

Conception de bases de données

BINV-2040-B

B. Lehmann

2022-2023

Contraintes d'intégrité

Bases de données relationnelles

Conception-Slides2-a.1-ContraintesIntégrité.pptx

Identifiants

Rappel: identifiant

- Un identifiant est un attribut ou un ensemble d'attributs qui permet de distinguer de manière unique une entité.
- (Contrainte d'unicité) Le SGBD impose qu'à tout instant, les tuples aient des valeurs distinctes d'identifiant.
- S'il y a plusieurs identifiants possibles, le plus représentatif est déclaré identifiant primaire. Les autres sont des identifiants secondaires.
- Un identifiant primaire doit être constitué de colonnes obligatoires.

Quelques propriétés des identifiants

- L'ensemble des attributs d'une entité est un identifiant.
- Tout ensemble d'attributs dont une partie stricte est un identifiant est aussi un identifiant (mais *non minimal*).

Ν	uméro	Année	Difficultés	Enoncé	Sujet	Nbre tables
				Organisation des ventes de		
	1	2010	DSD	la société Y	Vente	12
					Firme nettoyage - gestion circuits	
	2	2010	DSD	Euroclean	aller-retour panier	6
	3	2010	DSD	Hopital St Lambert	Complexe!	16
	1	2012	DSD	ToutEnbois	Tarification et ensuite commandes	9+7
					Cumul de petites difficultés en 1	
	2	2012	DSD	RépertoireBibliographique	DSD.	18
	1	2013	Appro	Offres & devis	Relation 1-1	n/a

Quelques propriétés des identifiants

• Un attribut peut appartenir à plusieurs identifiants.

N°co	N°Ligne	Produit	Quantité	Prix unitaire en €
108	1	1456	7	7,89
108	2	1702	2	8,56
108	3	1515	4	9,45
108	4	789	1	14,00

- N°co, N°Ligne est un identifiant.
- N°co, Produit est un identifiant.

Identifiant minimal

- Un identifiant minimal est un identifiant dont on ne peut retirer aucun attribut.
- On ne choisit que des identifiants minimaux pour chaque table d'un schéma conceptuel de base de données.

Numéro	Année	Difficultés	Enoncé	Sujet	Nbre tables
4	2040		Organisation des ventes de		42
1	2010	DSD	la société Y	Vente	12
				Firme nettoyage - gestion circuits	
2	2010	DSD	Euroclean	aller-retour panier	6
3	2010	DSD	Hopital St Lambert	Complexe!	16
1	2012	DSD	ToutEnbois	Tarification et ensuite commandes	9+7

Non minimaux

Recherche des identifiants

Recherche des identifiants : contexte

 On nous donne des données « en vrac », non organisées, sans information suffisante sur les contraintes métiers (exemple : liste des arbres dans un livre).

nom commun	genre	espèce	botaniste
Alisier blanc	Sorbus	aria	L.
Châtaignier	Castanae	sativa	Miller
Cotonéaster commun	Cotoneaster	vulgaris	L.
Epine-vinette	Berberis	vulgaris	L.
Frêne commun	Fraxinus	excelsior	L.
Erable champêtre	Acer	campestre	L.
Bouleau verruqueux	Betula	verrucosa	Ehrh.
Bouleau verruqueux	Betula	pendula	Roth
Sorbier des oiseleurs	Picea	abies	(L.) Karsten
Sorbier	Sorbus	domestica	L.

• On va utiliser le concept de « superclé ».

Superclé

- Une superclé est un groupe d'attributs qui, quand on fixe une valeur pour ce groupe, apparaît dans au plus un tuple.
- L'ensemble des attributs d'une entité est la plus grande superclé.

