

Base de données

Séance 3
Normalisation

Normalisation

- Dans la phase de conception sur le **modèle entités-associations** et dans le **modèle relationnel** pour éviter :
 - la redondance de données
 - la perte de données
 - les incohérences
 - l'effondrement de performance des traitements

 - **Dépendance fonctionnelle !**
-

Relation en extension (table)

- Soit R la relation définie sur les ensembles A_1, A_2, \dots, A_n

$$R \subseteq A_1 \times A_2 \times \dots \times A_n$$

- L'extension de R est l'ensemble de tuples de la relation à une instance donnée :

A1	A2	...	An
a_{11}	a_{21}	...	a_{n2}
a_{11}	a_{21}	...	a_{nk}
...
a_{12}	a_{2j}	...	a_{nk}

$(a_{11}, a_{21}, \dots, a_{nk})$ est un n-uplet (ou tuple)

Relation en intension (MLD)

□ Schéma relationnel :

$$R(A1:D1, A2:D2, \dots, An:Dn)$$

avec :

- R: nom de la relation
 - A1, A2, ..., An : nom des attributs de la relation
 - D1, D2, ..., Dn : domaine des attributs
 - et des contraintes d'intégrités éventuelles
-

Définition

- **Dépendance fonctionnelle (DF):**
 - Soit $R(A_1, A_2, \dots, A_n)$ un schéma de relation, et X et Y des sous-ensembles de A_1, A_2, \dots, A_n
 - On dit que X **détermine** Y ou que Y **dépend fonctionnellement** de X si, et seulement si, des valeurs identiques de X impliquent des valeurs identiques de Y
 - On le note : $X \rightarrow Y$
-

Exemple

□ Ville(numville, nom, code_postal, population)

- numville → nom
 - numville → code_postal
 - numville → population
 - code_postal ~~→~~ nom
-

Définition

□ Dépendance fonctionnelle élémentaire

- Une DF élémentaire est une DF de la forme $X \rightarrow A$
 - où A est un attribut unique n'appartenant pas à X
 - et où il n'existe pas X' inclus au sens strict dans X (i.e. $X' \subset X$) tel que $X' \rightarrow A$
-

Exemple

- Soit la relation de schéma

Reduction(cru, client, type, remise) :

CRU	TYPE	CLIENT	REMISE
CHENAS	A	C1	3%
MEDOC	A	C2	5%
JULIENAS	B	C1	4%
CHENAS	A	C2	4%

- **cru, client** → **remise** est élémentaire :

- ~~cru~~ → remise : (CHENAS, A, C1, 3%) et (CHENAS, A, C2, 4%)

- ~~client~~ → remise : (CHENAS, A, C1, 3%) et (JULIENAS, B, C1, 4%)

Exemple

- Soit la relation de schéma

Reduction(cru, client, type, remise) :

CRU	TYPE	CLIENT	REMISE
CHENAS	A	C1	3%
MEDOC	A	C2	5%
JULIENAS	B	C1	4%
CHENAS	A	C2	4%

- **cru, client** → **type**

- est une DF

- mais n'est pas élémentaire : cru → type

Définitions

□ Dépendance fonctionnelle directe

- Une DF $X \rightarrow A$ est une DF directe s'il n'existe aucun attribut B tel que l'on puisse avoir $X \rightarrow B$ et $B \rightarrow A$
-

Exemple

□ Soit la relation de schéma

Voiture(nv, marque, type, puissance, couleur) :

NV	MARQUE	TYPE	PUISSANCE	COULEUR
AF234567	Renault	RME8	5	rouge
XN228756	Citroën	4ZX2	7	grise
DE998031	Peugeot	P307B	7	verte
ANA22201	Opel	OCA6	5	noir
FJ000075	Ford	FFI9	5	jaune
SI141481	Citroën	4ZX2	7	verte

Exemple

- Soit la relation de schéma **Voiture(nv, marque, type, puissance, couleur) :**

NV	MARQUE	TYPE	PUISSANCE	COULEUR
AF234567	Renault	RME8	5	rouge
XN228756	Citroën	4ZX2	7	grise
DE998031	Peugeot	P307B	7	verte
ANA22201	Opel	OCA6	5	noir
FJ000075	Ford	FFI9	5	jaune
SI141481	Citroën	4ZX2	7	verte

- On a $NV \rightarrow MARQUE$, $NV \rightarrow TYPE$, $NV \rightarrow PUISSANCE$, $NV \rightarrow COULEUR$
 - mais aussi $TYPE \rightarrow MARQUE$ et $TYPE \rightarrow PUISSANCE$
 - donc $NV \rightarrow MARQUE$, $NV \rightarrow PUISSANCE$ ne sont pas directes
 - et $NV \rightarrow TYPE$, $NV \rightarrow COULEUR$ sont directes
-

