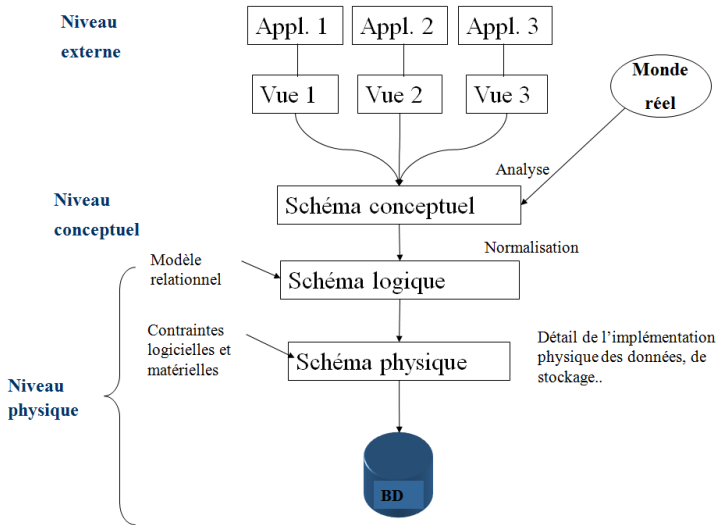


Normalisation

Dr. Mamadou Camara⁽¹⁾

⁽¹⁾ESP, Cheikh Anta Diop University, Dakar, Senegal
mamadou.camara@ucad.edu.sn

Module OMGL3



Problème

- ▶ Comment regrouper les attributs à l'intérieur des différents schémas de relation ?
- ▶ Un mauvais choix peut conduire à un certain nombre de problèmes tels que :
 - ▶ valeurs redondantes
 - ▶ Anomalies d'insertion,
 - ▶ Anomalies de suppression
 - ▶ Anomalies et de modification

Solution

- ▶ Il existe des concepts formels et théoriques permettant de définir plus précisément la qualité des schémas de relation.
- ▶ Formes normales définies à partir du concept de :
 - ▶ dépendances fonctionnelles : 1FN, 2FN, 3FN et FN de Boyce Codd
 - ▶ dépendances multi-valuées et dépendances de jointure (4FN et 5FN)

- ▶ Les dépendances fonctionnelles s'appliquent sur des schémas de relation individuels.
- ▶ Dépendance fonctionnelle DF = contrainte entre 2 ensembles. Soit une relation R , X et Y deux ensembles d'attributs et r l'ensemble des tuples de R.
- ▶ $X \rightarrow Y$ ssi $\forall (t_1, t_2) \in r^2, t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$
- ▶ On dit alors qu'il y a dépendance fonctionnelle de X vers Y ou encore que Y dépend fonctionnellement de X.
- ▶ Remarque : si X est une clé candidate de R alors $X \rightarrow Y$ pour tout sous-ensemble Y de R.

Exemple 1 :

A	B	C	D
a1	b1	c1	d1
a1	b2	c1	d2
a2	b2	c2	d2
a2	b3	c2	d3
a3	b3	c2	d4

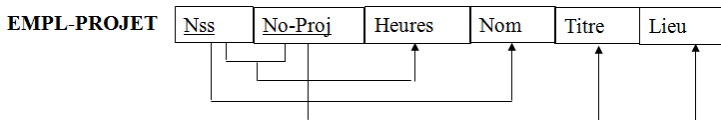
$A \rightarrow C$

~~$C \rightarrow A$~~

$AB \rightarrow D$

- Remarque : Une DF ne peut pas être automatiquement déduite d'une instance r de relation mais doit être explicitement définie par une personne maîtrisant la sémantique des attributs de R .

Exemple 1



$\{Nss, No-Projet\} \rightarrow Heures$

$No-Projet \rightarrow \{Titre, Lieu\}$

$Nss \rightarrow Nom$

$\underbrace{\hspace{10em}}_{\text{Membre gauche X}} \downarrow \text{DF} \underbrace{\hspace{10em}}_{\text{Membre droit Y}}$

- ▶ Chaque forme normal est caractérisée par un ensemble de propriétés que doit vérifier un schéma de relation.
- ▶ Pour chaque schéma de relation :
 - ▶ Choisir le degré de normalisation
 - ▶ Vérifier si le schéma de relation possède les propriétés correspondantes
 - ▶ si oui passer au degré suivant
 - ▶ si non traiter le schéma puis passer au degré suivant.

