

Bases de données

Normalisation

Xavier Tannier
xavier.tannier@sorbonne-universite.fr

Dépendance fonctionnelle

Dépendance fonctionnelle

- Si $X = (A_i, A_{i+1}, \dots)$ et $Y = (A_j, A_{j+1}, \dots)$ sont des ensembles d'attributs
- Il existe une dépendance fonctionnelle $X \rightarrow Y$ si et seulement si une valeur de X « détermine pleinement » la valeur de Y
- En d'autres termes :
 - $X \rightarrow Y$ si chaque valeur de X est associée avec exactement une valeur dans Y
 - Deux n-uplets partageant les mêmes valeurs de X auront les mêmes valeurs de Y
 - Si les valeurs des attributs de X sont connues, alors on peut connaître les valeurs des attributs de Y en regardant un n-uplet contenant ces valeurs de X .
 - Y est une fonction de X

Dépendance fonctionnelle

- Si $X = (A_i, A_{i+1}, \dots)$ et $Y = (A_j, A_{j+1}, \dots)$ sont des ensembles d'attributs
- Il existe une dépendance fonctionnelle $X \rightarrow Y$ si et seulement si une valeur de X « détermine pleinement » la valeur de Y

Série	Saison	Épisode	Titre	Créateur	Pays de diffusion	Chaîne	Date 1 ^{ère} diffusion
Weeds	1	5

(Série) \rightarrow (Créateur)

(Série, Pays de diffusion) \rightarrow (Chaîne)

(Série, Saison, Épisode) \rightarrow (Titre)

(Série, Saison, Épisode, Pays) \rightarrow (Date 1^{ère} diffusion)

Dépendance fonctionnelle

- Si $X = (A_i, A_{i+1}, \dots)$ et $Y = (A_j, A_{j+1}, \dots)$ sont des ensembles d'attributs
- Il existe une dépendance fonctionnelle $X \rightarrow Y$ si et seulement si une valeur de X « détermine pleinement » la valeur de Y

Série	Saison	Épisode	Titre	Créateur	Pays de diffusion	Chaîne	Date 1 ^{ère} diffusion
Weeds	1	5

~~(Créateur) \rightarrow (Série)~~ : non car un créateur peut créer plusieurs séries

~~(Série) \rightarrow (Pays de diffusion)~~ : non car une série peut être diffusée dans plusieurs pays

Dépendance fonctionnelle

- Dépendances fonctionnelles = **relations** entre les objets dans le monde réel
- Une approche complémentaire de la définition du diagramme de classes UML
- Une DF n'est pas « vraie » ou « fausse », elle est ou pas adaptée au problème
- Le schéma de la base doit être cohérent avec la définition des DF



Propriétés des DF

- Propriétés des dépendances fonctionnelles :

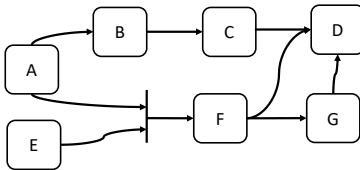
- Réflexivité : si B est un sous-ensemble de A, alors $A \rightarrow B$
- Augmentation : si $A \rightarrow B$ alors $\{A, C\} \rightarrow \{B, C\}$
- Transitivité : si $A \rightarrow B$ et $B \rightarrow C$ alors $A \rightarrow C$

Axiomes
d'Armstrong

- Décomposition : si $A \rightarrow \{B, C\}$ alors $A \rightarrow B$ et $A \rightarrow C$
- Union : si $A \rightarrow B$ et $A \rightarrow C$ alors $A \rightarrow \{B, C\}$
- Pseudo-transitivité : si $A \rightarrow B$ et $\{B, C\} \rightarrow D$ alors $\{A, C\} \rightarrow D$
- Composition : si $A \rightarrow B$ et $C \rightarrow D$ alors $\{A, C\} \rightarrow \{B, D\}$

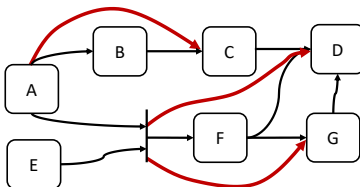
Dépendance fonctionnelle

- Les DF sont **transitives**
 - Si $X \rightarrow Y$ et $Y \rightarrow Z$ alors $X \rightarrow Z$
- Fermeture transitive** : déduction de toutes les DF implicites par application de la transitivité