Règles de normalisation

- Quatre formes normales essentielles
 - 1FN, 2FN, 3FN et BCNF.
 - Définissent la façon d'utiliser les entités, les attributs ou certaines contraintes sur les dépendances.
 - Pour les modèles complexes : 4FN, 5FN, 6FN, FNDC
-

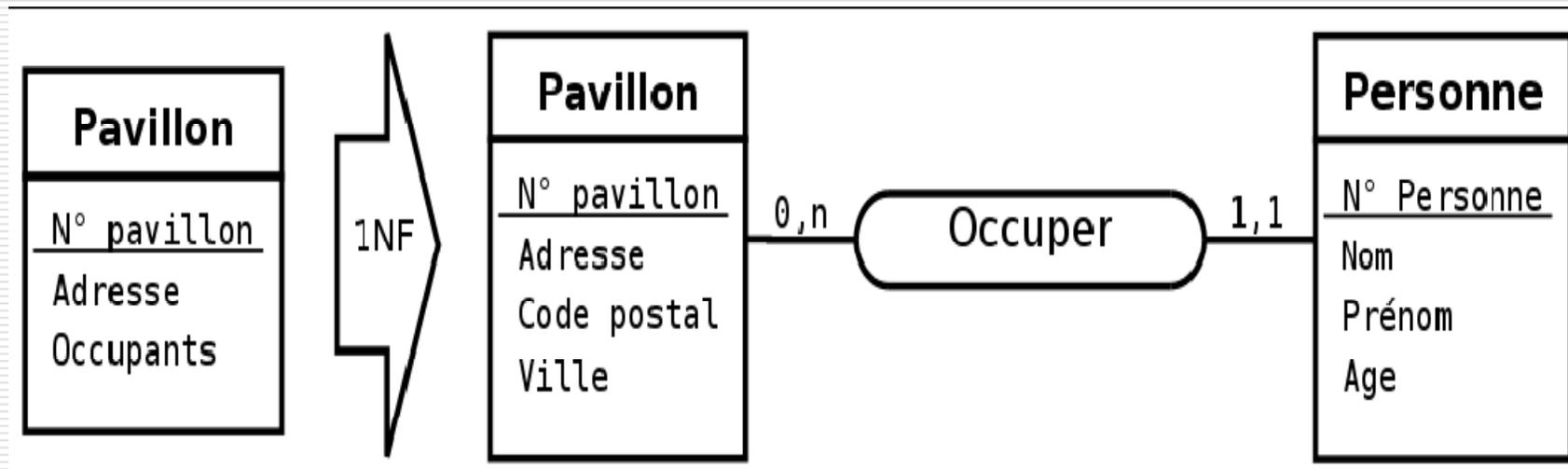
Première forme normale (1FN)

- Une relation est en 1FN si, et seulement si, tout attribut contient une valeur atomique (non multiple, non composée).

 - Exemple non en 1FN :
 - `Personne(numpersonne, nom, prenom, rue_et_ville, prenoms_enfants)`

 - `Personne(numpersonne, nom, prenom, rue, ville)`
 - `Prenom(numprenom, prenom)`
 - `Prenom_enfant(numpersonne, numprenom)`
-

Première forme normale (1FN)



Deuxième forme normale (2FN)

- Une relation est en 2FN si, et seulement si
 - elle est en 1FN
 - et si toutes les dépendances fonctionnelles entre la clé et les autres attributs sont élémentaires
- Exemple non en 2FN :
 - **Reduction(cru, client, type, remise)**

CRU	TYPE	CLIENT	REMISE
CHENAS	A	C1	3%
MEDOC	A	C2	5%
JULIENAS	B	C1	4%
CHENAS	A	C2	4%

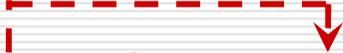
- **Remise(cru, client, remise) et Type(cru, type)**
-

Deuxième forme normale (2FN)

- Une relation est en 2FN si, et seulement si
 - elle est en 1FN
 - et si toutes les dépendances fonctionnelles entre la clé et les autres attributs sont élémentaires

- Exemple non en 2FN :

