individuelle, relais)
- Distance



# Modéli°–2 de chatgpt, 20/04/24 : discutable !

## 3. Table "Résultats":

- ID (clé primaire)
- ID de la compétition (clé étrangère faisant référence à la table "Compétitions")
- ID de l'athlète (clé étrangère faisant référence à la table "Athlètes")
- Position
- Temps de course
- Points
- Pénalités

## 4. Table "Pistes":

- ID (clé primaire)
- Nom de la piste
- Lieu
- Longueur
- Description

## 5. Table "Équipes":

- ID (clé primaire)
- Nom de l'équipe
- Pays
- Liste des membres de l'équipe (référence aux ID des athlètes)

# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

## Un exemple complet

Introduction au biathlon

Premier modèle par IA

Données

Normalisation

Manipulation des données





## Stade

- ▶ `stade_id` : identifiant du stade
- ▶ `stade_nom` : nom du stade
- ▶ Ville et Pays : ville et pays du stade
- ▶ `DF_STADE` : `stade_id` -> `stade_nom`, Ville, Pays

## Discipline

- ▶ `disci_id` : identifiant de la discipline
- ▶ `disci_nom` : nom de la discipline (sprint, relais mixte...)
- ▶ `nb_tours`, `nb_tirs` : # tours et # tirs à réaliser pour cette discipline
- ▶ `penalite_dist`, `penalite_duree` : distance ou durée de la pénalité
- ▶ `distance_tour` : distance en ski pour chaque tour (dépend du genre)
- ▶ `DF_DISCI_1` :  
`disci_id`->`disci_nom`,`nb_tours`,`nb_tirs`,`penalite_dist`,`penalite_duree`
- ▶ `DF_DISCI_2` : `disci_id`,`genre` -> `distance_tour`





## Biathlete

- ▶ `ath_id` : identifiant du · de la biathlète
- ▶ `ath_nom`, `ath_prenom`, `ath_naissance`, `genre`
- ▶ `ath_equipe` : équipe pour laquelle court le · la biathlète
- ▶ `DF_ATH` : `ath_id->ath_nom,ath_prenom,ath_naissance,genre,ath_equipe`

## Evénement

- ▶ `evt_id` : identifiant de l'événement
- ▶ `evt_nom` : nom de l'événement (Annecy le Grand Bornand)
- ▶ `evt_debut`, `evt_fin` : dates des début et fin de l'événement
- ▶ `DF_EV` : `evt_id -> evt_nom, evt_debut, evt_fin, stade_id`





## Epreuve (qui compose un événement)

- ▶ epr\_id : identifiant de l'épreuve
- ▶ epr\_date, epr\_heure : date et heure de l'épreuve
- ▶ DF\_EPR : epr\_id -> evt\_id, disci\_id, epr\_date, epr\_heure

## Participation

- ▶ part\_id : identifiant de la participation
- ▶ rang : rang de départ du · de la biathlète dans cette épreuve
- ▶ ski\_tour\_id, tir\_tour\_id : numéro du tour de ski ou de tir
- ▶ tir\_duree, noires, pioches : durées du tir, # cibles noires à la fin, # pioches utilisées
- ▶ ski\_durée : durée du tour à ski
- ▶ DF\_PART\_1 : part\_id -> epr\_id, ath\_id, rang
- ▶ DF\_PART\_2 : part\_id, tir\_tour\_id -> tir\_duree, noires, pioches
- ▶ DF\_PART\_3 : part\_id, ski\_tour\_id -> ski\_duree



# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

## Un exemple complet

Introduction au biathlon

Premier modèle par IA

Données

**Normalisation**

Manipulation des données



# Normalisation

## Relation UNIVERSELLE

[disci\_id, disci\_nom, genre, nb\_tours, nb\_tirs, distance\_tour, penalite\_dist, penalite\_duree, nb\_pioches, epr\_id, evt\_id, epr\_date, epr\_heure, evt\_nom, evt\_debut, evt\_fin, stade\_id, stade\_nom, Ville, Pays, part\_id, ath\_id, rang, ath\_nom, ath\_prenom, ath\_naiss, ath\_equipe, ski\_tour\_id, ski\_duree, tir\_tour\_id, tir\_duree, noires, pioches]

