MASTER INFORMATIQUE
DEPARTEMENT D'INFORMATIQUE
(FSEA)

Année académique 2023-2024

Chapitre 5

DEPENDANCES FONCTIONNELLES et NORMALISATION DES SCHEMAS

Que faire alors?

- **Objectif de la normalisation**
  - supprimer les anomalies précédentes
  - élimination des redondances (éviter les incohérences + minimiser l'espace de stockage)

- **Normaliser une relation** consiste à *décomposer une relation ayant des anomalies en plusieurs sous-relations sans anomalies*

- **Normalisation** s'inscrit dans la partie conception d'une base de données -> peut être vu comme un outil théorique de vérification

- **Normalisation** repose sur les liens sémantiques entre attributs -> dépendances fonctionnelles

Dépendances Fonctionnelles(DF)

- [Déf] Soit R une relation, X et Y 2 ss-ens d'attributs,
  Y dépend fonctionnellement de X
  ou X détermine fonctionnellement Y
  ssi $\forall$r extension de R, $\forall$ t1,t2 tuples de r, on
  $\Pi_X(t1) = \Pi_X(t2) \Rightarrow \Pi_Y(t1) = \Pi_Y(t2)$

- A chaque valeur de X dans R correspond exactement 1 seule valeur de Y dans R

- [Notation]

X détermine Y : $X \rightarrow Y$

X,Y ou XY correspond à $X \cup Y$

Redondances et Normalisation

Soit le schéma suivant:

Emprunt_Amis (<u>N°DVD</u>, <u>Nom</u>, <u>Prénom</u>, Contact, Date_emp)

	A	B	C	D	E
1	N°DVD	Nom	Prénom	Contact	date_Emp
2	120	Jean	Djirangar	66324567	12/09/2021
3	230	Abakar	Mahamat	66887654	13/09/2022
4	201	Djimadoum	Mbiossoum	98765432	14/09/2021
5	145	Fatimé	Mahamat	90765432	15/09/2019

Pourquoi cette relation a t-elle des anomalies?

Quelles sont ces anomalies?

	A	B	C	D	E
1	N°DVD	Nom	Prénom	Contact	date_Emp
2	120	Jean	Djirangar	66324567	12/09/2021
3	230	Abakar	Mahamat	66887654	13/09/2022
4	201	Djimadoum	Mbiossoum	98765432	14/09/2021
5	145	Fatimé	Mahamat	90765432	15/09/2019

- **Anomalie de mise à jour**– Abakar Mahamat a changé de numéro de portable– mise à jour dans tous les tuples concernés
- **Anomalie d'insertion**– Nouvel ami : Jean Djirangar– je ne peux l'entrer dans la base que lorsqu'il m'empruntera un DVD
- **Anomalie de suppression** – J'ai perdu le DVD n°230 – si je le supprime de la base, je perds les informations sur Abakar Mahamat

Notion de Dépendance Fonctionnelle

Notion:

- Le salaire dépend de la qualification
- Le montant de la vignette automobile dépend du département où l'on réside
- Le numéro NCI d'une personne correspond exactement à son nom, son prénom, à son âge …

Notation

- N°DVD → Titre
- N°DVD → Réalisateur
- Titre, Réalisateur → Genre

Propriétés des DF

Règles d'Amstrong : X, Y et Z des groupes d'attributs d'une relation R

- [Réflexivité] $Y \subseteq X$ alors $X \rightarrow Y$
- [Augmentation] Si $X \rightarrow Y$ alors $X, Z \rightarrow Y, Z$
- [Transitivité] Si $X \rightarrow Y$ et $Y \rightarrow Z$ alors $X \rightarrow Z$
-

Propriétés des DF

•

Exemple : propriétés des DF

- [Réflexivité]: (N°DVD, Titre) → Titre
- [Augmentation]: DVD → Titre alors (N°DVD, Genre) → (Titre, Genre)
- [Transitivité]: N°DVD → Titre et Titre → Genre alors N°DVD → Genre

Exercice: Soit R(A,B,C,D,E) avec comme DF
{A → B et D → E}. Montrer que (A,D) → (B,E).

Propriétés dérivées

X,Y,Z,W 4 groupes d'attributs d'une relation R

- **[Union]** Si X → Y et X → Z alors X → (Y,Z)
  N°DVD → Titre et N°DVD → Réalisateur alors N°DVD → (Titre, Réalisateur)
- **[Décomposition]** Si X → (Y,Z) alors X → Y et X → Z
  N°DVD → (Titre, Réalisateur) alors N°DVD → Titre et N°DVD → Réalisateur
- **[Pseudo-transitivité]** Si X → Y et W,Y → Z alors W,X → Z
  N°DVD → Titre et (Titre, Réalisateur) → Genre alors (N°DVD, Titre) → Genre

Exercices

Exercice

1. Démontrer que les règles précédentes peuvent se déduire à l'aide des règles d'Amstrong

2. Soit R(A,B,C,D,E,F) avec comme dépendances fonctionnelles {A → BC, B → E, CD → EF}

Montrer que AD → E peut être déduite à l'aide des propriétés.

DF élémentaire

Une Dépendance Fonctionnelle Elémentaire (DFE) notée :

- $X \rightarrow A$

est une DF où :

- A : attribut unique tel que $A \not\subset X$
- $\nexists X' \subset X \ , \ X' \rightarrow A$

Remarques sur les DFE :

- la cible (A) est un attribut unique
- la source (X) ne comporte pas d'attributs superflus
- transitivité : seule règle d'inférence qui s'applique aux DFE

Exemples de DFs élémentaires

- AB→C est élémentaire si ni A, ni B pris individuellement ne déterminent C.

  Ex: Nom, DateNaissance, LieuNaissance → Prénom est élémentaire.

- AB→A n'est pas élémentaire car A est incluse dans AB.