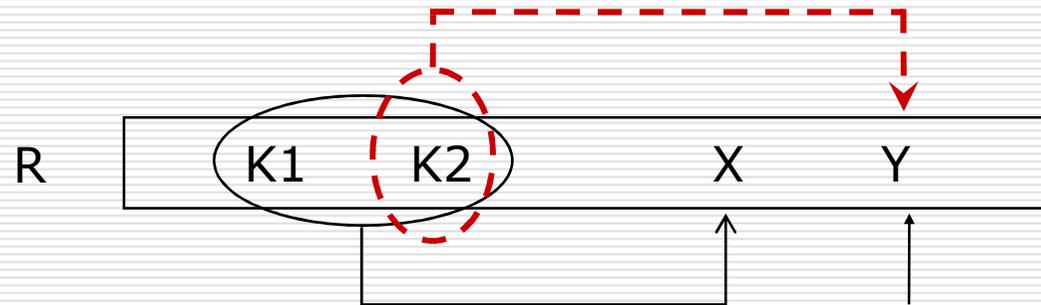
- Reduction(cru, client, type, remise)



CRU	TYPE	CLIENT	REMISE
CHENAS	A	C1	3%
MEDOC	A	C2	5%
JULIENAS	B	C1	4%
CHENAS	A	C2	4%

- Remise(cru, client, remise) et Type(cru, type)
-

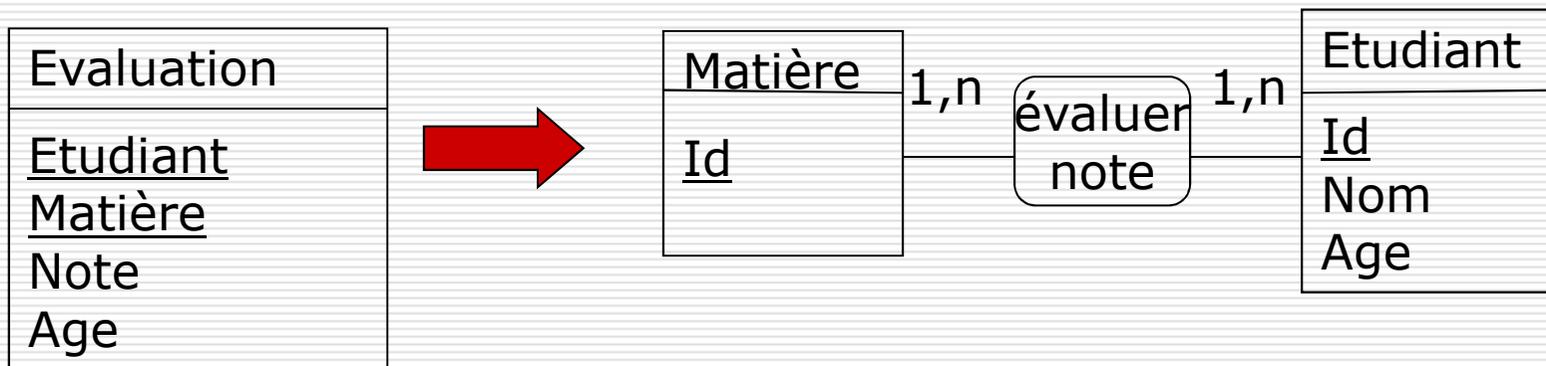
Relation non en 2FN



Une telle relation doit être décomposée en

$R1(\underline{K1}, \underline{K2}, X)$ et $R2(\underline{K2}, Y)$

Deuxième forme normale (2FN)



Troisième forme normale (3FN)

- Une relation est en 3FN si, et seulement si
 - elle est en 2FN
 - et si toutes les dépendances fonctionnelles entre la clé et les autres attributs sont directes

 - Exemple non en 3FN :
 - Voiture(nv, marque, type, puissance, couleur)

 - Vehicule(nv, type, couleur)
 - Modele(type, marque, puissance)
-

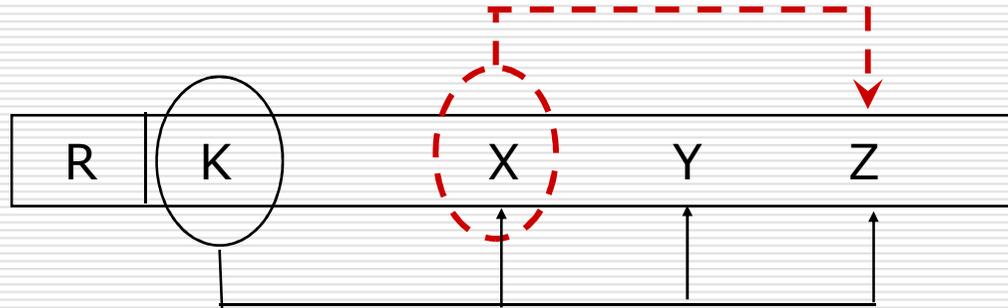
Troisième forme normale (3FN)

- Une relation est en 3FN si, et seulement si
 - elle est en 2FN
 - et si toutes les dépendances fonctionnelles entre la clé et les autres attributs sont directes