Recherche des superclés

re	espèce	botaniste
ous	aria	L.
tanae	sativa	Miller
oneaster	vulgaris	L.
beris	vulgaris	L.
kinus	excelsior	L.
r	campestre	L.
ula	verrucosa	Ehrh.
ula	pendula	Roth
ea	abies	(L.) Karsten
ous	domestica	L
t	ous anae oneaster oeris inus · ula ula	aria anae sativa oneaster vulgaris oeris vulgaris inus excelsior campestre ula verrucosa ula pendula a abies

- Toutes les colonnes : nom commun, genre, espèce, botaniste.
- Uniquement trois colonnes :
 - nom commun, genre, espèce
 - nom commun, genre, botaniste
 - nom commun, espèce , botaniste
 - genre, espèce, botaniste

- Uniquement deux colonnes :
 - nom commun, genre
 - nom commun, espèce
 - nom commun, botaniste
 - genre, espèce
 - genre, botaniste
 - espèce, botaniste
- Une seule colonne : impossible

Clé candidate = identifiant minimal

 Une clé candidate ou identifiant minimal est une superclé minimale : chacun de ses attributs/colonnes est nécessaire pour garantir la contrainte d'unicité.

superclés

nom commun, genre, espèce, botaniste nom commun, genre, espèce nom commun, genre, botaniste nom commun, espèce, botaniste genre, espèce, botaniste nom commun, espèce nom commun, botaniste genre, espèce

Clés minimales



nom commun, espèce nom commun, botaniste genre, espèce

Identifiant minimal, primaire

contexte

• On choisit l'identifiant minimal le plus représentatif, qu'on déclare primaire. nom commun, espèce nom commun, botaniste genre, espèce

 Un attribut qui ne fait pas partie de la clé primaire est un attribut non premier, quelle que soit la clé primaire choisie.

Exercice

Révision des concepts

Syllabus

Recherche des identifiants

Agenda dentistes

- Superclés
- Clés candidates
- Clé primaire
- Attributs premiers / non premiers

Date_séance	Heure	No_Patient	Code_Dentiste	Montant_dû	Montant_payé
8/11/2012	8:30	80236	Dubois	60,00	60,00
8/11/2012	10:00	40152	Dubois	75,00	0,00
8/11/2012	8:30	60143	Moreau	135,00	75,00
8/11/2012	9:30	60143	Moreau	107,00	0,00
15/11/2012	8:30	80236	Dubois	75,00	75,00
15/11/2012	10:00	40178	Dubois	135,00	0,00
16/11/2012	8:30	80236	Moreau	75,00	75,00
16/11/2012	8:30	40152	Dubois	75,00	75,00

Superclés Clés candidates

Clé primaire

Attributs premiers / non premiers

Date_séance	Heure	No_Patient	Code_Dentiste	Montant_dû	Montant_payé
8/11/2012	8:30	80236	Dubois	60,00	60,00
8/11/2012	10:00	40152	Dubois	75,00	0,00
8/11/2012	8:30	60143	Moreau	135,00	75,00
8/11/2012	9:30	60143	Moreau	107,00	0,00
15/11/2012	8:30	80236	Dubois	75,00	75,00
15/11/2012	10:00	40178	Dubois	135,00	0,00
16/11/2012	8:30	80236	Moreau	75,00	75,00
16/11/2012	8:30	40152	Dubois	75,00	75,00

1. Recherche des superclés

Date_séance, heure, no_patient, code_dentiste, montant_dû, montant_payé
Date_séance, heure, no_patient, code_dentiste, montant_dû, montant_payé

Date_séance	Heure	No_Patient	Code_Dentiste	Montant_dû	Montant_payé
8/11/2012	8:30	80236	Dubois	60,00	60,00
8/11/2012	10:00	40152	Dubois	75,00	0,00
8/11/2012	8:30	60143	Moreau	135,00	75,00
8/11/2012	9:30	60143	Moreau	107,00	0,00
15/11/2012	8:30	80236	Dubois	75,00	75,00
15/11/2012	10:00	40178	Dubois	135,00	0,00
16/11/2012	8:30	80236	Moreau	75,00	75,00
16/11/2012	8:30	40152	Dubois	75,00	75,00

1. Recherche des superclés (suite)

à trois colonnes au départ Date_séance, heure, no_patient, code_dentiste

Date_séance, heure, no_patient (pas de code_dentiste)

Date_séance, heure, code_dentiste (pas de no_patient)

Date_séance, no_patient, code_dentiste (pas d'heure) — impossible

Heure, no_patient, code_dentiste (pas de date_séance) — impossible

Date_séance	Heure	No_Patient	Code_Dentiste	Montant_dû	Montant_payé
8/11/2012	8:30	80236	Dubois	60,00	60,00
8/11/2012	10:00	40152	Dubois	75,00	0,00
8/11/2012	8:30	60143	Moreau	135,00	75,00
8/11/2012	9:30	60143	Moreau	107,00	0,00
15/11/2012	8:30	80236	Dubois	75,00	75,00
15/11/2012	10:00	40178	Dubois	135,00	0,00
16/11/2012	8:30	80236	Moreau	75,00	75,00
16/11/2012	8:30	40152	Dubois	75,00	75,00