- AB→CB n'est pas élémentaire car CB n'est pas un attribut, mais un groupe d'attributs.

- N°SS→ Nom, Prénom n'est pas élémentaire.

DF élémentaires

- On peut toujours réécrire un ensemble de DF en un ensemble de DFE, en supprimant les DF triviales obtenues par réflexivité et en décomposant les DF à partie droite non atomique en plusieurs DFE.

- **Réécriture de DF en DFE**

  - On peut réécrire les DF non élémentaires de l'exemple précédent en les décomposant DFE :

    - AB→A n'est pas considérée car c'est une DF triviale obtenu par réflexivité.

    - AB→CB est décomposée en AB→C et AB→B, et AB→B n'est plus considérée car triviale.

    - N°SS→Nom, Prénom est décomposée en N°SS→Nom et N°SS→Prénom.

Fermeture d'un ensemble de DF

- **[Définition]** L'ensemble des DF, noté **S^+**, obtenues par application des règles d'Amstrong s'appelle la **fermeture de S**
    - DF initiales enrichies de toutes les DF obtenues par *transitivité et augmentation*
    - Le calcul est fastidieux.
- **[Définition]** Deux ensembles de DF S1 et S2 sont dits équivalents ssi ils ont même fermeture i.e. $S1^+ = S2^+$

On a vu qu'à partir de certaines DF, à l'aide des règles d'Amstrong, on peut en déduire d'autres.

Notion de Fermeture Transitive des Dfs

- **Fermeture transitive**

  On appelle fermeture transitive F+ d'un ensemble F de DFE, l'ensemble de toutes les DFE qui peuvent être composées par transitivité à partir des DFE de F.

- Soit l'ensemble F = {A→B, B→C, B→D, A→E}.

  La fermeture transitive de F est

  F+ = { A→B, B→C, B→D, A→E, A→C, A→D }

Fermeture d'un ensemble de DF

- Pour identifier les bonnes clés d'un schéma relationnel on utilise les **axiomes d'Armstrong** (règles d'inférences) qui permettront de déduire des dépendances fonctionnelles à partir de celles existantes.

- L'ensemble de ces règles d'inférences déterminera une **fermeture** (on ne peut plus trouver de nouvelles DF) des dépendances fonctionnelles.

- A partir de ces DF on ne retiendra que les DF élémentaires (DFE) auxquelles ne s'appliqueront que les axiomes de transitivité ou pseudo transitivité.

- Ce qui permettra de déterminer une **fermeture transitive** à partir de laquelle on pourra extraire une ou plusieurs **couvertures minimales** (les DFE qui suffiront pour retrouver toutes les informations).

Fermeture d'un attribut

- Savoir si une DF **X → Y** appartient à la **fermeture de S**, ensemble de DF de départ. On calcule la **fermeture de X**, noté **X⁺**, *i.e.* **l'ensemble des attributs qui dépendent de X.**

- **Algorithme de saturation**

```
Données: X groupe d'attributs, S ens. de DF
Sortie: X+ fermeture de X
Res := X
Tant que Res évolue
faire
  Pour chaque DF Y → Z de S
  Si Y est dans Res alors Res := Res ∪ Z
finfaire
```

Exemple

- Soit R(A,B,C,D,E,F) avec
S={A → BC, E → CF, B → E, CD → EF}

Calculer AB^+

On trouve AB^+ = {A,B,C,E,F}

[Corollaire] Pour un ensemble S de DF, une DF X → Y appartient à S^+ ssi Y est un sous ensemble de X^+.

Fermeture d'un ensemble d' attributs

- La fermeture d'un ensemble d'attributs U, notée U^+, représente l'ensemble des attributs d'un schéma de relations R qui peuvent être déduits de U à partir d'une famille de dépendances fonctionnelles F en appliquant les axiomes d'Armstrong.

- Algorithme de calcul de la fermeture d'un ensemble d'attributs U :

  1. initialiser l'ensemble d'attributs U^+ avec U
  2. trouver une DF qui a en attributs source (à gauche de la DF) des attributs inclus dans U^+
  3. ajouter dans U^+ les attributs placés en partie droite de la dépendance fonctionnelle
  4. répéter les étapes dexu étapes précédentes jusqu'à ce que U^+ n'évolue plus.

Exemple

- Soit R(A,B,C,D,E,F) avec
S={A → BC, E → CF, B → E, CD → EF}

Calculer AB^+

On trouve AB^+ = {A,B,C,E,F}

Fermeture d'un ensemble d' attributs

3. ajouter dans U^+ les attributs placés en partie droite de la dépendance fonctionnelle

Soit F={A→D, AB→E, BI→E, CD→I, E→C}

Calculer la fermeture sous F, de U=(A,E) :

- A→D⟹(A,E)+={A,E,D}
- E→C⟹(A,E)+={A,E,D,C}
- CD→I⟹(A,E)+={A,E,D,C,I}

Calculer la fermeture sous FF, de U=(B,E) :

- E→C⟹(B,E)+={B,E,C}

Couverture minimale

- C'est un sous-ensemble minimal de la Fermeture Transitive
  - un seul attribut à droite de la DF
  - aucune DF ne peut être supprimée
  - aucun attribut à gauche de la DF ne peut-être enlevé

- La couverture minimale est l'ensemble minimal des DFE :
  - un seul attribut à droite $A \rightarrow BC \Rightarrow A \rightarrow B, A \rightarrow C$
  - il n'y a pas de DFE redondantes $\forall (X \rightarrow A) \in \bar{F}, (\bar{F} - \{X \rightarrow A\}) \neq \bar{F}$
  - les parties gauches sont "dégrossies" : $\forall (X \rightarrow A) \in \bar{F}, \nexists Z \subset X, ((\bar{F} - \{X \rightarrow A\}) \cup \{Z \rightarrow A\}) \neq \bar{F}$