- ▶ En général quand une BD est en 3FN ou en BCNF, elle est considérée comme normalisée.

clé primaire

- ▶ Soit une relation $R(A_1, A_2, \dots, A_N)$
- ▶ Une super-clé S de R est un ensemble d'attributs de R tel que
 - ▶ $\forall \text{ instance } r \text{ de } R, \neg \exists (t_1, t_2) \in r^2 \text{ tel que } t_1[S] = t_2[S]$
- ▶ Une clé de R est une super-clé minimale.
- ▶ Si un schéma de relation possède plusieurs clés minimales, chacune d'elles est appelée clé candidate.
- ▶ L'une des clés candidates est choisie arbitrairement pour devenir la clé primaire de R tandis que les autres deviennent des clés secondaires.
- ▶ Chaque schéma de relation doit posséder une clé primaire.

attribut premier

- Un attribut d'un schéma de relation est dit premier s'il participe à une clé quelconque de R.

- └ Les trois premières formes normales
- └ Première forme normale 1FN :

- ▶ L'objectif de la 1FN est d'interdire :
 - ▶ les attributs multi-valués
 - ▶ Les attributs composés
 - ▶ La combinaison de ces derniers

- ▶ Exemple : Avec l'hypothèse qu'un même département peut être localisé à différents endroits → Ce schéma de relation n'est pas en 1FN puisque l'attribut Lieux-Dept n'est pas atomique.

DEPARTEMENT	Nom-Dept	<u>No-Dept</u>	Nss-Resp	Lieux-Dept
--------------------	----------	----------------	----------	------------

- ▶ 2 interprétations possibles de l'attribut "Lieux" :

- └ Les trois premières formes normales
- └ Première forme normale 1FN :

- Soit son domaine contient des valeurs atomiques, mais certains tuples peuvent posséder un ensemble de ces valeurs
 $\Rightarrow \text{No-Dept} \multimap \longrightarrow \text{Lieux}$
 - L'attribut est dit multi-valué

Nom-Dept	<u>No-Dept</u>	Nss-Resp	Lieu1	Lieu2	Lieu3
Informatique	1	123456	Dakar	Bambey	
Gestion	2	167426	Dakar		
Marketing	3	103308	Dakar	Saint-Louis	Ziguinchor

- └ Les trois premières formes normales
- └ Première forme normale 1FN :

- ▶ Soit chaque valeur de son domaine est constituée par un ensemble de valeurs et l'attribut n'est donc pas atomique \Rightarrow No-Dept \rightarrow Lieux
 - ▶ L'attribut est dit composé

Nom-Dept	<u>No-Dept</u>	Nss-Resp	Lieux
Informatique	1	123456	Dakar-Bambey
Gestion	2	167426	Dakar
Marketing	3	103308	Dakar-Saint-Louis-Ziguinchor

- └ Les trois premières formes normales
- └ Première forme normale 1FN :

- Transformation de DEPARTEMENT en 1FN : Éclater DEPARTEMENT en deux schémas de relations.

DEPARTEMENT (Nom-Dep, No-Dept, Nss-Resp, Lieu~~x~~-Dept)



DEPARTEMENT (Nom-Dept, No-Dept, Nss-Resp)
DEPT-LIEU (No-Dept, Lieu~~x~~u)

- └ Les trois premières formes normales
- └ Première forme normale 1FN :

Cas des attributs composés multi-valués

- Ce type d'attributs donne lieu à des relations imbriquées non autorisées en 1FN.

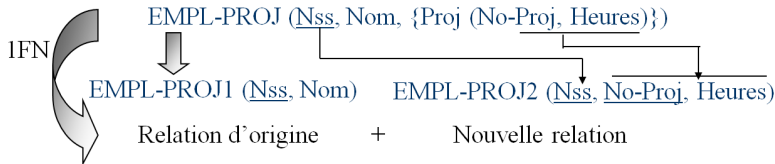
- └ Les trois premières formes normales
- └ Première forme normale 1FN :

Exemple 1

MAISON (Prometteur, Modèle)
MODELE (Style, Prix) $\xrightarrow{1FN}$ MAISON (Prometteur, Style, Prix)

- └ Les trois premières formes normales
- └ Première forme normale 1FN :

Exemple 2

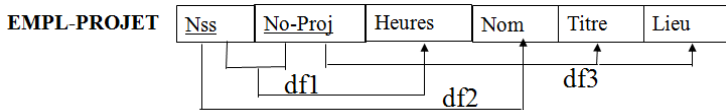


- └ Les trois premières formes normales
 - └ Première forme normale 1FN :

- Normalisation de ce type de schéma : La solution la plus consiste à reprendre le GDF.