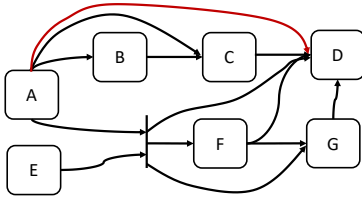
Dépendance fonctionnelle

- Les DF sont **transitives**
 - Si $X \rightarrow Y$ et $Y \rightarrow Z$ alors $X \rightarrow Z$
- Fermeture transitive** : déduction de toutes les DF implicites par application de la transitivité



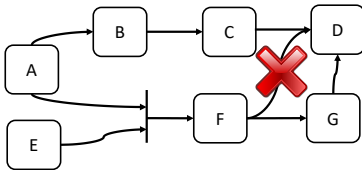
Dépendance fonctionnelle

- Les DF sont **transitives**
 - Si $X \rightarrow Y$ et $Y \rightarrow Z$ alors $X \rightarrow Z$
- Fermeture transitive** : déduction de toutes les DF implicites par application de la transitivité



Dépendance fonctionnelle

- Couverture minimale** : un sous-ensemble minimal de DF permettant de déduire toutes les autres par transitivité



Couverture minimale

- Pour trouver la couverture minimale (version courte)
 - On fait la fermeture transitive, puis
 - On décompose les DF (que des attributs uniques à droite)
 - On vérifie que les DF sont élémentaires, c'est-à-dire que la partie de droite dépend **entièrement** de la partie de gauche
 - On cherche à supprimer toutes les DF que l'on peut obtenir à partir des autres, en utilisant les propriétés des DF


Couverture minimale

- Exemple :

$\{A,B\} \rightarrow C$
 $C \rightarrow A$
 $\{B,C\} \rightarrow D$
 $\{A,C,D\} \rightarrow B$
 $D \rightarrow \{E,G\}$
 $\{B,E\} \rightarrow C$
 $\{C,G\} \rightarrow \{B,D\}$
 $\{C,E\} \rightarrow \{A,G\}$


Couverture minimale

1. Décomposition pour avoir un seul attribut à droite

$\{A,B\} \rightarrow C$ $C \rightarrow A$ $\{B,C\} \rightarrow D$ $\{A,C,D\} \rightarrow B$ $D \rightarrow \{E,G\}$ $\{B,E\} \rightarrow C$ $\{C,G\} \rightarrow \{B,D\}$ $\{C,E\} \rightarrow \{A,G\}$	<i>décomposition</i> 	$\{A,B\} \rightarrow C$ $C \rightarrow A$ $\{B,C\} \rightarrow D$ $\{A,C,D\} \rightarrow B$ $D \rightarrow E$ $D \rightarrow G$ $\{B,E\} \rightarrow C$ $\{C,G\} \rightarrow B$ $\{C,G\} \rightarrow D$ $\{C,E\} \rightarrow A$ $\{C,E\} \rightarrow G$
--	---	---

Couverture minimale

2. Suppression des attributs en surnombre à gauche

$\{A,B\} \rightarrow C$ $C \rightarrow A$ $\{B,C\} \rightarrow D$ $\{A,C,D\} \rightarrow B$ $D \rightarrow E$ $D \rightarrow G$ $\{B,E\} \rightarrow C$ $\{C,G\} \rightarrow B$ $\{C,G\} \rightarrow D$ $\{C,E\} \rightarrow A$ $\{C,E\} \rightarrow G$	<i>transitivité</i> 	$\{A,B\} \rightarrow C$ $C \rightarrow A$ $\{B,C\} \rightarrow D$ $\{C,D\} \rightarrow B$ $D \rightarrow E$ $D \rightarrow G$ $\{B,E\} \rightarrow C$ $\{C,G\} \rightarrow B$ $\{C,G\} \rightarrow D$ $\{C,E\} \rightarrow A$ $\{C,E\} \rightarrow G$
---	--	---