- Critères pour déterminer un ensemble minimal de DF à partir d'un ensemble F de DF :
  - **enlever les DF non-élémentaires :**
    - une DF X→A est DFE si A ne dépend pas d'un sous-ensemble de X
  - **enlever les DF redondantes :**
    - Une DF X→A de l'ensemble F des DF est redondante si elle est une conséquence F−{X→A}

Couverture minimale

- **Couverture minimale**
  - La couverture minimale d'un ensemble de DFE est un sous-ensemble minimum des DFE permettant de générer toutes les autres DFE.
  - **Synonymes** : Famille génératrice
- Tout ensemble de DFE (et donc tout ensemble de DF) admet au moins une couverture minimale (et en pratique souvent plusieurs).
- L'ensemble F = {A→B, A→C, B→C, C→B} admet les deux couvertures minimales :
  CM1 = {A→C, B→C, C→B} et
  CM2 = {A→B, B→C, C→B}

DF non redondante

- A partir d'un ensemble SE de DF élémentaires initial, l'ensemble de *toutes* les DF élémentaires sont obtenues uniquement par la règle de transitivité.
- Soit SE un ensemble de DF élémentaires. Une DF élémentaire est dite redondante si **elle se déduit des autres par la règle de transitivité**.
- Dans S ={A → B, A → C, B → C}, la DF A → C est redondante car elle peut être obtenue par transitivité des 2 autres.
- Deux méthodes pour éliminer les DF redondantes: méthode directe ou méthode suivante: La DF élémentaire X → A de l'ensemble de DF S est elle redondante? On calcule X+ sur S \ {X → A}. Si on obtient encore A alors X → A est redondante.

Exemple de couverture minimale

- Considérons l' ensemble F de DF :
  - $A \to B$
  - $(B,C) \to D$
  - $(A,C) \to (B,D,E)$
  - $D \to E$
- **Construire une couverture minimale**
- On réduit à droite des DF à un attribut (axiome de décomposition) :
  - $A \to B$
  - $(B,C) \to D$
  - $(A,C) \to B$
  - $(A,C) \to D$
  - $(A,C) \to E$
  - $D \to E$

La DF $(A,C) \to B$ n est pas élementaire du fait de la DF

- $A \to B$

Autrement dit il suffit de connaitre l attribut A pour retrouver B

$\Longrightarrow$ On peut donc éliminer la DF $(A,C) \to B$

La DF $(A,C) \to D$ est conséquence de $A \to B$

- $(A,C) \to (B,C)$ (axiome d'augmentation)
- $(A,C) \to D$ (transitivité avec $(B,C) \to D$)

$\Longrightarrow$ On peut donc éliminer la DF $(A,C) \to D$

De même $(A,C) \to E$ se déduira de

- $A \to B$
- $(B,C) \to D$
- $D \to E$

$\Longrightarrow$ On peut donc éliminer la DF $(A,C) \to E$

Au final on trouve donc une couverture minimale

- $A \to B$
- $(B,C) \to D$
- $D \to E$

constitué des DFE qui permettront de retrouver l ensemble des DF initiales

Exercice retrouver les DF initiales à partir de la couverture minimale

Construction de la couverture minimale

- Soit S un ensemble de DF initial. Pour construire la couverture minimale de S (irr(S)), on simplifie cet ensemble: $\forall$ s : X $\rightarrow$ Y de S,
- [R1] rendre le membre droit de s atomique
- [R2] rendre le membre gauche de s irréductible
- Puis [R3] éliminer les DF redondantes de l'ensemble obtenu par R1 et R2.

Pour 1. On applique la règle de décomposition

Pour 2. On ne considère que les membres gauche X de s ayant plus d'un attribut. On applique l'algorithme de saturation sur X\{un attribut composant X}. Si on retrouve le membre droit de s alors le membre gauche n'était pas irréductible.

- **Pour 3**. On applique la méthode décrite page précédente (éliminer les DF redondantes) ou par observation directe.

Graphe des DFs

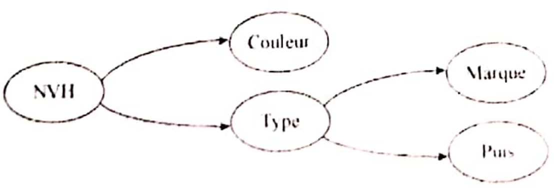
- On peut représenter un ensemble de DFE par un graphe orienté (ou plus précisément un réseau de Pétri), tel que les nœuds sont les attributs et les arcs les DFE (avec un seul attribut en destination de chaque arc et éventuellement plusieurs en source).

- **Relation Voiture**

Soit la relation Voiture(NVH, Marque, Type, Puis, Couleur) avec l'ensemble des DF

F = {NVH→Type, Type→Marque, Type→Puis, NVH→Couleur}.

On peut représenter F par le graphe ci-dessous :

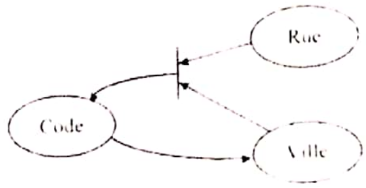


Graphe des DFE de la relation Voiture

Graphe des DFs

Relation CodePostal

- Soit la relation CodePostal (Code, Ville, Rue)
- avec l'ensemble des DF F={Code→Ville, (Ville,Rue)→Code}.
- On peut représenter F par le graphe ci-dessous :



Graphe des DFE de la relation CodePostal

• avec l'ensemble des DF F={Code→Ville, (Ville,Rue)→Code}.

Définition Formelle d'une clé

Clé

- Soient une relation R(A1,A2,...,An) et K un sous-ensemble de A1,A2,... ,An.
- K est une clé de R si et seulement si :
  1. K→ A1, A2, ..., An
  2. et il n'existe pas X inclus dans K tel que X→ A1, A2 ,..., An.
- Une clé est donc un ensemble minimum d'attributs d'une relation qui détermine tous les autres.

