



INITIATION AUX BASES DE DONNÉES

MODULE : Base de données et application
(Semestre 2)

Faculté Polydisciplinaire de Ouarghazat
<http://mural.uv.es/mone>

1

FPO: SIL

Mohamed Nemiche



Plan

- 1) Introduction
- 2) Le Modèle Relationnel
- 3) Normalisation et décomposition
- 4) Utilisation de relations
- 5) Algèbre relationnel
- 6) Conclusion

2

FPO: SIL

Mohamed Nemiche



1) INTRODUCTION

Constat

- Comment gérer (mémoriser et traiter) un ensemble d'étudiants (n°, nom, prénom, groupe, filière...)?

Première solution : utilisation des Fichiers

Mais comment traiter les requêtes (répondre aux questions) suivantes ?

- n° de Youssef Alami ?
- les étudiants du groupe 51 ?
- ...



1) INTRODUCTION

A chaque nouvelle question, de nouvelles procédures ;

Traitement dépendant de la structure des données stockées et du langage de programmation.



1) INTRODUCTION

Travailler directement sur un fichier présente plusieurs inconvénients :

- Manipulation de données lourde et compliquée. Il faut être expert en programmation
- Le programmeur doit connaître la localisation physique des fichiers, la structure physique des enregistrements, le mode d'accès à ces fichiers
- Toute modification de la structure des enregistrements (ajout d'un champ par exemple) entraîne la réécriture de tous les programmes qui manipulent ces fichiers

5

FPO: SIL

Mohamed Nemiche



Dupont
Symptomes : y
Turlututu : sqj
Symptomes : y
Turlututu : sdd
Analyses : xxx

Dupond
Turlututudjjsk
Symptom: yyy
Analyses xxx
Turlututudhjsd
Analyses xxx



Problèmes

- Difficultés de gestion
- Incohérence des données
- Coûts élevés
- Maintenance difficile
- Gestion de pannes ???
- Partage des données ???
- Confidentialité ???

L'APPROCHE SYSTEMES DE FICHIERS



Duhpon
Symptomes : yy
Analyses : xxx
Symptomes : yy

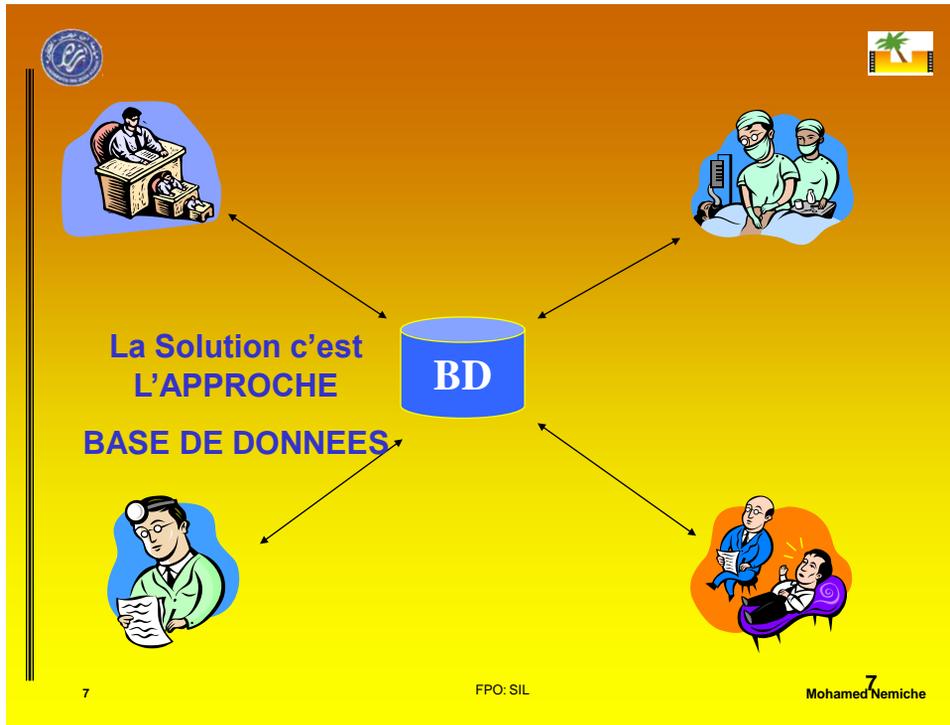
Duipont
Turlututu : sq
Symptomyyyy
Analysesxxxx
Turlututudhjsd



6

FPO: SIL

Mohamed Nemiche



1) INTRODUCTION

Solution: L'approche "Bases de données"