## Dépendances fonctionnelles

DF\_STADE : stade\_id -> stade\_nom, Ville, Pays

DF\_DISCI\_1 : disci\_id -> disci\_nom, nb\_tours, nb\_tirs, penalite\_dist, penalite\_duree

DF\_DISCI\_2 : disci\_id, genre -> distance\_tour

DF\_ATH : ath\_id -> ath\_nom, ath\_prenom, ath\_naissance, genre, ath\_equipe

DF\_EV : evt\_id -> evt\_nom, evt\_debut, evt\_fin, stade\_id

DF\_EPR : epr\_id -> evt\_id, disci\_id, epr\_date, epr\_heure

DF\_PART\_1 : part\_id -> epr\_id, ath\_id, rang

DF\_PART\_2 : part\_id, tir\_tour\_id -> tir\_duree, noires, pioches

DF\_PART\_3 : part\_id, ski\_tour\_id -> ski\_duree

## Clé candidate

▶ : \_\_\_\_\_

# Transitivités entre DF partir de `part_id`

## Dépendances fonctionnelles

DF\_STADE : `stade_id` -> `stade_nom`, Ville, Pays

DF\_DISCI\_1 : `disci_id`->`disci_nom`,`nb_tours`,`nb_tirs`,`penalite_dist`,`penalite_duree`

DF\_DISCI\_2 : `disci_id`,`genre` -> `distance_tour`

DF\_ATH : `ath_id`->`ath_nom`,`ath_prenom`,`ath_naissance`,`genre`,`ath_equipe`

DF\_EV : `evt_id` -> `evt_nom`, `evt_debut`, `evt_fin`, `stade_id`

DF\_EPR : `epr_id` -> `evt_id`, `disci_id`, `epr_date`, `epr_heure`

DF\_PART\_1 : `part_id` -> `epr_id`, `ath_id`, `rang`

DF\_PART\_2 : `part_id`, `tir_tour_id` -> `tir_duree`, `noires`, `pioches`

DF\_PART\_3 : `part_id`, `ski_tour_id` -> `ski_duree`

- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ `epr_id`, `evt_id`, `evt_nom`, `evt_debut`, `evt_fin`, `stade_id`, `stade_nom`,  
Ville, Pays, `disci_id`, `disci_nom`,`nb_tours`,`nb_tirs`, `penalite_dist`,  
`penalite_duree`, `epr_date`, `epr_heure`, `ath_id`,`ath_nom`,`ath_prenom`,  
`ath_naissance`,`genre`,`ath_equipe`,`rang`, `distance_tour`



# UNIVERSELLE 2NF ?

A partir de la DF trouvée au transparent précédent

part\_id -> epr\_id, evt\_id, evt\_nom, evt\_debut, evt\_fin, stade\_id, stade\_nom, Ville, Pays, disci\_id, disci\_nom, nb\_tours, nb\_tirs, penalite\_dist, penalite\_duree, epr\_date, epr\_heure, ath\_id, ath\_nom, ath\_prenom, ath\_naissance, genre, ath\_equipe, rang, distance\_tour



## Décomposition d'UNIVERSELLE

R1 : [part\_id, epr\_id, evt\_id, evt\_nom, evt\_debut, evt\_fin, stade\_id, stade\_nom, Ville, Pays, disci\_id, disci\_nom, nb\_tours, nb\_tirs, pénalité\_dist, nb\_pioches, epr\_date, epr\_heure, ath\_id, ath\_nom, ath\_prenom, ath\_naissance, genre, ath\_equipe, rang , distance\_tour]

R2 : [part\_id, ski\_tour\_id, ski\_duree, tir\_tour\_id, tir\_duree, noires, pioches]

# Focus sur R2

## Données concernant R2

R2 : [part\_id, ski\_tour\_id, ski\_duree, tir\_tour\_id, tir\_duree, noires, pioches] de clé candidate {part\_id, tir\_tour\_id, ski\_tour\_id}