2. Recherche des clés candidates (en fonction des 4 superclés)

Date_séance, heure, no_patient Date_séance, heure, code_dentiste

3. Choix de la clé primaire

Une des deux clés candidates en fonction du contexte \rightarrow ici,

Date séance, heure, code dentiste

09-11-22 Conception 18

Problème des redondances internes

Problème des redondances internes

R(No-frn,	Nom,	Ville,	No-produit,	Prix)
1	Antoine	Bruxelles	1	3,00
2	Bernard	Anvers	5	6,20
1	Antoine	Bruxelles	5	5,95
•••				

Les données « Nom » et « Ville » sont répétées autant de fois qu'il y a de « No-frn »!

On nomme ce phénomène redondance d'information puisqu'une même information est enregistrée plusieurs fois.

Problèmes induits

R(No-frn,	Nom,	Ville,	No-produit,	Prix)
1	Antoine	Bruxelles	1	3,00
2	Bernard	Anvers	5	6,20
1	Antoine	Bruxelles	5	5,95
•••				

Problèmes:

- Il y a un gaspillage d'espace.
- La modification des données du fournisseur doit être faite sur tous les tuples de la table.
- Il y a potentiellement des orthographes différentes ou encore des numéros différents pour le même fournisseur ou un manque de cohérence.

Et:

- Le fournisseur n'est pas une entité isolée, ce qui rompt le principe fondateur des DB relationnelles : toute entité du domaine doit être enregistrée une seule fois.
- On ne peut pas définir un nouveau fournisseur s'il n'a pas fourni de produit (pas de prospection possible).
- On perd les renseignements du fournisseur si on supprime les produits qu'il livre.

Détection et correction des situations de redondance

Pour détecter et corriger les situations de redondance, on introduit une nouvelle forme de

contrainte d'intégrité : la dépendance fonctionnelle.

Dépendance fonctionnelle

Contrainte d'intégrité : dépendance fonctionnelle

R(No-frn, Nom, Ville, No-produit, Prix)

Si deux lignes ont le même No-frn, alors elles doivent avoir les mêmes valeurs de Nom et de Ville.

No-frn détermine Nom, Ville

 Il y a une dépendance fonctionnelle de No-frn vers Nom et Ville

Dépendance fonctionnelle

Un ensemble d'attributs X détermine fonctionnellement un ensemble d'attributs Y SI

Dans tout tuple ayant les mêmes valeurs pour les attributs de X, on a les mêmes valeurs pour les attributs de Y.

Il y a une dépendance fonctionnelle de X vers Y,

Notation: $X \longrightarrow Y$

- X détermine Y (X est un déterminant de Y)
- Y est déterminé par X (il est dépendant de X).

Dépendance fonctionnelle

En base de données relationnelles, il doit donc exister dans chaque table une DF entre l'identifiant de cette table et chacune des colonnes de celle-ci.

Détection des redondances internes

• Il y a donc redondance interne dès qu'il existe un déterminant qui n'est pas un identifiant de la relation.

R(<u>No-frn</u> ,	Nom,	Ville,	No-produit,	Prix)
1	Antoine	Bruxelles	1	3,00
2	Bernard	Anvers	5	6,20
1	Antoine	Bruxelles	5	5,95
•••				

- Dans notre exemple,
 - No-frn est un déterminant dans R() mais il n'en est pas l'identifiant
 - No-frn est l'identifiant du fournisseur mais pas de R().
 - Il existe donc une redondance interne.