Normalisation des entités

- Les formes normales ont pour objectif de définir la décomposition des schémas relationnels, tout en préservant les DF et sans perdre d'informations, afin de représenter les objets et associations canoniques du monde réel de façon non redondante.

- On peut recenser les 6 formes normales suivantes, de moins en moins redondantes :
  - la première forme normale
  - la deuxième forme normale
  - la troisième forme normale
  - la forme normale de Boyce-Codd
  - la quatrième forme normale
  - la cinquième forme normale

- La troisième forme normale est généralement reconnue comme étant la plus importante à respecter.

Principe de la décomposition

- L'objectif de la décomposition est de "casser" une relation en relations plus petites afin d'en éliminer les redondances et sans perdre d'information.

importante à respecter.

- **Décomposition**
  - La décomposition d'un schéma de relation R(A1,A2,...,An) est le processus de remplacement de ce schéma par une collection de schémas R1,R2,...,Rn telle qu'il est possible de reconstruire R par des opérations relationnelles de jointure sur R1,R2,...,Rn.

- **Décomposition préservant les DF**
  - Une décomposition d'une relation R en relations R1,R2,...Rn préserve les DF si la fermeture transitive F+ des DF de R est la même que celle de l'union des fermetures transitives des DF de R1,R2,...,Rn.

La première forme normale(1FN)

- Une relation est en 1NF si elle possède au moins une clé et si tous ses attributs sont atomiques

- **Attribut atomique**
- Un attribut est atomique si il ne contient qu'une seule valeur pour un tuple donné, et donc s'il ne regroupe pas un ensemble de plusieurs valeurs.

- **Avoir plusieurs métiers**
- Soit la relation Personne instanciée par deux tuples :

Personne(#Nom, Profession)

(Dupont, Géomètre)
Durand, Ingénieur-Professeur)

- La relation n'est pas en 1NF, car l'attribut Profession peut contenir plusieurs valeurs.

La première forme normale(1FN)

Personne(#Nom, Profession)

(Dupont, Géomètre)
(Durand, Ingénieur-Professeur)

- Pour que la relation soit en 1NF, on pourrait par exemple ajouter Profession à la clé et faire apparaître deux tuples pour Durand, on obtiendrait :

Personne(#Nom, #Profession)

(Dupont, Géomètre)
(Durand, Ingénieur)
(Durand, Professeur)

- Une autre solution aurait été d'ajouter un attribut *ProfessionSecondaire*. On obtiendrait ainsi : *Personne(#Nom, Profession, ProfessionSecondaire)*

(Dupont, Géomètre, Null)
(Durand, Ingénieur, Professeur)

Normalisation des entités

- **deuxième forme normale:**

  Les propriétés d'une entité ne doivent dépendre que de l'identifiant de l'entité et non d'une partie de cet identifiant. En gros une relation est NF2 si elle est NF1 et les attributs non-clé ne doivent pas dépendre que d'une partie de la clé mais de sa totalité.

R(A,B,C,D,E) avec S ={AB →CDE, B →C }

R (<u>A</u> <u>B</u> C D E)

Soit *Employe*(*idEmp*, *idService*, *nom*, *salaire*, *nomService*) avec [*idEmp*, *idService*] étant la clé de la table et on sait que *idEmp* $\rightarrow$ *idService*, *nom*, *salaire*, *nomService* Est-ce que cette relation est NF2 ?

La deuxième forme normale (2FN)

- La deuxième forme normale permet d'éliminer les dépendances entre des parties de clé et des attributs n'appartenant pas à une clé.
- Une relation est en 2NF si elle est en 1NF et si tout attribut qui n'est pas dans une clé ne dépend pas d'une partie seulement d'une clé .C'est à dire encore que toutes les DF issues d'une clé sont élémentaires.

- Soit la relation Personne : Personne(#Nom, #Profession, Salaire)
- Soit les DF suivantes sur cette relation :

  Nom,Profession→Salaire et Profession→Salaire

- On note alors que la première DF est issue de la clé et qu'elle n'est pas élémentaire (puisque Profession détermine Salaire) et donc que le schéma n'est pas en 2NF.

La deuxième forme normale (2FN)

Personne(#Nom, #Profession, Salaire)

- Pour avoir un schéma relationnel en 2NF, il faut alors décomposer Personne en deux relations :

  Personne(#Nom, #Profession=>Profession, Prenom)
  Profession(#Profession, Salaire)

- On remarque que ce **schéma est en 2NF** (puisque Salaire dépend maintenant fonctionnellement d'une clé et non plus d'une partie de clé).

- On remarque aussi que la décomposition a préservé les DF, puisque nous avons à présent :
  - Profession→Salaire (DF de la relation Profession)
  - Nom,Profession→Profession (par Réflexivité)
  - Nom,Profession→Salaire (par Transitivité)

Troisième Forme Normale

- Une relation est NF3 si elle est NF2 et si les attributs non clés ne dépendent pas d'un attribut non clé

- Une relation est NF3 si tous les attributs non clé sont en DFE directe avec la clé. Il ne doit pas y avoir de dépendance fonctionnelle entre des attributs non clé

R(A,B,C,D,E) avec S ={AB →CDE, D →E }

R (<u>A</u> <u>B</u> C D E)

Soit $R(Nom_F, Adresse_F, Produit, Prix)$ et $DF = \{Nom_F \rightarrow Adresse_F ; Nom_F, Produit \rightarrow Prix\}$ La clé de R étant $[Nom_F, Produit]$. La relation est-elle NF3 ?

Troisième forme normale (3FN)

- La troisième forme normale permet d'éliminer les dépendances entre les attributs n'appartenant pas à une clé.