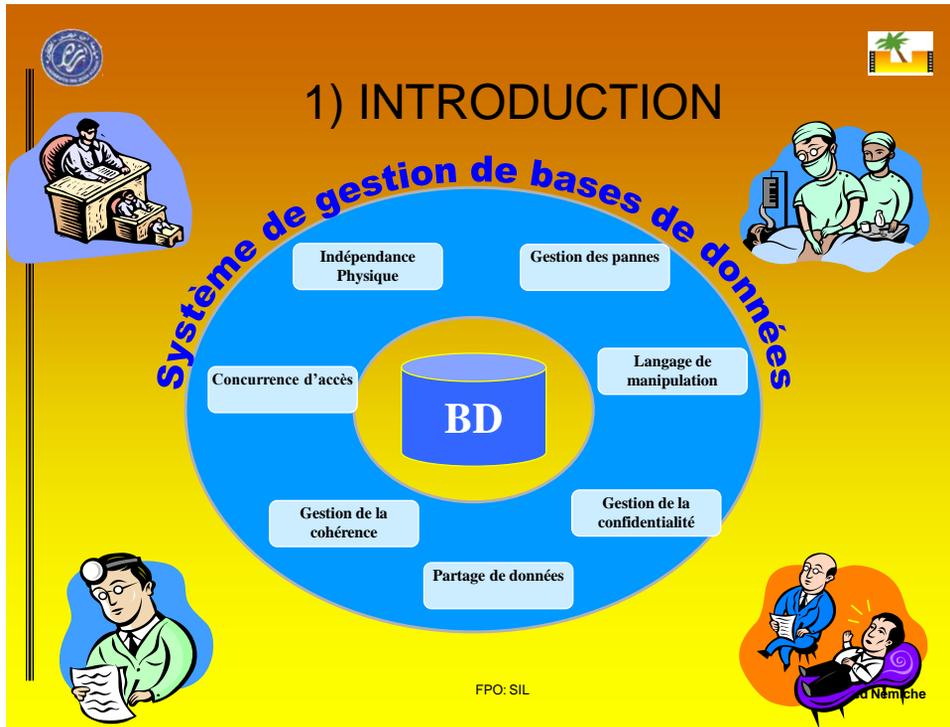
Définition intuitive : une base de données est un ensemble d'informations, (fichiers), partagé par plusieurs utilisateurs. Ces informations sont interrogées et mises à jour par l'intermédiaire d'un logiciel.

- **Modélisation des données**
 - Eliminer la **redondance** de données
 - **Centraliser** et **organiser** correctement les données
- **Logiciel «Système de Gestion de Bases de Données»**
 - **Factorisation** des modules de contrôle des applications
 - Interrogation, cohérence, partage, gestion de pannes, etc...
 - Administration facilitées des données

8

FPO: SIL

Mohamed Némiche



1) INTRODUCTION

Exemples d 'application

- Système de gestion de la clientèle d'une banque
- Catalogue électronique d'une bibliothèque

10

FPO: SIL

Mohamed Némiche



1) INTRODUCTION

Définition : une base de données est un ensemble structuré de données (1) enregistrées sur des supports accessibles par l'ordinateur (2) pour satisfaire simultanément plusieurs utilisateurs (3) de manière sélective (4) en un temps opportun(5).

- (1) : Organisation et description de données
- (2) : Stockage sur disque
- (3) : Partage des données
- (4) : Confidentialité
- (5) : Performance



1) INTRODUCTION

Modélisation et modèles

Définition (Modélisation): Expression du monde réel observé utilisant les concepts d'un *modèle* de représentation.

- Représentation de la partie du monde « utile » à l'application.
 - Par exemple, la modélisation d'une « personne » sera différente dans une application gérant l'inscription d'étudiants à des diplômes, et dans la BDD de la sécurité sociale.

Il existe plusieurs modèles de représentation.

- 1960 Modèle hiérarchique, modèle réseau.
- 1970 Modèle relationnel
- 1990 Modèle objet



Le Modèle Relationnel



HISTORIQUE SUR LE MODELE RELATIONNEL

Conçu en 1970 par E.F CODD (alors chercheur chez IBM-SAN JOSE-CALIFORNIE), Ce modèle qui se développe le plus actuellement pour des raisons de simplicités de représentation.



Basé sur la théorie des ensemble et les mathématiques relationnelles, il permet d'appliquer aux données tous les opérateurs ensemblistes et relationnels.

Le grand apport de ce modèle est la normalisation des données.

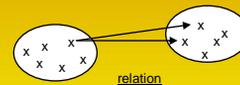
15

FPO: SIL

15
Mohamed Némiche

1) Introduction

- **Modèle de description des informations**
- **Modèle très utilisé, quoique ancien**
- **Fondement mathématique**
- **Modèle associé à l'algèbre relationnelle**



... qui repose sur 4 notions principalement

domaine
relation
Dépendance fonctionnelle
clé

16

FPO: SIL

Mohamed Némiche



2) Notions de base

2.1) Notions de domaine, d'attribut

Un attribut est une information qui contribue à la description de l'objet que l'on veut modéliser. Exemple nom, prénom, adresse ...