 - Exemple non en 3FN :
 - Voiture(nv, marque, type, puissance, couleur)

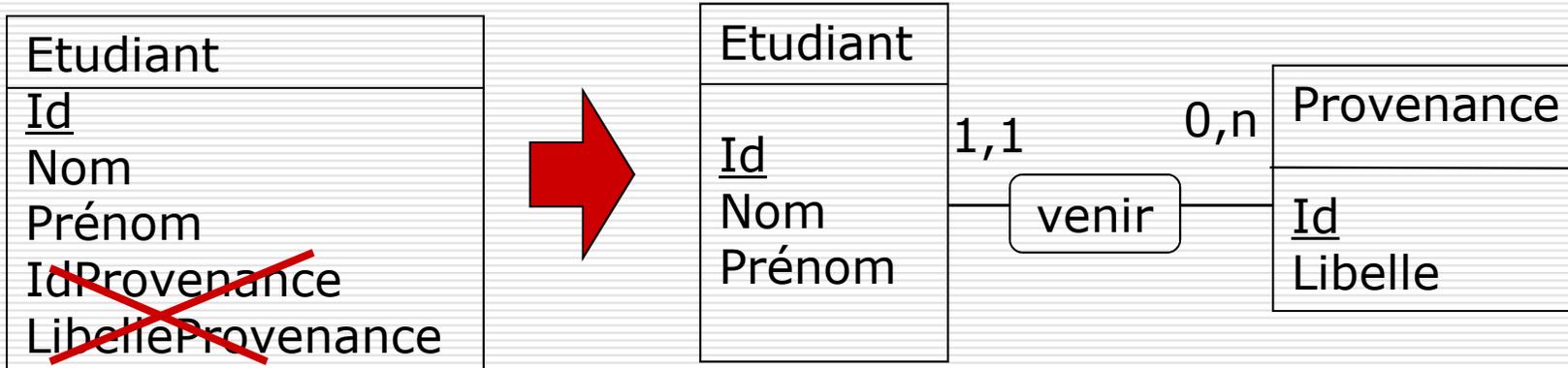
 - Vehicule(nv, type, couleur)
 - Modele(type, marque, puissance)
-

Relation non en 3FN



Une telle relation doit être décomposée en
 $R_1(\underline{K}, X, Y)$ et $R_2(\underline{X}, Z)$

Troisième forme normale (3FN)



libelleProvenance ne dépend pas directement de la clé mais d'un autre attribut

Forme normale de Boyce-Codd (BCNF)

- Une relation est en BCNF si, et seulement si
 - elle est en 3FN
 - et si les seules dépendances fonctionnelles élémentaires sont celles dans lesquelles une clé détermine un attribut.

 - Exemple
 - Université(numetudiant, nummatiere, numenseignant, note)
 - Si un enseignant n'enseigne qu'une seule matière :
numenseignant → nummatiere
-

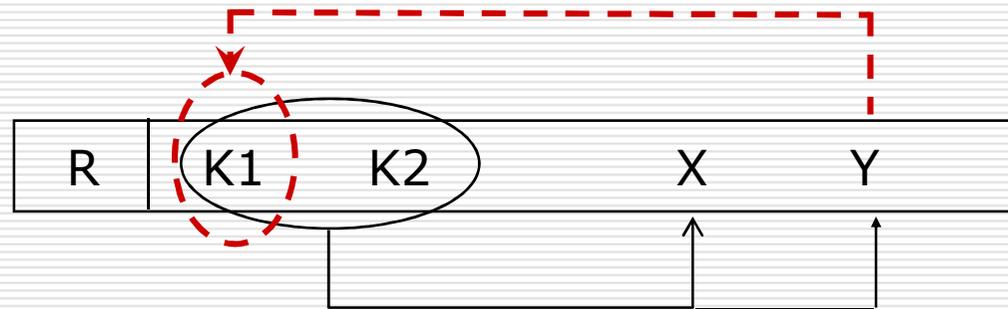
Forme normale de Boyce-Codd (BCNF)

- Université(numetudiant, nummatiere, numenseignant, note) n'est pas en BCNF => redondances de données

numetudiant	nummatiere	numenseignant	note
1	5	4	8
2	5	4	14
3	5	4	9,5
1	7	10	13

- Evaluation(numetudiant, nummatiere, note)
 - Enseignement(numenseignant, nummatiere)
-

Relation en 3FN mais non en BCNF



Une telle relation peut être décomposée en

$R1(\underline{K1}, \underline{K2}, \underline{X})$ et $R2(\underline{Y}, \underline{K1})$

Forme normale de Boyce-Codd (BCNF)