Couverture minimale

3. Suppression des DF redondantes

$\{A,B\} \rightarrow C$
 $C \rightarrow A$
 $\{B,C\} \rightarrow D$
 $\{C,D\} \rightarrow B$
 $D \rightarrow E$
 $D \rightarrow G$
 $\{B,E\} \rightarrow C$
 $\{C,G\} \rightarrow B$
 $\{C,G\} \rightarrow D$
 $\{C,E\} \rightarrow A$
 $\{C,E\} \rightarrow G$

Couverture minimale

3. Suppression des DF redondantes

$\{A,B\} \rightarrow C$
 $C \rightarrow A$
 $\{B,C\} \rightarrow D$
 ~~$\{C,D\} \rightarrow B$~~
 $D \rightarrow E$
 $D \rightarrow G$
 $\{B,E\} \rightarrow C$
 $\{C,G\} \rightarrow B$
 $\{C,G\} \rightarrow D$
 $\{C,E\} \rightarrow A$
 $\{C,E\} \rightarrow G$

$\{D,C,G\} \rightarrow \{D,B\}$
 augmentation
 $\{D,C\} \rightarrow \{D,B\}$
 transitivité
 $\{D,C\} \rightarrow B$
 décomposition

$\{A,B\} \rightarrow C$
 $C \rightarrow A$
 $\{B,C\} \rightarrow D$
 $D \rightarrow E$
 $D \rightarrow G$
 $\{B,E\} \rightarrow C$
 $\{C,G\} \rightarrow B$
 $\{C,G\} \rightarrow D$
 $\{C,E\} \rightarrow A$
 $\{C,E\} \rightarrow G$

Couverture minimale

3. Suppression des DF redondantes

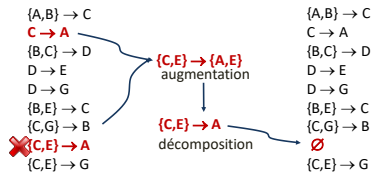
$\{A,B\} \rightarrow C$
 $C \rightarrow A$
 $\{B,C\} \rightarrow D$
 $D \rightarrow E$
 $D \rightarrow G$
 $\{B,E\} \rightarrow C$
 $\{C,G\} \rightarrow B$
 ~~$\{C,G\} \rightarrow D$~~
 $\{C,E\} \rightarrow A$
 $\{C,E\} \rightarrow G$

$\{C,G\} \rightarrow D$
 pseudo-transitivité

$\{A,B\} \rightarrow C$
 $C \rightarrow A$
 $\{B,C\} \rightarrow D$
 $D \rightarrow E$
 $D \rightarrow G$
 $\{B,E\} \rightarrow C$
 $\{C,G\} \rightarrow B$
 $\{C,E\} \rightarrow A$
 $\{C,E\} \rightarrow G$

Couverture minimale

3. Suppression des DF redondantes



Couverture minimale

4. Résultat :

$\{A, B\} \rightarrow C$
 $C \rightarrow A$
 $\{B, C\} \rightarrow D$
 $D \rightarrow E$
 $D \rightarrow G$
 $\{B, E\} \rightarrow C$
 $\{C, G\} \rightarrow B$
 $\{C, E\} \rightarrow G$

Décomposition

Anomalies de structure

Client	Adresse Client	TypeProduit Acheté	Référence Produit	Quantité	Date
J. Dupont	Paris	Livre	LYF526	1	2016-01-20
J. Dupont	Paris	Feuilles	FA459F	10	2016-02-12
P. Durand	Lyon	Téléphone	Asu153	1	2016-01-20
T. Martin	Marseille	Feuilles	FA459F	5	2016-01-27
J. Dupont	Paris	Livre	LDE901	1	2016-03-03
P. Durand	Lyon	Livre	LTY314	2	2016-01-20
L. Dubois	Marseille	Carnet	CDSF93	3	2016-03-03

Anomalies de structure

• Anomalies de mise à jour

- Mettre à jour l'adresse de J. Dupont ?