2) Notions de base

2.1) Notions de domaine, d'attribut

Ces attributs peuvent être :

- calculés à partir d'autres attributs,
- non calculables

+
élémentaires,
+ structurés.

Dans ce cas-là, en général, ils ne sont pas mémorisés.

On garde les attributs élémentaires.

Le modèle relationnel, en soi, ne fait pas la différence entre les 3 !!!

Dans ce cas-là, en général, ils ne sont pas mémorisés. On garde la formule de calcul et les attributs élémentaires.

$$\text{Taille_moyenne} = (\text{mini} + \text{maxi}) / 2$$



2) Notions de base

2.1) Notions de domaine, d'attribut

Un domaine : ensemble fini ou infini de valeurs distinctes que peut prendre un attribut (entiers, booléens...)

Un domaine peut être partagé par plusieurs attributs.



2) Notions de base

2.2) Notion de relation

Il y a deux façons de « voir » les choses :

- sous l'angle mathématique,
- sous l'angle pragmatique.

Une relation est un sous-ensemble du produit cartésien des domaines.

Une relation, c'est une table dans laquelle les colonnes correspondent aux attributs et les lignes aux occurrences.




2) Notions de base

2.2) Notion de relation

Il y a deux façons de « voir » les choses :

- sous l'angle mathématique,
- sous l'angle pragmatique.

Une **relation** est un sous-ensemble du produit cartésien des **domaines**.

Notions de base

Une relation, c'est une table dans laquelle les colonnes correspondent aux **attributs** et les lignes aux occurrences.

21 FPO: SIL Mohamed Nemiche




2) Notions de base

2.2) Notion de relation

Une relation, c'est une table dans laquelle les colonnes correspondent aux attributs et les lignes aux occurrences :

nom	prénom	adresse	filière

un étudiant pragmatique

un autre étudiant

NB1 : L'ordre des lignes et des colonnes est sans importance ;

Relation Etudiants

22 FPO: SIL Mohamed Nemiche



2) Notions de base

clé, identifiant

2.3) Notions de clé

Clé: attribut ou ensemble d'attributs dont les valeurs identifient de manière unique chaque tuple de la relation.

N°	NOM	PRENOM	ADRESSE	FILIERE
a12	Alami	Youssef	xx1	IGE
a13	toto	Mohamed	yy1	Tourisme
a14	titi	Ali	zz1	Cimena

- jamais de tuple en double !
- Il existe donc toujours un identifiant : dans le cas le pire, c'est l'ensemble des attributs de la relation !
- Il peut exister plusieurs identifiants

23

FPO: SIL

Mohamed Nemiche



Concepts du modèle relationnel

- Une **relation** est caractérisée par :
 - un nom R
 - un ensemble d'attributs A_1, A_2, \dots, A_n
- Si une relation a n attributs, n est son **arité**.

- Notation d'une relation

$$R (A_1, A_2, \dots, A_n)$$

ex. **Produit** (numProd, libellé, pu)

24

FPO: SIL

Mohamed Nemiche



Relations

- Un **attribut** est caractérisé par :
 - un nom A_i
 - un domaine noté $\text{dom}(A_i)$, ensemble des valeurs possibles de A_i
ex. : $\text{dom}(pu) =]0, 10000]$
- Valeur **nulle** (notée **NULL**) : valeur particulière indiquant que la valeur d'un attribut n'est pas connue ou que l'attribut ne s'applique pas.
 - ex1. Cas un client dont on ignore la ddn.
 - ex2. Cas d'un employé ne possédant pas de téléphone.



Relations

- Un **tuple** d'une relation $R(A_1, A_2, \dots, A_n)$ est un ensemble de valeurs $\langle v_1, v_2, \dots, v_n \rangle$ telles que
 - $v_i \in \text{dom}(A_i)$
 - ou
 - $v_i = \text{NULL}$

ex. $\langle 36, nokia, 500 \rangle$ est un tuple de la relation *Produit*.