- Une relation est en 3NF si elle est en 2NF et si tout attribut n'appartenant pas à une clé ne dépend pas d'un autre attribut n'appartenant pas à une clé. C'est à dire encore que toutes les DFE C'est à dire encore que toutes les DFE vers des attributs n'appartenant pas à une clé, sont issues d'une clé

- **Clé candidate** La définition concerne toutes les clés candidates et non uniquement la clé primaire

Troisième forme normale (3FN)

- Soit la relation Profession : **Profession(#Profession, Salaire, Prime)**

- Soit les DF suivantes sur cette relation :
  - Profession→Salaire
  - Profession→Prime
  - Salaire→Prime

- Cette relation **n'est pas en 3NF** car **Salaire, qui n'est pas une clé, détermine Prime**.

- Pour avoir un schéma relationnel en 3NF, il faut décomposer Profession :

  Profession(#Profession, Salaire=>Salaire)
  Salaire(#Salaire, Prime)

- Ce schéma est en 3NF, car Prime est maintenant déterminé par une clé.

2FN et 3FN

- Une relation en 3NF est forcément en 2NF car :
  - Toutes les DFE vers des attributs n'appartenant pas à une clé sont issues d'une clé, ce qui implique qu'il n'existe pas de DFE, issues d'une partie de clé vers un attribut qui n'appartient pas à une clé.
  - Il ne peut pas non plus exister de DFE issues d'une partie de clé vers un attribut appartenant à une clé, par définition de ce qu'une clé est un ensemble minimum.
- On n'en conclut qu'il ne peut exister de DFE, donc a fortiori pas de DF, issues d'une partie d'une clé, et donc que toutes les DF issues d'une clé sont élémentaires.

Théorème de Heat

Soit la relation $R\{A, B, C\}$ dans laquelle A, B et C sont des sous-ensembles d'attributs de R. Si R satisfait la DF $A \rightarrow B$, alors R est égale à la jointure de ses projections sur $\{A, B\}$ et $\{A, C\}$.

Décomposition

- Soit F un ensemble de DF définies sur l'ensemble des attributs de la relation
- Déterminer la couverture minimale MIN(F) de F
- Construire la relation universelle R, relation composée de tous les attributs
- Déterminer la clé primaire de R à partir de MIN(F)
- Pour chaque DF, tant que la DF ne contient pas tous les attributs de la relation, décomposer la relation en deux nouvelles relations en utilisant le théorème de Heat :
  - Si la relation R(A,B,C) avec la DF B→C n'est pas en NF3, elle sera décomposée en R1(A,B) et R2(B,C)
- Appliquer ce processus de décomposition sur les relations jusqu'à l'obtention de relations en NF3. La décomposition peut être représenté sous forme d'arbre, les feuilles de l'arbre constituent les relations de la base.

- Soit F un ensemble de DF définies sur l'ensemble des attributs de la relation

Exemple de Décomposition

- **Décomposer la relation pour qu'elle soit en NF3**

  Relation universelle R : R(A,B,C,D,E,F,G,H,I,J,K)

  Couverture minimale :F={AB→CE; A→F; F→DGH; D→IJK}

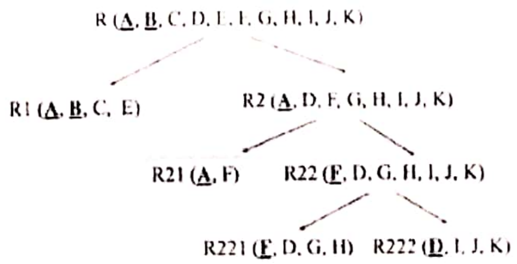
- **Solution:**

  A,B→C,E;

  A→F;

  F→D,G,H;

  D→I,J,K



[Th] Toute relation admet une décomposition en 3NF sans perte d'information et avec preservation des dépendances fonctionnelles.

Forme Normale de Boyce-Codd

- Pour les identifiants composés de plusieurs propriétés, ces dernières ne doivent pas être dépendantes d'une autre propriété de l'entité.
- Une relation est en BCNF si elle est en 3NF et si les seules DFE existantes sont celles pour lesquelles une clé candidate détermine un attribut.
- Soit la relation Personne : `Personne(#N°SS, #Pays, Nom, Région)`
- Soit les DF suivantes sur cette relation : N°SS, Pays→Nom; N°SS,Pays→Région; Région→Pays
- **Il existe une DFE qui n'est pas issue d'une clé et qui détermine un attribut appartenant à une clé**. Cette relation est en 3NF, **mais pas en BCNF** (car en BCNF toutes les DFE sont issues d'une clé).

Forme Normale de Boyce-Codd

- Soit la relation Personne :

```
Personne(#N°SS, #Pays, Nom, Région)
```

- Soit les DF suivantes sur cette relation : N°SS,Pays→Nom; N°SS,Pays→Région; Région→Pays
- Il existe une DFE qui n'est pas issue d'une clé et qui détermine un attribut appartenant à une clé. Cette relation est en 3NF, mais pas en BCNF (car en BCNF toutes les DFE sont issues d'une clé).
- Pour avoir un schéma relationnel en BCNF, il faut décomposer Personne :

```
Personne(#N°SS, #Region=>Region, Nom)
Region(#Region, Pays)
```

- La BCNF est la forme normale la plus facile à appréhender intuitivement et formellement, puisque les seules DFE existantes sont de la forme K→A où K est une clé.
- **Algorithme de décomposition en Forme Normale de Boyce-Codd**
  - Entrée R(a_1, a_2, ..., a_n) et F un ensemble minimal de DF

  1. Si R n'est pas en FNBC, alors il existe X→A tel que X n'est pas une surclé de R
  2. Remplacer R par (X,A) et R\{A} et recommencer

Forme Normale de Boyce-Codd

- **Synthèse:** Une relation est en forme normale de BOYCE-CODD (BCNF) si, et seulement si : elle est en troisième forme normale et les seules dépendances fonctionnelles élémentaires sont celles dans lesquelles une clé détermine un attribut.