Suppression des redondances internes par décomposition

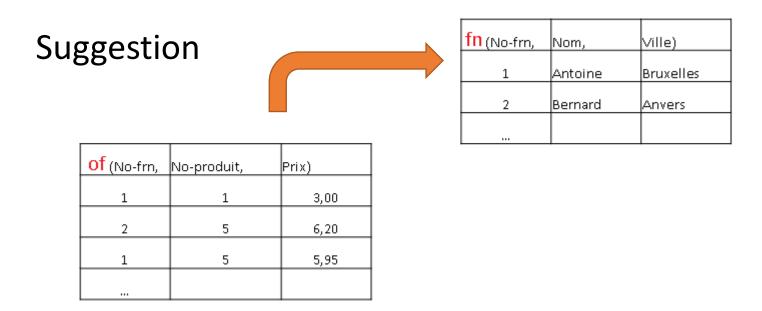
Redondance détectée

R(<u>No-frn</u> ,	Nom,	Ville,	No-produit,	Prix)
1	Antoine	Bruxelles	1	3,00
2	Bernard	Anvers	5	6,20
1	Antoine	Bruxelles	5	5,95
•••				

Cette table contient des données de deux entités.

- L'entité fournisseurs
- L'entité offres_prix d'un fournisseur pour un produit.

Suppression des redondances internes



La dépendance fonctionnelle permet de détecter les situations de redondances; la décomposition permet de les éviter.

Suppression des redondances internes (2)

- En décomposant R en deux fragments R1 et R2 :
 - R1(déterminant 1, déterminé)
 - R2 (déterminant_2, résidu)
 - R2.déterminant 1 est une clé étrangère vers R1.

- Décomposer une table de manière à éliminer ses dépendances anormales consiste à normaliser cette table.
- La décomposition se fait sans perte d'information. On peut reconstituer la table T initiale. (« offres_prix » dans l'exemple).
- Il est essentiel que toutes les tables d'une base de données soient normalisées.

Calcul des identifiants d'une relation

Calcul des identifiants

Les identifiants s'imposent : ils se calculent à partir de l'ensemble des DF dont la relation est le siège.

Relation entre identifiants et DF:

 Tout identifiant d'une relation détermine chaque attribut de cette relation.

Relation client:

- no client → nom, prénom, adresse, tel, email
- Si un groupe d'attributs détermine chacun des attributs de la relation, alors on peut déclarer ce groupe identifiant de la relation.

Relation lignes de commande :

- no com, no produit → qte, prix_unitaire_co
 - no com → date, no client
 - no_produit → description, prix_unitaire

Normalisation

Normalisation

Normaliser une relation consiste à lui appliquer une ou plusieurs décompositions préservant les données afin d'éliminer les problèmes de redondance interne dont elle est éventuellement le siège.

Hainaut, 2018, Chapitre 3 – p 90

Ensemble de règles qui permettent de décomposer la relation de sorte qu'il n'y ait pas de redondances dans les données et que la mise à jour soit simplifiée.

Première forme normale

1FN: Chaque attribut n'a que des valeurs atomiques

Pas de listes comme valeur

On veut éviter la situation suivante :

D(Nom,	Ville,	Chef,	Employés)
Compta	Bruxelles	Jean	René, Paul, Marie
Marketing	Namur	Paul	Pierre, Jules
•••			

D(Nom,	Ville,	Chef,	Employés)
Compta	Bruxelles	Jean	René, Paul, Marie
Marketing	Namur	Paul	Pierre, Jules
•••			



D(Nom,	Ville,	Ch	Employé1,	Employé2	Employé3)
Compta	Bruxelles	Ję		Da	Marie
Marketing	Namur	Paul			

D(Nom,	Ville,	Chef,	Employés)
Compta	Bruxelles	Jean	René, Paul, Marie
Marketing	Namur	Paul	Pierre, Jules
•••			

1FN



D(<u>Nom</u> ,	Ville,	Chef,	<u>Employé</u>)
Compta	Bruxelles	Jean	René
Compta	Bruxelles	Jean	Paul
Compta	Bruxelles	Jean	Marie
Marketing	Namur	Paul	Pierre
Marketing	Namur	Paul	Jules
•••			

Remarque:

• Pour les propriétés composées qui ont du sens conceptuellement, l'éclatement des valeurs concaténées peut être traité après toutes les étapes de normalisation.

Exemple : adresse

Deuxième forme normale

- Une relation est en 2FN si :
 - 1FN +
 - Elle ne contient pas de DF à une partie stricte de la clé.

Qu'est-ce une partie stricte de clé?

A est une partie stricte de B, si A est incluse dans B et A est différente de B. Ce qui implique que A est nécessairement plus petit que B.