EXEMPLE:



Soit la relation « Ouvrage »:

- Dans sa représentation tabulaire, elle se représente ainsi :

OUVRAGE:

Nom de la relation Ou table

Tuple ou Occurrence de la relation

N°Ouvrage	Titre	Auteur	Qte_stock
5412	Le rêve Mexicain	XXXX	120
7318	Le Parfum	XXXX	57
8417	De près de loin	XXXX	140
9003	Jour de silence à Tanger	XXXX	95
9075	L'avenir radieux	XXXX	110

Valeur de l'attribut « Qte_stock »

27
Mohamed Nemiche



- Dans sa représentation en extension, la relation « Ouvrage » se représente ainsi :

Ouvrage (N°Ouvrage ,Titre, Auteur, Qte_stock)

- Degré :

Correspond aux nombres d'attributs de la relation. (la relation « Ouvrage » est de degré 4)

- Cardinalité :

Elle représente le nombre de tuples ou d'occurrences ou de lignes de la relation. (la cardinalité de « Ouvrage » est 5)



Relations

L'**intention** d'une relation est définie par son **schéma** :

- nom de la relation
- liste des attributs + domaines
- contraintes d'intégrité

ex. Produit (numProd : nombre entier,
 libellé : chaîne de caractères,
 pu : nombre réel)

Deux contraintes : 1) clé primaire : numProd
 2) $0 < pu \leq 10\ 000$



Contraintes d'intégrité

- **Clé d'une relation** : Groupe d'attributs **minimum** qui identifie de manière **unique** un tuple dans une relation
 - Notion d'identifiant de type d'entité
- Toute relation doit avoir au moins une clé documentée, c'est la **clé primaire**.
 - ex. numProd : clé primaire de Produit.
- Notation : la clé primaire est **soulignée** dans le schéma.
 - ex. Produit (numProd, libellé, pu)



Contraintes d'intégrité

- **Clé étrangère d'une relation** : attribut(s) constituant la clé primaire d'une autre relation.
- Les clés étrangères définissent les CI référentielles
- Notation : la clé étrangère est en **italique** dans le schéma.



Contraintes d'intégrité

- **Contraintes liées au domaine** : les données doivent vérifier certaines conditions pour être cohérentes.

ex. $pu > 0$ ET $pu \leq 10000$



3) Dépendance fonctionnelle et normalisation

Un des buts de la normalisation est de passer d'une « grosse » relation à un ensemble de plus petites, plus « propres ».



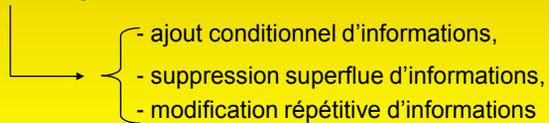
3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

Si l'on fait tout ce travail, c'est parce qu'on va y gagner !
Normaliser une relation (ou un ensemble de relations), c'est éviter des problèmes :

- de redondance,

- de stockage.



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

Considérons une extension de la relation Étudiants

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'entreprise
Belhaj	Amina	IGE	Informatique et gestion d'entreprise
Ouahbi	Fouad	IGE	Informatique et gestion d'entreprise

Dans ce tableau (bien entendu, nous l'avons choisi pour cela !), il y a 5 fois le fait que la filière informatique et gestion d'entreprise, le code de la filière IGE

redondance

35 FPO: SIL Mohamed Nemiche

3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

Relation Etudiants1

Nom	Prénom	Code Filière
Alami	Ali	IGE
Slaoui	Rachid	IGE
Nasri	Hind	IGE
Belhaj	Amina	IGE
Ouahbi	Fouad	IGE

Relation Filières

Code Filière	Filière
IGE	Informatique et gestion d'entreprise

Après normalisation, le système de tables n'aura plus que l'attribut Code Filière comme redondance. C'est ce que l'on pourrait appeler les coûts fixes. Le prix de la non-redondance, en quelque sorte.

Clé étrangère

~~redondance~~

36 FPO: SIL Mohamed Nemiche



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

ajout
conditionnel

Reprenons l'extension précédente de la relation
Étudiants

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'entreprise
Belhaj	Amina	IGE	Informatique et gestion d'entreprise
Ouahbi	Fouad	IGE	Informatique et gestion d'entreprise

Supposons que l'on vienne de connaître l'existence d'un autre étudiant, par exemple Ait Taleb Samir de la filière Techniques cinématographiques (TECC).

37

FPO: SIL

Mohamed Nemiche



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

ajout
conditionnel

Relation Etudiants1

Nom	Prénom	Code Filière
Alami	Ali	IGE
Slaoui	Rachid	IGE
Nasri	Hind	IGE
Belhaj	Amina	IGE
Ouahbi	Fouad	IGE

Relation Filières

Code Filière	Filière
IGE	Informatique et gestion d'entreprise
TECC	Techniques Cinématographiques

Après normalisation, le nouvelle filière sera enregistré avant même le premier étudiant ...

