

Persistence des données I

DON2

Denis Boigelot, Geneviève Cuvelier, Selim Rexhep, Yannick Voglaire



Haute École Bruxelles-Brabant
École Supérieure d'Informatique

Année académique 2020 / 2021

Plan du cours

- 0 – Présentation
- 1 – Introduction
- 2 – **Dépendance fonctionnelle**
- 3 – Schéma conceptuel
- 4 – Projection et sélection
- 5 – Jointure
- 6 – Agrégat
- 7 – Sous-requête
- 8 – Fichiers

2 – Dépendance fonctionnelle

- 1 Vocabulaire (Table, Ligne, Colonne)
- 2 Redondance interne
- 3 Dépendances fonctionnelles

Table

CLIENT					
NCLI	NOM	ADRESSE	LOCALITE	(CAT)	COMPTE
B062	GOFFIN	72, r. de la Gare	Namur	B2	-3200
B112	HANSENNE	23, r. Dumont	Poitiers	C1	1250
B332	MONTI	112, r. Neuve	Genève	B2	0
B512	GILLET	14, r. de l'Eté	Toulouse	B1	-8700
C003	AVRON	8, r. de la Cure	Toulouse	B1	-1700
C123	MERCIER	25, r. Lemaître	Namur	C1	-2300
C400	FERARD	65, r. du Tertre	Poitiers	B2	350
D063	MERCIER	201, bvd du Nord	Toulouse		-2250
F010	TOUSSAINT	5, r. Godefroid	Poitiers	C1	0
F011	PONCELET	17, Clos des Erables	Toulouse	B2	0
F400	JACOB	78, ch. du Moulin	Bruxelles	C2	0
K111	VANBIST	180, r. Florimont	Lille	B1	720
K729	NEUMAN	40, r. Bransart	Toulouse		0
L422	FRANCK	60, r. de Wépion	Namur	C1	0
S127	VANDERKA	3, av. des Roses	Namur	C1	-4580
S712	GUILLAUME	14a, ch. des Roses	Paris	B1	0

Tables, Lignes, Colonnes

CLIENT					
NCLI	NOM	ADRESSE	LOCALITE	(CAT)	COMPTE
B062	GOFFIN	72, r. de la Gare	Namur	B2	-3200
B112	HANSENNE	23, r. Dumont	Poitiers	C1	1250
B332	MONTI	112, r. Neuve	Genève	B2	0
B512	GILLET	14, r. de l'Eté	Toulouse	B1	-8700
C003	AVRON	8, r. de la Cure	Toulouse	B1	-1700
C123	MERCIER	25, r. Lemaître	Namur	C1	-2300
C400	FERARD	65, r. du Tertre	Poitiers	B2	350
D063	MERCIER	201, bvd du Nord	Toulouse		-2250
F010	TOUSSAINT	5, r. Godefroid	Poitiers	C1	0
F011	PONCELET	17, Clos des Erables	Toulouse	B2	0
F400	JACOB	78, ch. du Moulin	Bruxelles	C2	0
K111	VANBIST	180, r. Florimont	Lille	B1	720
K729	NEUMAN	40, r. Bransart	Toulouse		0
L422	FRANCK	60, r. de Wépion	Namur	C1	0
S127	VANDERKA	3, av. des Roses	Namur	C1	-4580
S712	GUILLAUME	14a, ch. des Roses	Paris	B1	0

schéma

ligne

colonnes obligatoires

colonnes facultatives

données

Domaine, attribut, relation

Nous allons à présent introduire la terminologie spécifique au **modèle relationnel** de base de donnée.

Il s'agit simplement de donner une définition légèrement plus précise des notions de table, ligne, colonne. Est-ce vraiment nécessaire ? Réponse : oui ! Cela nous permettra d'étudier des questions relatives à notre modèle de donnée (l'élimination de la redondance notamment).

Domaine, attribut, relation

Nous considérerons que chaque colonne d'une table est définie sur un certain **domaine**. Un domaine est simplement un ensemble prédéfini de valeurs : l'ensemble des chaînes de caractères, l'ensemble de nombres entiers, l'ensemble des nombres réels, l'ensemble des dates, ...