Clé	Partie stricte
No_client	
No_commande, no_produit	No_commande et no_produit

La relation R indique l'offre de prix faite par un fournisseur pour un produit.

L'identifiant de la relation, toujours souligné, est produit, fournisseur.

R(produit, fournisseur, nom, ville, prix) est en 1FN

En regardant les DF,

- Produit, fournisseur → prix
- Fournisseur → nom
- Fournisseur → ville

Y a-t-il une (ou +) partie stricte de clé?

On observe que la relation n'est pas en 2FN : il y a des dépendances fonctionnelles à une partie stricte de la clé : fournisseur.

R(produit, fournisseur, nom, ville, prix) pas en 2FN



Décomposition:

R1(<u>fournisseur</u>, nom, ville) en 2FN

R2(produit, fournisseur, prix) en 2FN

Troisième forme normale

- Une relation est en 3FN si :
 - 2FN +
 - Elle ne contient pas de DF transitive.
 - Elle ne contient pas de DF telle qu'un attribut non premier ne dépende pas d'un autre attribut non premier
 - OU
 - Chaque déterminant est un identifiant !

Dans le cas suivant, un produit est vendu par un seul fournisseur. L'identifiant de la relation, toujours souligné, est <u>produit</u>.

R (produit, description, fournisseur, nom, ville, prix)

En regardant les DF,

- Produit → description
- Produit → fournisseur
- Produit \longrightarrow prix
- Fournisseur → nom
- Fournisseur → ville

On observe que la relation n'est pas en 3FN : il y a des dépendances fonctionnelles transitives.

R (<u>produit</u>, description, fournisseur, nom, ville, prix) en 2FN



- R1 (fournisseur, nom, ville)
- R2 (produit, description, fournisseur, prix)

Autres règles?

forme normale de Boyce-Codd

4e forme normale

5e forme normale

...

Implications

Si la DB est normalisée,

 Un identifiant doit déterminer toutes les colonnes de la table.

• Tout déterminant doit être un identifiant d'une table.

Questions?

Exercice

Révision des concepts

Syllabus

Exercice sur les DF Commandes

- Dépendances fonctionnelles
- Clés candidates
- 1FN
- 2FN
- 3FN

Commande

- 1. Dépendances fonctionnelles
- 2. Clés candidates
- 3. 1FN

No client	Nom	Adresse	Nocom	Date	No article	Qte	Prixu
159	Dubois	Bruxelles	17	7/11/2013	7	12	7,59€
159	Dubois	Bruxelles	17	7/11/2013	14	6	6,68€
159	Dubois	Bruxelles	17	7/11/2013	21	3	4,56€
215	Durand	Liège	18	8/11/2013	7	5	7,59€
218	Dupont	Namur	19	8/11/2013	21	4	4,56€
159	Dubois	Bruxelles	20	9/11/2013	21	23	4,56€
159	Dubois	Bruxelles	20	9/11/2013	14	5	6,68€
175	Du Pond	Bruxelles	21	10/11/2013	14	7	6,68€
183	Dubois	Gand	22	10/11/2013	14	11	6,68€
183	Dubois	Gand	22	10/11/2013	8	7	7,87€

1. Recherche des DF

No	Nom	Adresse	Nocom	Date	No	Qte	Prixu
client					article		

a. DF avec déterminant « simple » (une colonne déterminante)

- Noclient → Nom
- Noclient → Adresse
- Nocom \longrightarrow Date
- Nocom → Noclient
- Noarticle → Prixu

un client a un nom

un client a une adresse

une commande est passée à une date

une commande est émise pour un client

un article a un prix unitaire

Vérification que toutes les colonnes participent au moins à 1 DF

Qte ne participe à aucune DF

c. DF avec déterminant « complexe » (plusieurs colonnes déterminantes)

Nocom, Noarticle

 — Qte dans une commande, il y a une quantité par article

2. Calcul de l'identifiant

No	Nom	Adresse	Nocom	Date	No	Qte	Prixu
client					article		

Avant de voir si la relation est normalisée, nous allons lui donner un identifiant.

Pour cela, nous allons nous baser sur les DF.

Les déterminants trouvés sont :

- Nocom
- Noclient
- Noarticle
- Nocom, noarticle

- Trouver la superclé en formant une composition des déterminants : Noclient, Nocom, noarticle .
- Trouver les clés candidates par simplification!