Etape n° 1

38

FPO: SIL

Mohamed Nemiche




3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

ajout conditionnel

Relation Etudiants1

Nom	Prénom	Code Filière
Alami	Ali	IGE
Slaoui	Rachid	IGE
Nasri	Hind	IGE
Belhaj	Amina	IGE
Ouahbi	Fouad	IGE
Ait Taleb	Samir	TECC

Relation Filières

Code Filière	Filière
IGE	Informatique et gestion d'entreprise
TECC	Techniques Cinématographiques

Etape n° 2

39 FPO: SIL Mohamed Nemiche




3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

suppression superflue

Reprenons notre extension favorite de la relation Étudiants ajoutons-y l'information concernant Ait Taleb.

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'entreprise
Belhaj	Amina	IGE	Informatique et gestion d'entreprise
Ouahbi	Fouad	IGE	Informatique et gestion d'entreprise
Ait Taleb	Samir	TECC	Techniques cinématographiques

Supposons maintenant que cet étudiant soit le seul de la filière TECC dans la base de données.

40 FPO: SIL Mohamed Nemiche




3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

suppression
superflue

Supprimer les informations concernant le Ait Taleb nous fait OBLIGATOIREMENT perdre celles sur la filière

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'entreprise
Belhaj	Amina	IGE	Informatique et gestion d'entreprise
Ouahbi	Fouad	IGE	Informatique et gestion d'entreprise
Ait Taleb	Samir	TECC	Techniques cinématographiques

suppression de l'étudiant → suppression de la filière !!

41 FPO: SIL Mohamed Nemiche




3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

suppression
superflue

Relation Etudiants1

Nom	Prénom	Code Filière
Alami	Ali	IGE
Slaoui	Rachid	IGE
Nasri	Hind	IGE
Belhaj	Amina	IGE
Ouahbi	Fouad	IGE
Ait Taleb	Samir	TECC

Relation Filières

Code Filière	Filière
IGE	Informatique et gestion d'entreprise
TECC	Techniques Cinématographiques

Après normalisation, on pourra conserver les références de la filière TECC tout en ayant enlevé celles de l'étudiant. On n'aura supprimé que ce qui était strictement nécessaire.

42 FPO: SIL Mohamed Nemiche



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
répétitive

Reprenons encore une fois (la dernière ???) l'extension de la relation Étudiants:

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'organisation
Slaoui	Rachid	IGE	Informatique et gestion d'organisation
Nasri	Hind	IGE	Informatique et gestion d'organisation
Belhaj	Amina	IGE	Informatique et gestion d'organisation
Ouahbi	Fouad	IGE	Informatique et gestion d'organisation
Ait Taleb	Samir	TECC	Techniques cinématographiques

et supposons que nous ayons mal écrit le libelle de la filière informatique

43

FPO: SIL

Mohamed Nemiche



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
répétitive

Rectifier cette erreur va nécessiter la modification de TOUTES les occurrences de la relation (de toutes les lignes de la table) correspondant à la filière informatique (ie. des milliers !!).

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'organisation
Nasri	Hind	IGE	Informatique et gestion d'organisation
Belhaj	Amina	IGE	Informatique et gestion d'organisation
Ouahbi	Fouad	IGE	Informatique et gestion d'organisation
Ait Taleb	Samir	TECC	Techniques cinématographiques

44

FPO: SIL

Mohamed Nemiche



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
répétitive

Rectifier cette erreur va nécessiter la modification de TOUTES les occurrences de la relation (de toutes les lignes de la table) correspondant à des papillons décrits par Carl Linné (ie. des milliers !!).

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'organisation
Belhaj	Amina	IGE	Informatique et gestion d'organisation
Ouahbi	Fouad	IGE	Informatique et gestion d'organisation
Ait Taleb	Samir	TECC	Techniques cinématographiques



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
répétitive

Rectifier cette erreur va nécessiter la modification de TOUTES les occurrences de la relation (de toutes les lignes de la table) correspondant à des papillons décrits par Carl Linné (ie. des milliers !!).

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'entreprise
Belhaj	Amina	IGE	Informatique et gestion d'organisation
Ouahbi	Fouad	IGE	Informatique et gestion d'organisation
Ait Taleb	Samir	TECC	Techniques cinématographiques



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
répétitive

Rectifier cette erreur va nécessiter la modification de TOUTES les occurrences de la relation (de toutes les lignes de la table) correspondant à des papillons décrits par Carl Linné (ie. des milliers !!).