Exemple : Dans la table CLIENT des slides précédents, nous pourrions considérer que la colonne COMPTE est définie sur l'ensemble des nombres entiers. La colonne NCLI est définie sur l'ensemble des chaînes de caractères de longueur 4 (vu qu'un client est représenté par une chaîne de 4 caractères).

Domaine, attribut, relation

Nous pouvons à présent voir une table T de n colonnes (avec $n \geq 1$ un nombre entier) définie sur les domaine D_1, \dots, D_n comme un sous-ensemble du produit cartésien :

$$T \subseteq D_1 \times D_2 \dots \times D_n.$$

Nous dirons que T est une **relation** définie sur les domaines D_1, \dots, D_n .

Illustrons immédiatement ceci sur un exemple !

Domaine, attribut, relation

Considérons la table ACHAT ci-dessous :

ACHAT		
CLIENT	PRODUIT	PRIX
André	Sucre	2,2
Marc	Sucre	2,2
Marc	Sel	1,4
Anne	Savon	1,4
Anne	Sel	1,4

Nous pouvons considérer que les domaines sont

$D_1 = D_2 =$ l'ensemble des chaînes de caractères,

$D_3 =$ les nombres réels positifs.

Ainsi, $ACHAT \subseteq D_1 \times D_2 \times D_3$. Les lignes de *ACHAT* sont identifiées aux éléments de la relation : (André, Sucre, 2,2), (Marc, Sucre, 2,2), ...

Domaine, attribut, relation

Les colonnes de la table forment donc les composantes de la relation, on appelle ces colonnes des **attributs**. Par exemple : NOM, PRODUIT, PRIX sont les attributs de la relation ACHAT, et on note :

$$\text{ACHAT}(\text{NOM}, \text{PRODUIT}, \text{PRIX}).$$

Si on souhaite être plus précis et préciser les domaines de chaque attribut, on pourrait aussi noter

$$\text{ACHAT}(D1 : \text{NOM}, D2 : \text{PRODUIT}, D3 : \text{PRIX})$$

où $D1, D2, D3$ sont définis comme au slide précédent.

Domaine, attribut, relation

Les lignes d'une relation donnée ayant n attributs sont parfois appelées des **n -uples**. Par exemple :

(B062,GOFFIN, 72 r. de la Gare, Namur, B2, -3200)

est un 6-uple de la relation

CLIENT(NCLI,NOM,ADRESSE,LOCALITE,CAT,COMPTE).

Domaine, attribut, relation

Les relations sont donc des **ensembles** (au sens mathématique) ce qui implique que :

- 1 Deux lignes d'une relation ne sont jamais identiques,
- 2 Il ne peut y avoir d'attribut facultatif.

La notion de relation est donc un peu plus restrictive que la notion intuitive de table. Nous verrons cependant que le modèle relationnel est très utile dans l'étude de certains aspects des base de données (à condition de prendre garde à ces deux restrictions).

Redondance interne

Table répertoriant les livres d'une bibliothèque

LIVRE					
NUMERO	TITRE	AUTEUR	ISBN	DATE_ACHAT	EMPL
1029	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1030	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1032	Mercure	Nothomb A.	2 253 14911 X	14/10/2008	G5
1045	Eva Luna	Allende I.	2 253 05354 6	22/2/2009	F3
1067	Mercure	Nothomb A.	2 253 14911 X	24/2/2009	G5
1022	Mercure	Nothomb A.	2 253 14911 X	3/10/2008	G6

Redondance interne

Observation :

Les données TITRE et AUTEUR sont répétées autant de fois qu'il existe de livres identiques.

LIVRE					
NUMERO	TITRE	AUTEUR	ISBN	DATE_ACHAT	EMPL
1029	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1030	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1032	Mercure	Nothomb A.	2 253 14911 X	14/10/2008	G5
1045	Eva Luna	Allende I.	2 253 05354 6	22/2/2009	F3
1067	Mercure	Nothomb A.	2 253 14911 X	24/2/2009	G5
1022	Mercure	Nothomb A.	2 253 14911 X	3/10/2008	G6

La table viole le 1^{er} principe des bases de données

Tout **fait** du domaine d'application est enregistré une et une seule fois.