L'identifiant <u>Nocom, noarticle</u> détermine l'entièreté des attributs de la relation.

3. Normalisation

Avec l'identifiant calculé Nocom, noarticle

R (Noclient, Nom, Adresse, <u>Nocom</u>, Date, <u>Noarticle</u>, Qte, Prixu)

La relation Commande est-elle en 1FN ? Oui La relation Commande est-elle en 2FN ? Non

- 1. Les parties strictes de la clé sont :
 - Nocom
 - Noarticle
- 2. Il y a des DF à ces deux parties strictes de clé.
- 3. Il faut donc décomposer.

3.a. Transformation en 2FN

Commande (Noclient, Nom, Adresse, <u>Nocom</u>, Date, <u>Noarticle</u>, Qte, Prixu)



Décomposition:

- R1 (Nocom, Noclient, Nom, Adresse, Date)
- R2 (<u>Noarticle</u>, Prixu)
- R3 (Nocom, Noarticle, Qte)

3.a. Transformation en 2FN

- R1 (Nocom, Noclient, Nom, Adresse, Date)
- R2 (Noarticle, Prixu)
- R3 (Nocom, Noarticle, Qte)

Les relations ainsi obtenues sont-elles en 2FN ? Oui Sont-elles en 3FN ? Non

- 1. Il y a une DF transitive dans la R1.
- 2. Il faut donc décomposer.

3.b. Transformation en 3FN

- R1 (Nocom, Noclient, Nom, Adresse, Date)
- R2 (Noarticle, Prixu)
- R3 (Nocom, Noarticle, Qte)

3FN



- R1a (<u>Noclient</u>, Nom, Adresse)
- R1b (Nocom, Noclient, Date)
- R2 (Noarticle, Prixu)
- R3 (Nocom, Noarticle, Qte)

Les relations ainsi obtenues sont-elles en 3FN? Oui

Etudier et connaître par cœur

Les définitions de tous les concepts des bases de données relationnelles :

- Contraintes d'intégrité.
- Superclé, clé candidate et identifiant minimal.
- Calcul des identifiants.
- Dépendances fonctionnelles.
- Détection des redondances internes.
- Suppression des redondances par décomposition.
- Processus de normalisation.
- Formes normales.

Quelques exercices

Déterminez si les relations suivantes sont en 1FN. Normalisez la relation si elle n'est pas en 1FN.

_ joueurs (<u>matricule</u>, nom, prénom, surnoms)

<u>matricule</u>	nom	prénom	surnoms
F365	Dupont	Jean	Jeannot, Lapin
F789	Durant	Marc	Marco
F357	Térieur	Alain	Al, Capone, Terreur
F456	Albert	Jean	Terreur

_ animaux (<u>nom</u>, propriétaire, race)

Pour chaque transformation illustrée dans les slides suivants (par une flèche bleue),

- Indiquez de quelle(s) transformation(s) il s'agit (1FN et/ou 2FN et/ou 3FN).
- Indiquez si le résultat <u>de ce passage-là</u> est correct (peu importe si le résultat n'est pas le résultat final qui pourrait être obtenu).
- En cas d'erreur, décrivez l'erreur et corrigez-la.

Exercice catégorie

Si DF:

- no_article → nom, description, catégorie
- catégorie \rightarrow tva

R (no article, nom, description, catégorie, tva)



Quelle transformation? Est-elle correcte? Sinon, corrigez-la.

- R1 (catégorie, tva)
- R2 (<u>no article</u>, nom, description, catégorie)

Exercice prime de nuit

R (<u>no machine fixe, no ouvrier, atelier</u>, nom_ouvrier, montant-prime)



Quelle transformation? Est-elle correcte? Sinon, corrigez-la.

- R1 (<u>no ouvrier</u>, nom_ouvrier)
- R2 (<u>no machine fixe, no ouvrier, atelier</u>, montantprime)

Exercice affectations

R (<u>no progr, no projet</u>, equipe, nom, adresse, nom_projet, bâtiment)



Quelle transformation? Est-elle correcte? Sinon, corrigez-la.

- Programmeur (no progr, nom, adresse)
- Projet (no projet, nom_projet)
- Affectation (no progr, no projet, equipe)