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'entreprise
Belhaj	Amina	IGE	Informatique et gestion d'entreprise
Ouahbi	Fouad	IGE	Informatique et gestion d'organisation
Ait Taleb	Samir	TECC	Techniques cinématographiques



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
répétitive

Rectifier cette erreur va nécessiter la modification de TOUTES les occurrences de la relation (de toutes les lignes de la table)

Nom	Prénom	Code Filière	Filière
Alami	Ali	IGE	Informatique et gestion d'entreprise
Slaoui	Rachid	IGE	Informatique et gestion d'entreprise
Nasri	Hind	IGE	Informatique et gestion d'entreprise
Belhaj	Amina	IGE	Informatique et gestion d'entreprise
Ouahbi	Fouad	IGE	Informatique et gestion d'entreprise
Ait Taleb	Samir	TECC	Techniques cinématographiques




3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
~~répétitive~~

Relation Etudiants1

Nom	Prénom	Code Filière
Alami	Ali	IGE
Slaoui	Rachid	IGE
Nasri	Hind	IGE
Belhaj	Amina	IGE
Ouahbi	Fouad	IGE
Ait Taleb	Samir	TECC

Relation Filières

Code Filière	Filière
IGE	Informatique et gestion d'organisati
TECC	Techniques Cinématographique

Après normalisation, si erreur il y a, on pourra la rectifier en une seule écriture modificative. Le gain de temps peut être appréciable.

49 FPO: SIL Mohamed Nemiche




3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

modification
~~répétitive~~

Relation Etudiants1

Nom	Prénom	Code Filière
Alami	Ali	IGE
Slaoui	Rachid	IGE
Nasri	Hind	IGE
Belhaj	Amina	IGE
Ouahbi	Fouad	IGE
Ait Taleb	Samir	TECC

Relation Filières

Code Filière	Filière
IGE	Informatique et gestion d'entreprise
TECC	Techniques Cinématographique

Après normalisation, si erreur il y a, on pourra la rectifier en une seule écriture modificative. Le gain de temps peut être appréciable.

50 FPO: SIL Mohamed Nemiche



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

- Théorie de la normalisation : limiter ces problèmes Analyse des dépendances entre les attributs qui sont à l'origine des phénomènes redondances
- Poser des méthodes systématiques pour décomposer une relation en plusieurs relations plus adaptées pour la gestion du problème
 - ◇ La relation d'origine doit pouvoir être retrouvée par jointures



3) Dépendance fonctionnelle et normalisation

3.1 Intérêt de la normalisation

- Idée : spécifier les relations entre les attributs (notion de dépendance entre les informations).
- Définition : Soit $R(\Delta)$ une relation, Δ ses attributs, et $X \subseteq \Delta$ et $Y \subseteq \Delta$, il existe une dépendance fonctionnelle (DF) entre X et Y (on dit aussi que X détermine Y), notée $X \rightarrow Y$, si dans la relation chaque valeur de X détermine une et une seule valeur de Y .
- X est appelé source et Y cible.
- i.e. $X \rightarrow Y$: si deux tuples ont même valeur sur X alors ils ont même valeur sur Y .



3) Dépendance fonctionnelle et normalisation

3.2) Notion de dépendance fonctionnelle

A	B	C	D	E
a1	b1	c1	d3	e2
a1	b1	c3	d4	e3
a2	b2	c4	d2	e1
a3	b1	c1	d3	e2
a2	b2	c4	d2	e1

	\rightarrow ou \nrightarrow ?	
A	\rightarrow	B
A	\nrightarrow	C
B,C	\rightarrow	D
A,C	\rightarrow	D
B	\nrightarrow	D
D	\rightarrow	E
A,C	\rightarrow	E

53

FPO: SIL

Mohamed Némiche



3) Dépendance fonctionnelle et normalisation

3.2) Notion de dépendance fonctionnelle

Règles de Armstrong

Tous les attributs sont pris dans un ensemble Δ :

- Si $Y \subseteq X \subseteq \Delta$, alors $X \rightarrow Y$ (réflexivité)
- Si $X \rightarrow Y$ et $Z \subseteq \Delta$, alors $XZ \rightarrow YZ$ (augmentation)
NB ($XZ = X \cup Z$)
- Si $X \rightarrow Y$ et $Y \rightarrow Z$, alors $X \rightarrow Z$ (transitivité)

54

FPO: SIL

Mohamed Némiche



3) Dépendance fonctionnelle et normalisation

3.2) Notion de dépendance fonctionnelle

Exemple:

Démontrer que $AD \rightarrow BE$ en ayant les dépendances fonctionnelles suivantes:

$$A \rightarrow B$$

$$B, C \rightarrow D$$

$$A, C \rightarrow E$$

$$D \rightarrow E$$

$$A, C \rightarrow E$$

Dem:

$$\begin{array}{l} (1) A \rightarrow B \xrightarrow{\text{(Augmentation)}} AD \rightarrow BD \\ (2) D \rightarrow E \xrightarrow{\text{(Augmentation)}} BD \rightarrow BE \\ (3) (1) + (2) \xrightarrow{\text{(Transitivité)}} AD \rightarrow BE \end{array}$$



3) Dépendance fonctionnelle et normalisation

3.2) Notion de dépendance fonctionnelle

Pour simplifier, 3 autres règles qui peuvent se déduire des 3 premières.