- Informations **redondantes**
- Risque d'**incohérences**

Client	Adresse Client	TypeProduit Acheté	Référence Produit	Quantité	Date
J. Dupont	Paris	Livre	LYF526	1	2016-01-20
J. Dupont	Paris	Feuilles	FA459F	10	2016-02-12
P. Durand	Lyon	Téléphone	Asu153	1	2016-01-20
T. Martin	Marseille	Feuilles	FA459F	5	2016-01-27
J. Dupont	Paris	Livre	LDE901	1	2016-03-03
P. Durand	Lyon	Livre	LTY314	2	2016-01-20
L. Dubois	Marseille	Carnet	CDSF93	3	2016-03-03

Anomalies de structure

• Anomalies d'insertion

- Insérer un nouveau client qui n'a encore rien acheté ?

- Informations **manquantes**
- Clés **incomplètes**

Client	Adresse Client	TypeProduit Acheté	Référence Produit	Quantité	Date
J. Dupont	Paris	Livre	LYF526	1	2016-01-20
J. Dupont	Paris	Feuilles	FA459F	10	2016-02-12
P. Durand	Lyon	Téléphone	Asu153	1	2016-01-20
T. Martin	Marseille	Feuilles	FA459F	5	2016-01-27
J. Dupont	Paris	Livre	LDE901	1	2016-03-03
P. Durand	Lyon	Livre	LTY314	2	2016-01-20
L. Dubois	Marseille	Carnet	CDSF93	3	2016-03-03
J. Thomas	Toulouse	NULL	NULL	NULL	NULL

Anomalies de structure

- **Anomalies de suppression**

- Suppression des données du 2016-01-20 ?
 - **Perte** d'informations (Client P. Durand)

Client	Adresse Client	TypeProduit Acheté	Référence Produit	Quantité	Date
J. Dupont	Paris	Livre	LYF526	1	2016-01-20
J. Dupont	Paris	Feuilles	FA459F	10	2016-02-12
P. Durand	Lyon	Téléphone	Asu153	1	2016-01-20
T. Martin	Marseille	Feuilles	FA459F	5	2016-01-27
J. Dupont	Paris	Livre	LDE901	1	2016-03-03
P. Durand	Lyon	Livre	LTY314	2	2016-01-20
L. Dubois	Marseille	Carnet	CDSF93	3	2016-03-03

Décomposition

- Si $R = (X, Y, Z)$ où X, Y, Z sont des ensembles d'attributs
- Si $X \rightarrow Y$ alors on peut décomposer R en $R_1(X, Y)$ et $R_2(X, Z)$

Client	Adresse Client	TypeProduit Acheté	Référence Produit	Quantité	Date
--------	----------------	--------------------	-------------------	----------	------

DF : Client \rightarrow AdresseClient

Décomposition

Client	Adresse Client
J. Dupont	Paris
P. Durand	Lyon
T. Martin	Marseille
L. Dubois	Marseille
L. Thomas	Toulouse

Client	TypeProduit Acheté	Référence Produit	Quantité	Date
J. Dupont	Livre	LYF526	1	2016-01-20
J. Dupont	Feuilles	FA459F	10	2016-02-12
P. Durand	Téléphone	Asu153	1	2016-01-20
T. Martin	Feuilles	FA459F	5	2016-01-27
J. Dupont	Livre	LDE901	1	2016-03-03
P. Durand	Livre	LTY314	2	2016-01-20
L. Dubois	Carnet	CDSF93	3	2016-03-03

Décomposition

→ Décomposition

Client	Adresse Client
J. Dupont	Paris
P. Durand	Lyon
T. Martin	Marseille
L. Dubois	Marseille
L. Thomas	Toulouse

Client	TypeProduit Acheté	Référence Produit	Quantité	Date
J. Dupont	Livre	LYF526	1	2016-01-20
J. Dupont	Feuilles	FA459F	10	2016-02-12
P. Durand	Téléphone	Asu153	1	2016-01-20
T. Martin	Feuilles	FA459F	5	2016-01-27
J. Dupont	Livre	LDE901	1	2016-03-03
P. Durand	Livre	LTY314	2	2016-01-20
L. Dubois	Carnet	CDSF93	3	2016-03-03

Décomposition

Série	Saison	Épisode	Titre	Créateur	Pays de diffusion	Chaîne	Date 1 ^{ère} diffusion
-------	--------	---------	-------	----------	-------------------	--------	---------------------------------