- Ce qu'on pourrait écrire schématiquement et intuitivement ainsi : La relation **(A1, A2, A3, A4)** est en **BCNF s'il n'existe pas A3 - > A1** Ce cas est beaucoup plus rare.

- **La solution est la suivante : (A3, A1) (#A3, A2) (A1, A2, A4) si c'est le cas.**

Remarque On a ajouté la table (A3, A2), ce qui n'était pas évident. Cette table est nécessaire pour maintenir la dépendance initiale : A1 + A2 -> A3. En effet, cette dépendance (et donc la table de départ) pourra être reconstruite par la jointure entre (A3, A2) et (A1, A2, A4).

Forme Normale de Boyce-Codd

Exemple

- Des profs (P), des élèves (E), des matières (M), des salles (S).
- Un élève n'a qu'un prof par matière.
- Un élève ne suit une matière qu'une seule fois.
- Un prof n'enseigne qu'une matière.
- Une matière peut être enseignée par plusieurs professeurs.
- Un élève pour une matière donnée est dans une salle donnée.
- Dans une salle on peut avoir plusieurs enseignements. (E, M, P, S)
- Mais on a aussi prof -> matière

- Solution : (P, M) (E, M, S) (#P, E)

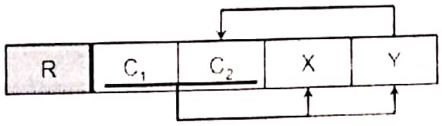
Forme Normale de Boyce-Codd

- Entrée : $R(a_1, a_2, \ldots, a_n)$ et F un ensemble minimal de DF

1. Si R n'est pas en FNBC, alors il existe $X \rightarrow A$ tel que X n'est pas une surclé de R
2. Remplacer R par (X,A) et R\{A} et recommencer.

- Forme Normale de Boyce Codd

  une relation est en BCNF (Boyce-Codd Normal Form) ssi
  - les seules dépendances fonctionnelles sont du type clé détermine attribut non clé



Une telle relation peut être décomposée en

$R_1(C_1, Y, X)$ et $R_2(Y, C_2)$

Exemple : **R (VILLE, DEPARTEMENT, CODE POSTAL) n'est pas en BCNF**

R peut être décomposée en : R1 (VILLE, CODE POSTAL) R2 (DEPARTEMENT, CODE POSTAL)
= Décomposition sans perte Mais qui ne préserve pas la DF : VILLE, DEPARTEMENT -> CODE. **R1 et R2 sont en BCNF.**

Forme Normale de Boyce-Codd

Une relation en BCNF peut comporter des redondances:

"L'étudiant de numéro NUMETUD pratique le sport SPORT et suit le COURS ".

– Pas de DF
– CLE = {NUMETUD, SPORT, COURS}
– R est en 3FN et en BCNF
– Cependant R contient des redondances

R	NUMETUD	SPORT	COURS
	100	Foot	Math
	100	Foot	Anglais
	200	Foot	Math
	200	Tennis	Anglais
	200	Foot	Anglais
	200	Tennis	Math

Dépendance multivaluée (DMV)

Restaurant	Pizza Variety	Delivery Area
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City

- Un restaurant délivre toutes ses variétés de pizzas dans toutes ses zones de livraison
- L'indépendance entre les variétés de pizzas offertes par un restaurant et les zones de livraison de ce restaurant entraîne des redondances dans la table

Dépendance multivaluée (DMV)

- Soit une relation R (A1, A2, ..., An)
  - X sous-ensemble de {A1, A2, ..., An}
  - Y sous-ensemble de {A1, A2, ..., An}

- X ->> Y : "**X multidétermine Y** "

  si, soit Z = R - X - Y,

  {(xyz) et (xy'z') ∈ R => (xy'z) et (xyz') ∈ R }

"A chaque valeur de X, il y a un ensemble de valeurs de Y associées et cet ensemble est indépendant des autres attributs.«

Dépendance multivaluée: Exemples

- **Exemple:** Un étudiant peut pratiquer plusieurs sports et un étudiant peut faire plusieurs cours.

  R (NUMETUD, SPORT, COURS)

  NUMETUD->>SPORT

  NUMETUD->>COURS

  <=> Pas de lien entre les cours suivis et les sports pratiqués.

- ***Contre-exemple :*** R (NUMEMP, LANGUE, PRODUIT)
  - "L'employé NUMEMP parle LANGUE et vend PRODUIT".
  - *Si pour vendre un produit, il faut parler la langue du pays où il est distribué, il n'y a pas de dépendance multivaluée.*

Dépendance multivaluée: Exemple

- Soit **(xyz)** = R25 Rouge Break
- Et **(xy'z')** = R25 Vert Normal
- Tous les deux sont des tuples de la relation affichée.

TYPE_VOITURE	COULEUR	Modele
R25	ROUGE	Break
R25	VERT	Normal
R25	ROUGE	Normal
R25	VERT	Break
MEGANE	VERT	Normal
MEGANE	BLEU	Break
MEGANE	VERT	Break
MEGANE	BLEU	Normal

- Alors **(xyz')** = R25 Rouge Normal et
- **(xy'z)** = R25 Vert Break sont aussi des tuples de R.
- Il faut juste une permutation entre z et z'.