- **Union:** $X \rightarrow Y$ et $X \rightarrow Z \implies X \rightarrow YZ$
- **Décomposition :** $X \rightarrow Y$ et $Z \subseteq Y$ alors $X \rightarrow Z$
- **Pseudo-transitivité:** $X \rightarrow Y$ et $WY \rightarrow Z \implies WX \rightarrow Z$



Typologie des dépendances fonctionnelles

Une dépendance $X \xrightarrow{R} Y$ est :

- *triviale* si $Y \subseteq X$
- *élémentaire* si pour tout $X' \subset X$, la dépendance fonctionnelle $X' \xrightarrow{R} Y$ n'est pas vraie (Y ne dépend pas fonctionnellement d'une partie de X)
- *canonique* si sa partie droite ne comporte qu'un seul attribut
- *directe* si (i) elle est élémentaire et si (ii) Y ne dépend pas transitivement de X

57

FPO: SIL

Mohamed Nemiche

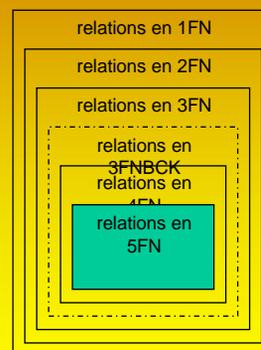


3) Dépendance fonctionnelle et normalisation

3.3) Formes normales

Le concept même de forme normale d'une relation a été introduit par E.D. CODD, qui a proposé une classification des relations en 3 catégories (appelées **formes normales**) reposant sur la nature des dépendances fonctionnelles des relations.

Cette classification a été, plus tard, enrichie. Toutes ces formes (il y en a maintenant 5) correspondent à des restrictions, de plus en plus fortes, pesant sur les dépendances.



58

FPO: SIL

Mohamed Nemiche



3) Dépendance fonctionnelle et normalisation

3.3) Formes normales

1FN A priori, toute relation dont les attributs sont élémentaires est en **première forme normale**.

2FN Une relation est en **deuxième forme normale** si :
 - elle est en première forme normale,
 - tous ses attributs non-clés sont en dépendance fonctionnelle pleine avec la clé.

pas de dépendance non-élémentaire

3FN Une relation est en **troisième forme normale** si :
 - elle est en deuxième forme normale,
 - tous ses attributs non-clés sont en dépendance fonctionnelle directe avec la clé.

pas de dépendance transitive

59

FPO: SIL

Mohamed Nemiche



Première forme normale

- Définition : une relation est dite en *première forme normale* (1NF) si tous ses attributs sont atomiques.
- Les attributs ne peuvent pas être décomposés du point de vue du contexte dans lequel est envisagé la relation.
- Soit la relation :

notes	matiere	etudiant	notes
15		Albert Paul	8, 12.5
15		Duplo Bertrand	2.5, 0, 18
- Pas un 1NF ! :
 notes(matiere,nom,prenom,note1,note2,note3) ou

60

FPO: SIL

Mohamed Nemiche



Deuxième forme normale

- Définition : une relation est dite en *deuxième forme normale* (2FN) si elle est en 1FN et qu'un attribut n'appartenant à aucune clé soit en DF élémentaire avec toutes les clés.
- i.e. : $\langle R(A, B, C), \{AB \rightarrow C, B \rightarrow C\} \rangle$ pas en 2FN.
- Pour passer en 2FN, il suffit d'appliquer une décomposition de la relation basée sur $B \rightarrow C$.
- D'où $\langle R_1(A, B), \{AB \rightarrow B\} \rangle$ et $\langle R_2(B, C), \{B \rightarrow C\} \rangle$

61

FPO: SIL

Mohamed Nemiche



Troisième forme normale

- Définition : une relation est dite en *troisième forme normale* (3FN) lorsqu'elle est en 2FN et que tout attribut n'appartenant pas à une clé ne dépend pas d'un attribut non clé.
- Toutes les DF sont directes.
- Tout attribut n'appartenant pas à une clé est en DF élémentaire directe avec la clé.
- Dès que deux attributs sont en DF et qu'ils ne font pas partie d'une clé alors ils ne sont pas en 3FN.

62

FPO: SIL

Mohamed Nemiche



Troisième forme normale

- i.e. : $\langle R(A, B, C), \{A \rightarrow BC, B \rightarrow C\} \rangle$ pas en 3FN.
- Pour passer en 3FN, il suffit de décomposer la relation avec $B \rightarrow C$
- D'où $\langle R_1(A, B), \{A \rightarrow B\} \rangle$ et $\langle R_2(B, C), \{B \rightarrow C\} \rangle$
- Toute relation admet une décomposition en 3FN à jonction conservative (sans perte) et avec préservation des DF.