- (Série) → Créateur
- (Série, Pays de diffusion) → Chaîne
- (Série, Saison, Épisode) → Titre
- (Série, Saison, Épisode, Pays) → Date 1^{ère} diffusion



Décomposition

Série	Saison	Épisode	Titre	Créateur	Pays de diffusion	Chaîne	Date 1 ^{ère} diffusion
-------	--------	---------	-------	----------	-------------------	--------	---------------------------------

(Série) → Créateur

Décomposition

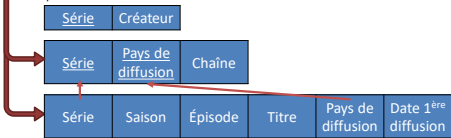


Décomposition

Série	Saison	Épisode	Titre	Créateur	Pays de diffusion	Chaîne	Date 1 ^{ère} diffusion
-------	--------	---------	-------	----------	-------------------	--------	---------------------------------

(Série, Pays de diffusion) → Chaîne

Décomposition

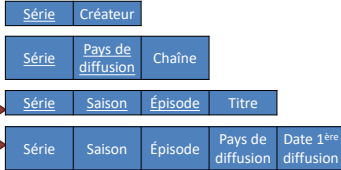


Décomposition

Série	Saison	Épisode	Titre	Créateur	Pays de diffusion	Chaîne	Date 1 ^{ère} diffusion
-------	--------	---------	-------	----------	-------------------	--------	---------------------------------

(Série, Saison, Épisode) → Titre

Décomposition

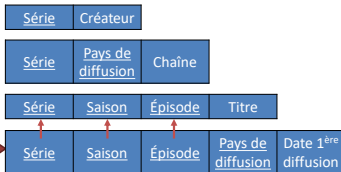


Décomposition

Série	Saison	Épisode	Titre	Créateur	Pays de diffusion	Chaîne	Date 1 ^{ère} diffusion
-------	--------	---------	-------	----------	-------------------	--------	---------------------------------

(Série, Saison, Épisode, Pays) → Date 1^{ère} diffusion

Décomposition



Décomposition

- La décomposition :
 - Préserve les informations
 - Préserve les dépendances fonctionnelles

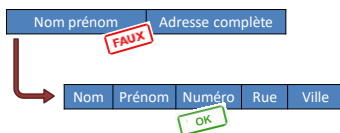
Formes normales

Normalisation

- Obtention d'une **forme normale** = transformation d'un ensemble de relations comportant des anomalies en un autre ensemble de relations :
 - Contenant les mêmes informations
 - Possédant les propriétés recherchées
- Un cadre permettant de :
 - Vérifier une structure (à l'issue du processus de modélisation)
 - Corriger une structure défailante
- Complémentaire de la modélisation UML puis relationnelle

Première forme normale : 1FN

- Une relation est en première forme normale (1FN) si :
 - Tous ses attributs sont atomiques



Deuxième forme normale : 2FN

- Une relation est en deuxième forme normale (2FN) si :
 - Elle est en 1FN
 - Aucun attribut non-clé n'est dépendant que d'une partie d'une clé

Fournisseur	Produit	Adresse Fournisseur	Prix
-------------	---------	---------------------	------

Fournisseur → AdresseFournisseur

(Fournisseur, Produit) → Prix

Décomposition

Fournisseur	Produit	Prix
Fournisseur	AdresseFournisseur	

FAUX

OK

Troisième forme normale : 3FN

- Une relation est en troisième forme normale (3FN) si :
 - Elle est en 2FN
 - Aucun attribut non-clé n'est dépendant d'un autre attribut non-clé

Immatriculation	Marque	Type	Puissance	Couleur
-----------------	--------	------	-----------	---------

Immatriculation → (Marque, Type)

Immatriculation → Couleur

(Marque, Type) → Puissance

Décomposition

Immatriculation	Marque	Type	Couleur
Marque	Type	Puissance	

FAUX

OK

Forme normale de Boyce-Codd: BCNF

- Une relation est en forme normale de Boyce-Codd si :
 - Elle est en 3FN
 - Aucun attribut clé n'est dépendant d'un attribut non-clé

A	B	C	D
---	---	---	---

(A, B) → C

(A, B) → D

C → A

Décomposition

A	B	D
C	A	

OK