- Donc **Type_voiture ->> Couleur**

- Et (xy'z') = R25 Vert Normal

R25	VERT	Normal

Raisonnement sur les Dépendance multivaluée (DMV)

1. DF :Si $X \rightarrow Y$ alors $X \rightarrow\rightarrow Y$

2. Complémentation :

   Si $X \rightarrow\rightarrow Y$ alors $X \rightarrow\rightarrow R \setminus \{X,Y\}$

3. Augmentation :

   Si $X \rightarrow\rightarrow Y$ et $W \supseteq Z$ alors $WX \rightarrow\rightarrow YZ$

4. Transitivité :

   Si $X \rightarrow\rightarrow Y$ et $Y \rightarrow\rightarrow Z$ alors $X \rightarrow\rightarrow (Z-Y)$

3. Augmentation :

Dépendance Multivaluée Elémentaire (DME)

C'est une Dépendance multivaluée X ->> Y d'une relation R telle que :

– 1. Y n'est pas vide et est disjoint de X

– 2. R ne contient pas une autre DM du type X' ->> Y'

telle que $X' \subset X$ et $Y' \subset Y$

Quatrième Forme Normale

- **une relation est en 4FN si les seules DM sont celles dans lesquelles une clé multi détermine un attribut.**

- ***Remarques :***
  - une dépendance fonctionnelle est un cas particulier de dépendance multivaluée
  - 4FN => 3FN et BCNF
  - en fait, on ne considère que les DM élémentaires (parties gauche et droite minimale).

Les Dfs sont des cas particuliers des DMs

- Une dépendance fonctionnelle est un cas particulier de dépendance multivaluée
- En effet:
  - (X → Y) => {(x,y,z) et (x,y',z') ∊ R => y=y'}
  - d'où : (X → Y) => {(x,y,z) et (x,y',z') ∊ R

    => (x,y',z) et (x,y,z') ∊ R}
  - Donc : (X → Y) => (X ->> Y)

Quatrième Forme Normale

- Une relation est en quatrième forme normale **si et seulement si les dépendances multivaluées**

  $X \rightarrow\rightarrow Y$ vérifient :

  ✓ $Y \subseteq X$ ou $XY = R$

  ✓ X est une surclé.

- Exemples :

Voiture (Type, Couleur, mod)

V1(Type, couleur)

V2(Type, Mod)

Algorithme de décomposition

- Utiliser le même algo que BCNF

1. Si R n'est pas en 4^e FN, alors il existe $X \rightarrow\rightarrow A$ tel que X n'est pas une surclé de R
2. Remplacer R par (X,A) et R\{A} et recommencer

✓ $Y \subset X$ ou $XY = R$

$X \rightarrow\rightarrow A$ tel que X n'est pas une surclé de

Algorithme de décomposition en 4FN

On utilise cet algorithme lorsque certaines dépendances sont des DMs

<u>Entrée</u> : Schéma R(X) contenant des DF et des DM

<u>Sortie</u> : Schéma {R1, R2, ... , Rn} avec Ri en 4NF pour tout i

<u>Etape 1</u> : (initialisation)

S = { R }.

<u>Etape 2</u> : (itération)

Si T est un schéma de S qui n'est pas en 4NF, alors chercher une DM non triviale W ->-> V de T telle que W ne contienne pas une clé de T. Remplacer le schéma T dans S par les deux schémas T1(W, V) et T2(X - V) munis des dépendances dérivées de la fermeture de celles du schéma T.

Répéter l'étape 2 si S contient un schéma qui n'est pas en 4NF

<u>Etape 3</u> : (élimination de la redondance)

Pour tout couple Ri et Rj de S, si $X_i \subset X_j$ alors éliminer Ri de S.

Entrée : Schéma R(X) contenant des DF et des DM

Revoyons la relation concernant les PIZZAS

Restaurant	Pizza Variety	Delivery Area
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City

- la relation **PIZZAS (RESTAURANT, TYPE_PIZZA, ZONE_LIVRAISON)** n'est pas en 4NF
- RESTAURANT ->> TYPE_PIZZA
- RESTAURANT ->> ZONE_LIVRAISON

Exemple de décomposition en 4FN

Restaurant	Pizza Variety
Vincenzo's Pizza	Thick Crust
Vincenzo's Pizza	Thin Crust
Elite Pizza	Thin Crust
Elite Pizza	Stuffed Crust
A1 Pizza	Thick Crust
A1 Pizza	Stuffed Crust

Restaurant	Delivery Area
Vincenzo's Pizza	Springfield
Vincenzo's Pizza	Shelbyville
Elite Pizza	Capital City
A1 Pizza	Springfield
A1 Pizza	Shelbyville
A1 Pizza	Capital City

Restaurant	Pizza Variety	Delivery Area
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City

Un exemple pour introduire!

- **La 4NF n'élimine pas toutes les redondances**

- **Exemple**: la relation **BUVEURJUS** (BUVEUR, JUS, PRODUCTEUR)

BUVEUR	JUS	PRODUCTEUR
ALI	COCA	ABAKAR
ALI	COCA	DJIMADOUM
ALI	FANTA	DJIMADOUM
JEAN	COCA	DJIMADOUM

- Cette **Relation est en 4NF** : il n'existe pas de DM :
  - BUVEUR ->> JUS **faux** : manque (ALI, FANTA, ABAKAR)
  - JUS ->> PRODUCTEUR **faux** : manque (JEAN, coca, ABAKAR)
  - PRODUCTEUR ->> BUVEUR **faux** : manque (JEAN, FANTA, DJIMADOUM)

Dépendances de Jointure (DJ)

- On considère les projections R1, R2, R3 de la relation **BUVEURJUS** sur deux attributs :