63

FPO: SIL

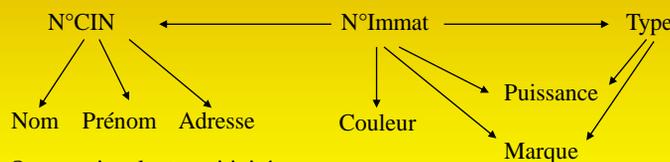
Mohamed Nemiche



Exemple



- Service d'immatriculation de voitures dans une préfecture
 - Soient les DF suivantes :
 - N°Immat \rightarrow Couleur, Type, Puissance, Marque
 - N°CIN \rightarrow Nom, Prénom, Adresse
 - N°Immat \rightarrow N°CIN
 - Type \rightarrow Marque, Puissance
 - On crée le graphe :



On supprime les transitivités
On obtient :

Personne (N°CIN, Nom, Prénom, Adresse)
Voiture (N°Immat, Couleur, Type*, N°CIN*)
Types (Type, Puissance, Marque)

64

FPO: SIL

Mohamed Nemiche

3) Dépendance fonctionnelle et normalisation
3.3) Formes normales

Une relation est en troisième forme normale de Boyce-Codd-Kent si :

3FNBCK

- elle est en troisième forme normale,
- chaque fois qu'une dépendance fonctionnelle non triviale de la relation R est vérifiée, alors sa source contient une clé de R.

ens. des attributs-clés

pas de dépendance vers la clé

Cette dépendance concrétise le fait que la relation n'est pas en 3FNBCK.

65 FPO: SIL Mohamed Nemiche

3) Dépendance fonctionnelle et normalisation
3.3) Formes normales

Une relation est en quatrième forme normale si :

4FN

- elle est en troisième forme normale,
- chaque fois qu'une dépendance multi-valuée non triviale de la relation R est vérifiée, alors sa source contient une clé de R.

pas dans ce cours

attribut X

attribut Y

A un élément de X correspond un sous-ensemble de Y.

X → Y

66 FPO: SIL Mohamed Nemiche



Algorithme de normalisation par synthèse simplifié

- 0 Ecrire les dépendances fonctionnelles (df)
- 1 Rendre canoniques et élémentaires les df qui ne le sont pas
- 2 Représenter les nouvelles df sous forme d'un graphe dont les nœuds sont les attributs impliqués dans les df et les arcs les df elles-mêmes
- 3 Eliminer les df non directes
- 4 Partitionner les df en G_1, \dots, G_n de façon à ce que toutes les df d'une partition aient la même partie gauche
- 5 Constituer une relation R_i par G_i , la partie gauche de G_i étant clé de R_i : chaque R_i est alors en 3NF

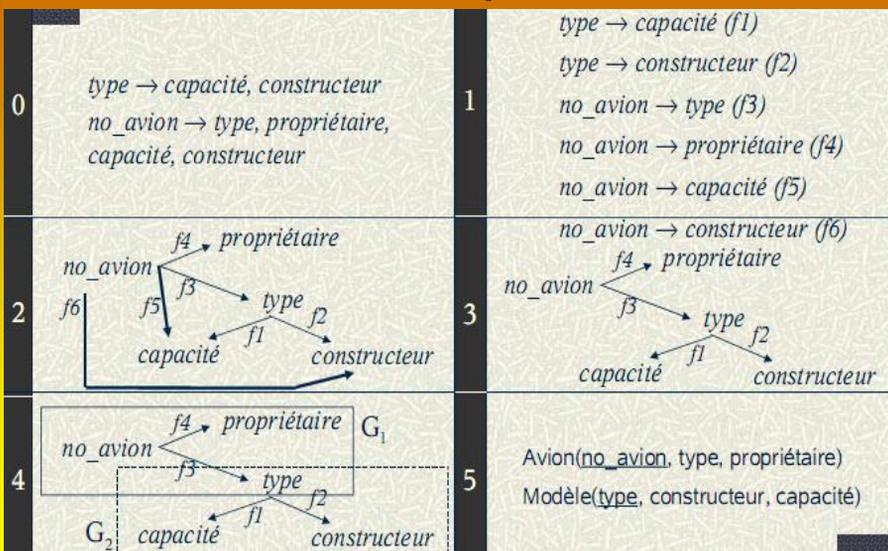
67

FPO: SIL

Mohamed Nemiche



Exemple



68

FPO: SIL

Mohamed Nemiche