DF\_PART\_2 : part\_id, tir\_tour\_id -> tir\_duree, noires, pioches

DF\_PART\_3 : part\_id, ski\_tour\_id -> ski\_duree



## Mise en 2NF de R2

- ▶ Décompositions successives de R2 en

SKI\_DUREE : [part\_id, ski\_tour\_id, ski\_duree]  
de clé candidate {part\_id, ski\_tour\_id}

TIR\_DUREE : [part\_id, tir\_tour\_id, tir\_duree, noires, pioches]  
de clé candidate {part\_id, tir\_tour\_id}

- ▶ Reste

R3 : [part\_id, ski\_tour\_id, tir\_tour\_id]  
de clé candidate {part\_id, ski\_tour\_id, tir\_tour\_id}



# 3NF-1 ?

## Ce qui est déjà 3NF

- ▶ SKI\_DUREE, TIR\_DUREE : \_\_\_\_\_  
\_\_\_\_\_
- ▶ R3 : \_\_\_\_\_

## Décomposition de R1 (pas 3NF en raison des transitivités) en

R1 : [part\_id, epr\_id, evt\_id, evt\_nom, evt\_debut, evt\_fin, stade\_id, stade\_nom, Ville, Pays, disci\_id, disci\_nom, nb\_tours, nb\_tirs, penalité\_dist, penalité\_duree, nb\_pioches, epr\_date, epr\_heure, ath\_id, ath\_no, ath\_prenom, ath\_naissance, genre, ath\_equipe, rang , distance\_tour]

DF\_STADE : stade\_id -> stade\_nom, Ville, Pays

---

DF\_DISCI\_1 : disci\_id->disci\_nom,nb\_tours,nb\_tirs,penalite\_dist,penalite\_duree

---

---

# 3NF-2 ?



## Poursuite de la décomposition à partir du reste

R4 : [part\_id, epr\_id, evt\_id, evt\_nom, evt\_debut, evt\_fin, stade\_id, disci\_id, epr\_date, epr\_heure, ath\_id, ath\_nom, ath\_prenom, ath\_naissance, genre, ath\_equipe, rang , distance\_tour]

DF\_DISCI\_2 : disci\_id, genre -> distance\_tour

---

Reste : R5 : [part\_id, epr\_id, evt\_id, evt\_nom, evt\_debut, evt\_fin, stade\_id, disci\_id, epr\_date, epr\_heure, ath\_id, ath\_nom, ath\_prenom, ath\_naissance, genre, ath\_equipe, rang]

DF\_ATH : ath\_id -> ath\_nom, ath\_prenom, ath\_naissance, genre, ath\_equipe

---

---



# 3NF-3 ?



## Fin de la décomposition

R6 : [part\_id, epr\_id, evt\_id, evt\_nom, evt\_debut, evt\_fin, stade\_id, disci\_id, epr\_date, epr\_heure, ath\_id, rang]

DF\_EV : evt\_id -> evt\_nom, evt\_debut, evt\_fin, stade\_id

---

Reste : R7: [part\_id, epr\_id, evt\_id, disci\_id, epr\_date, epr\_heure, ath\_id, rang]

DF\_EPR : epr\_id -> evt\_id, disci\_id, epr\_date, epr\_heure

---

Reste : \_\_\_\_\_





Tout est BCNF. Il suffit de vérifier chaque dépendance fonctionnelle incluse dans chaque relation.



# 4NF ? Focus sur R3

R3 : [part\_id, ski\_tour\_id, tir\_tour\_id] de clé candidate  
{part\_id,ski\_tour\_id, tir\_tour\_id}

## Intuition

Indépendance des valeurs de tir\_tour\_id par rapport à ski\_tour\_id

## Vérification (code python et valeurs)

```
import pandas as pd

df_biathlon = pd.read_csv("/home/biathlon.csv")
R3 = df_biathlon[['part_id', 'ski_tour_id',
                  'tir_tour_id']]
R3 = R3.drop_duplicates()
R3
```

part_id	ski_tour_id	tir_tour_id
9	1	1
9	1	2
9	2	1
9	2	2
9	3	1
9	3	2
1	1	1
1	1	2
1	2	1
1	2	2
1	3	1
1	3	2
10	1	1
10	1	2
10	2	1
10	2	2
10	3	1
10	3	2