- ▶ Basée sur le concept de dépendance fonctionnelle totale
- ▶ Une dépendance fonctionnelle $X \rightarrow Y$ est dite totale ssi
 - ▶ $\forall A \in X, (X - \{A\}) \rightarrow Y$
- ▶ Une dépendance fonctionnelle est dite partielle ssi
 - ▶ $\exists A \in X / (X - \{A\}) \rightarrow Y$
- ▶ Un schéma de relation R est en 2FN si chaque attribut non premier A de R est fonctionnellement dépendant de la clé primaire de R de façon totale.

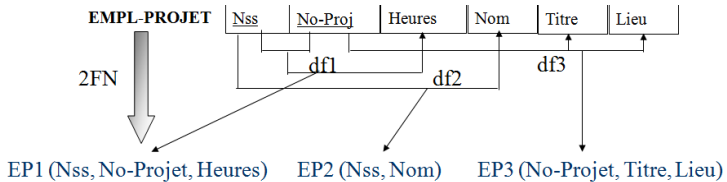
- └ Les trois premières formes normales
- └ Deuxième forme normale 2FN :



- Le schéma de relation EMPL-PROJ est en 1FN mais non en 2FN car :
 - L'attribut non premier "Nom" ne dépend pas totalement de la clé primaire d'après df2.
 - Les attributs non premiers "Titre" et "Lieu" ne dépendent pas totalement de la clé primaire d'après df3.

- └ Les trois premières formes normales
 - └ Deuxième forme normale 2FN :

- Un schéma de relation qui n'est pas en 2FN peut être normalisé en un ensemble de schémas de relations en 2FN
- Les attributs non premiers ne sont associés qu'à la partie de la clé primaire de laquelle ils dépendent fonctionnellement de façon totale.



- └ Les trois premières formes normales
- └ Deuxième forme normale 2FN :

TD

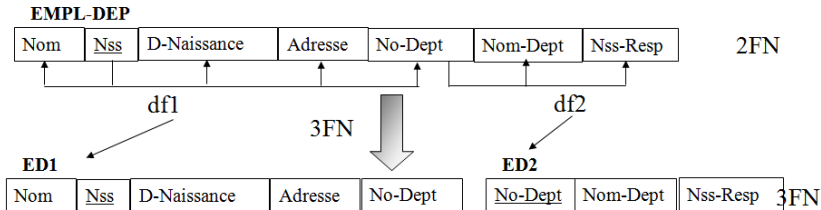
- ▶ TD 1 NF
- ▶ TD 2 NF

- ▶ Basée sur le concept de dépendances transitives.
- ▶ Une dépendance fonctionnelle $X \rightarrow Y$ dans un schéma de relation R est une dépendance transitive s'il existe un ensemble Z d'attributs tel que :
 - ▶ Z n'est ni la clé candidate ni un sous-ensemble d'une des clés de la relation.
 - ▶ $X \rightarrow Z$ et $Z \rightarrow Y$ sont vérifiées

- Un schéma de relation est en 3FN s'il est en 2FN et si aucun attribut non premier n'est en dépendance transitive avec la clé primaire.

Exemple : EMPL-DEPT

- ▶ Le schéma de relation EMPL-DEPT n'est pas en 3FN à cause des dépendances transitives de Nss sur Nss-Resp et de Nom-Dept via No-Dept.
 - ▶ No-Dept n'est pas un sous-ensemble de la clé primaire
- ▶ EMPL-DEPT a donc été décomposé en 2 schémas de relations ED1 et ED2 qui sont en 3FN.