R1

BUVEUR	JUS
ALI	COCA
ALI	FANTA
JEAN	COCA

R2

BUVEUR	PRODUCTEUR
ALI	ABAKAR
ALI	DJIMADOUM
JEAN	DJIMADOUM

R3

JUS	PRODUCTEUR
COCA	ABAKAR
COCA	DJIMADOUM
FANTA	DJIMADOUM

- On n'obtient pas la relation **BUVEURJUS** par jointure de ces relations deux à deux :
  - BUVEURJUS ≠ R1 |X| R2 : **(Ali, Fanta, Abakar) en trop**
  - BUVEURJUS ≠ R1 |X| R3 : **(Jean, coca, Abakar) en trop**
  - BUVEURJUS ≠ R2 |X| R3 : **(Jean, Fanta, Djimadoum) en trop**

Dépendances de Jointure (DJ)

- La relation BUVEURJUS présente des redondances :
  - Il apparaît deux fois que ALI boit du COCA et que DJIMADOUM produit du COCA
- Mais elle n'est pas décomposable en deux relations
- Il existe des relations non décomposables en deux relations, mais décomposables en N relations (N >= 3)
- **Exemple** : si la relation BUVEURJUS obéit à la contrainte :
  - Tout buveur ayant bu un Jus et ayant commandé à un producteur produisant ce jus a aussi commandé ce jus à ce producteur :
  - (b,j) ∈R1 et (b,p) ∈ R2 et (c,p) ∈ R3 => (b,j,p) ∈ R
  - Dans ce cas R sera la jointure de R1, R2 et R3

relations (N >= 3)

Dépendance de Jointure (DJ)

- *Dépendance de jointure :*

  R (A1, A2, …An)
  X1, X2, …, Xm sous-ensembles de {A1, A2, …, An}

  Il existe une dépendance de jointure *{X1, X2, …, Xm}

  si $R = \Pi_{X1}(R) \bowtie \Pi_{X2}(R) \bowtie \dots \bowtie \Pi_{Xm}(R)$

- **S'il existe une telle dépendance de jointure, alors R est décomposable en m relations X1, X2, …, Xm.**

Dépendances de Jointure

Exemple:

La relation BUVEURJUS obéit à la dépendance de jointure :

- *{(BUVEUR, JUS), (BUVEUR, PRODUCTEUR),(JUS, PRODUCTEUR)}
- Elle est donc décomposable en ces trois relations, car :
- Si (b,j), (b,p) et (j,p) appartiennent respectivement à R1, R2 et R3 alors (b,j,p) appartient à BUVEURJUS

- **Les DM sont des cas particuliers de DJ** :
  - Une relation R(X,Y,Z) vérifiant la DM : X ->> Y (et donc X ->> Z) satisfait la DJ *{(XY),(XZ)}

Cinquième Forme Normale

- La cinquième forme normale concerne les tables dont la clé est constituée d'au moins trois attributs et qui n'ont pas d'autres attributs que ceux de la clé primaire.
- Une table est en cinquième forme normale si aucun attribut non clé ne dépend fonctionnellement d'une partie de cette clé (tout attribut dépend de toute la clé).
- Ce qu'on pourrait écrire schématiquement et intuitivement ainsi : La relation (A1, A2, A3) est 5FN si il n'existe pas A1 ->-> A2, A3
- La correction consiste à créer deux tables : (A1, A2) (A1, A3)

Cinquième Forme Normale

Exemple Des clients, des courriers, plusieurs adresses par client, plusieurs courriers envoyés par client.

- on peut concevoir la table des **courriers** envoyés à toutes les **adresses** des **clients** : (CL, AD, CO)
- Cette table n'est pas en 5ème forme normale car: CL ->-> AD, CO
- La solution est la suivante : (CL, CO) (CL, AD)

Exercices

Soit S le schéma d'une base de données relationnelle suivant comportant la seule relation R, sur lequel on a défini une famille F de dépendances fonctionnelles.

S={R(A,B,C,D,E)}

F={DE → C, C → B, B → E }

1. F est- elle minimale? Justifiez votre réponse.
2. Quelles sont les clés minimales de R? montrez comment vous les obtenez
3. Donner 2 exemples de redondances susceptibles d apparaître dans une instance de R?
4. En quelle forme normale est S?
5. La décomposition de S en S'={R1(C,D,E) R2(A,B,D,E)} est elle sans perte d'information? Si oui prouvez-le si non montrez un contre exemple
6. Quelles sont les dépendances projetées sur R1 et sur R2? La décomposition de S en S' est- elle sans perte de dépendance?
7. Proposez une décomposition de S, sans perte d'information, sans perte de dépendance et en 3ème forme normale Le schéma ainsi obtenu est-il en forme normale de Boyce-Codd?

- Décomposer en 4e FN les schémas suivants
  1. R(A,B,C,D) avec A→→B et A→→C
  2. R(A,B,C,D) avec A→→B et B→→CD
  3. R(A,B,C,D) avec AB→→C et B→D
  4. R(A,B,C,D,E) avec A→→B AB→→C et A→D, AB→E

1. F={AB→C, B→D CD→E, CE→GH G→A}
   - A-t-on AB→E BG→C AB→G
2. Les deux ensembles de DF F et G sont-ils équivalents?
   - F={A→B, CE→H C→E, A→CH}
   - G={C→EH A→BC}

Soit la relation R(A,B,C,D,E,F,G,H) avec la liste de dépendances fonctionnelles suivantes

DF = {A→B, A→C, D→E, F→A, F→C, F→G, AD→C, AD→H}

1. Proposer une décomposition en 3e FN de R
2. FNBC?

- G={C→EH, A→BC}

Normalisation des relations

- Une association binaire du MCD/MOD sans propriétés dont une seule cardinalité est 0 ou 1 devient une relation du MLD.

- Une association du MCD/MOD porteuse de propriétés devient une entité du MLD.

- Une association qui a des cardinalités x,n x,n (ou x vaut 0 ou 1) devient une entité.

- Une association ternaire devient une entité. Cette entité (correspondant à une association) est reliée aux entités (correspondants à des individus)par des relations. Toutes les relations sont à priori de type 1-N mais une relation 1-1 reste possible.