# Décomposition de R3 en 4NF



## Dépendance multivaluée

- ▶ On a \_\_\_\_\_
- ▶ \_\_\_\_\_
- ▶ \_\_\_\_\_

## Intérêts de telles projections ?

- ▶  $[part\_id, ski\_tour\_id]_{R3}$  est \_\_\_\_\_  
SKI\_DUREE :  $[part\_id, ski\_tour\_id, ski\_duree]$
- ▶ similairement  $[part\_id, tir\_tour\_id]_{R3}$  est \_\_\_\_\_  
TIR\_DUREE :  $[part\_id, tir\_tour\_id, tir\_duree, noires, pioches]$
- ▶ Ces deux projections : \_\_\_\_\_





# Plan

Introduction à la normalisation

1NF, 2NF, 3NF, BCNF

4NF et 5NF

## Un exemple complet

Introduction au biathlon

Premier modèle par IA

Données

Normalisation

Manipulation des données



# Découper, exporter les csv à p. d'UNIVERSELLE

```
import pandas as pd

df_biathlon = pd.read_csv("/home/biathlon.csv")

TIR_DUREE = df_biathlon[['part_id', 'tir_tour_id', 'tir_duree',
                        'noires', 'pioches']]
TIR_DUREE = TIR_DUREE.drop_duplicates()
TIR_DUREE.to_csv("/home/TIR_DUREE.csv",
                quoting=csv.QUOTE_NONNUMERIC, index=False, header=False)

...

PARTICIPATION = df_biathlon[['part_id', 'epr_id', 'ath_id', 'rang']]
PARTICIPATION = PARTICIPATION.drop_duplicates()
PARTICIPATION.to_csv("/home/PARTICIPATION.csv",
                    quoting=csv.QUOTE_NONNUMERIC, index=False, header=False)

DISCIPLINE_DIST
```

# Import d'un CSV dans PhpMyAdmin-1



## Création de la structure de chaque table (STADE)

Navigation: [Parcourir](#) | [Structure](#) | [SQL](#) | [Rechercher](#) | [Insérer](#) | [Exporter](#) | [Importer](#) | [Opérations](#) | [Suivi](#) | [Déclencheurs](#)

Vue: [Structure de table](#) | [Vue relationnelle](#)

#	Nom	Type	Interclassement	Attributs	Null	Valeur par défaut	Commentaires	Extra	Action
<input type="checkbox"/>	1	stade_id	int		Non	Aucun(e)			<a href="#">Modifier</a> <a href="#">Supprimer</a> <a href="#">Plus</a>
<input type="checkbox"/>	2	stade_nom	varchar(256)	utf8mb4_0900_ai_ci	Non	Aucun(e)			<a href="#">Modifier</a> <a href="#">Supprimer</a> <a href="#">Plus</a>
<input type="checkbox"/>	3	Ville	varchar(256)	utf8mb4_0900_ai_ci	Non	Aucun(e)			<a href="#">Modifier</a> <a href="#">Supprimer</a> <a href="#">Plus</a>
<input type="checkbox"/>	4	Pays]	varchar(256)	utf8mb4_0900_ai_ci	Non	Aucun(e)			<a href="#">Modifier</a> <a href="#">Supprimer</a> <a href="#">Plus</a>



# Import d'un CSV dans PhpMyAdmin-2

## Import dans chaque table



## Importation dans la table « STADE »

### Fichier à importer :

Le fichier peut être compressé (gzip, bzip2, zip) ou non.

Le nom du fichier compressé doit se terminer par **.[format].[compression]**. Exemple : **.sql.zip**

Parcourir les fichiers :  STADE.csv (Taille maximale : 2 048kio)

Il est également possible de glisser-déposer un fichier sur n'importe quelle page.

Jeu de caractères du fichier :

### Importation partielle :

Permettre l'interruption de l'importation si la limite de temps configurée dans PHP est sur le point d'être atte

Ignorer ce nombre de requêtes (pour SQL), à partir du début :

### Autres options :

Activer la vérification des clés étrangères

### Format :