- └ Les trois premières formes normales
- └ Troisième forme normale 3FN

TD

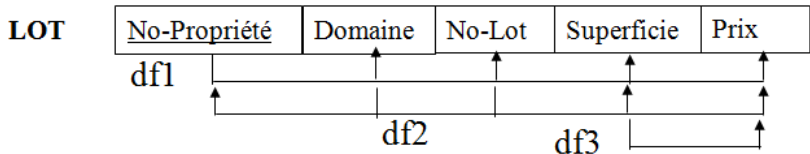
- ▶ TD 3 NF
- ▶ TD Décomposition

Définition

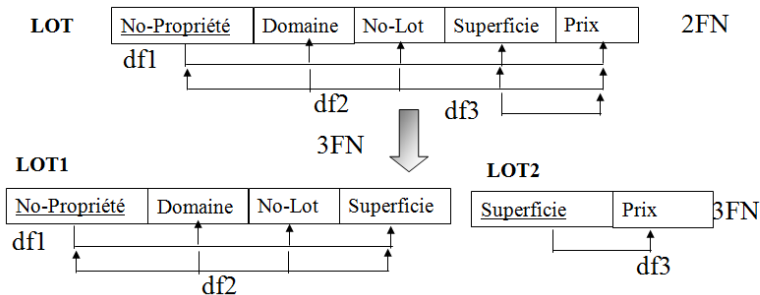
- ▶ C'est une définition plus générale de la 3FN (prise en compte des clés candidates et non seulement de la clé primaire).
- ▶ Un schéma de relation est en 3FN ssi
 - ▶ $\forall X \rightarrow A$ vérifiée sur R, "X est une super-clé de R"
 - ▶ ou "A est un attribut premier de R".
- ▶ Définition de la forme normale de Boyce Codd. Un schéma de relation est en BCNF ssi
 - ▶ $\forall X \rightarrow A$ vérifiée sur R, "X est une super-clé de R"

Exemple : Description

- ▶ Soit le schéma de relation et les DF suivantes. Hypothèses :
 - ▶ No-Propriété et (Domaine, No-Lot) sont 2 clés candidates.
 - ▶ No-Propriété est choisi arbitrairement comme clé primaire.
- ▶ LOT n'est pas en 3FN à cause de df3 puisque
 - ▶ "Superficie" n'est pas une super-clé de LOT
 - ▶ "Prix" n'est pas un attribut premier de LOT.



Exemple : Normalisation en 3 NF



Exemple : hypothèses additionnelles

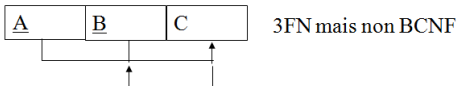
- ▶ Pour mieux percevoir la nécessité de la BCNF, supposons :
 - ▶ Qu'il y ait des centaines de lots de terrains répartis seulement sur deux domaines D1 et D2.
 - ▶ Que les superficies des lots prennent des valeurs dans 12 valeurs possibles :
 - ▶ 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 ha pour le domaine D1
 - ▶ 1.1, 1.2, 1.3, 1.4, 1.5, 1.6 ha pour le domaine D2.

Exemple : BCNF non respectée

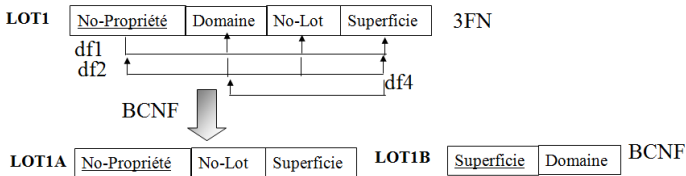
- ▶ DF additionnelle suivante df4 : Superficie \rightarrow Domaine
- ▶ LOT1 est toujours en 3FN puisque "Domaine" est un attribut premier.
- ▶ Intuitivement, on constate qu'il existait un schéma de relation R (Superficie, domaine) (12 tuples au plus) qui permettrait d'éviter la répétition de la même information des centaines de fois dans les tuples de LOT1.
 - ▶ Nécessité de définir une nouvelle forme normale qui permettrait de palier à l'insuffisance de la 3FN.
 - ▶ Dans l'exemple précédent LOT1 n'est pas en BCNF à cause de df4 car "Superficie" n'est pas une super-clé de LOT1.

- └ Les trois premières formes normales
- └ Forme normale de Boyce Codd

Exemple : Normalisation en 3 BCNF



Normalisation de LOT1 :



- ▶ La plupart des schémas de relation en 3FN le sont aussi en BCNF
- ▶ Lorsque ce n'est pas le cas le BCNF est préférable à la 3FN
- ▶ Les 2FN et 1FN ne sont pas considérées comme des degrés suffisants de normalisation.

Un schéma de relation R est en :

1FN	Si tous les attributs sont atomiques
2FN	Si chaque attribut non primaire A de R est fonctionnellement dépendant de la clé primaire de R de façon totale
3FN	S'il est en 2FN et si aucun attribut non primaire n'est en dépendance transitive avec la clé primaire
BCNF	Ssi $\forall X \rightarrow A$ vérifiée sur R, X est une super-clé de R

TD

► TD BCNF



References I