Redondance interne

LIVRE					
NUMERO	TITRE	AUTEUR	ISBN	DATE_ACHAT	EMPL
1029	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1030	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1032	Mercure	Nothomb A.	2 253 14911 X	14/10/2008	G5
1045	Eva Luna	Allende I.	2 253 05354 6	22/2/2009	F3
1067	Mercure	Nothomb A.	2 253 14911 X	24/2/2009	G5
1022	Mercure	Nothomb A.	2 253 14911 X	3/10/2008	G6

Problèmes rencontrés :

- ◇ gaspillage d'espace ;
- ◇ une modification de la valeur d'un titre implique la modification pour toutes les lignes similaires ;
- ◇ la suppression d'un exemplaire unique implique la perte des informations sur son auteur et son titre ;
- ◇ les titres et l'auteurs peuvent être orthographié différemment ;
- ◇ ...

Redondance interne

Suggestion

Rassembler les données communes
(ISBN, TITRE, AUTEUR) dans une table spécifique

OUVRAGE		
ISBN	TITRE	AUTEUR
2 02 033300 7	L'humanité perdue	Finkelkraut A.
2 253 14911 X	Mercure	Nothomb A.
2 253 05354 6	Eva Luna	Allende I.

EXEMPLAIRE			
NUMERO	ISBN	DATE_ACHAT	EMPL
1029	2 02 033300 7	14/10/2008	F3
1030	2 02 033300 7	14/10/2008	F3
1032	2 253 14911 X	14/10/2008	G5
1045	2 253 05354 6	22/2/2009	F3
1067	2 253 14911 X	24/2/2009	G5
1022	2 253 14911 X	3/10/2008	G6

Redondance interne

Deux questions :

- 1 Comment détecter les situations de redondance ?
- 2 Comment les corriger ?

La réponse à ces questions repose sur une forme de contrainte d'intégrité :
la dépendance fonctionnelle.

Dépendances fonctionnelles

Notion de dépendance fonctionnelle

LIVRE					
NUMERO	TITRE	AUTEUR	ISBN	DATE_ACHAT	EMPL
1029	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1030	L'humanité perdue	Finkelkraut A.	2 02 033300 7	14/10/2008	F3
1032	Mercure	Nothomb A.	2 253 14911 X	14/10/2008	G5
1045	Eva Luna	Allende I.	2 253 05354 6	22/2/2009	F3
1067	Mercure	Nothomb A.	2 253 14911 X	24/2/2009	G5
1022	Mercure	Nothomb A.	2 253 14911 X	3/10/2008	G6

$$ISBN \rightarrow \{TITRE, AUTEUR\}$$

Si deux lignes ont la même valeur de ISBN, alors elles ont aussi les mêmes valeurs de TITRE et d'AUTEUR

On dit que :

- ◇ il existe une dépendance fonctionnelle de ISBN vers TITRE et AUTEUR
- ◇ ISBN détermine ou est un déterminant de TITRE et AUTEUR
- ◇ TITRE et AUTEUR dépendent de ou sont déterminés par ISBN

Dépendances fonctionnelles

Dépendance fonctionnelle :

- ◇ contrainte d'intégrité très importante du modèle relationnel,
- ◇ proche de l'identifiant mais plus précis,
- ◇ à la base de la théorie de la normalisation.

Dépendances fonctionnelles

ACHAT		
CLIENT	PRODUIT	PRIX
André	Sucre	2,2
Marc	Sucre	2,2
Marc	Sel	1,4
Anne	Savon	1,4
Anne	Sel	1,4

Quelle que soit la ligne, à une même valeur de PRODUIT est toujours associée la même valeur de PRIX

On note :

ACHAT : PRODUIT \rightarrow PRIX

ou

PRODUIT \rightarrow PRIX

PRODUIT est un identifiant de ACHAT[PRODUIT,PRIX]

Dépendances fonctionnelles

Définition

Soit X et Y des ensembles d'attributs dans une relation R . Il existe une dépendance fonctionnelle $X \rightarrow Y$ si, à tout instant, deux lignes de R qui ont les mêmes valeurs pour X ont aussi les mêmes valeurs pour Y .

Abus de notations

On omettra les accolades liées à la notation ensembliste pour un unique attribut. $\{PRODUIT\} \rightarrow \{PRIX\}$ deviendra $PRODUIT \rightarrow PRIX$.

L'union de deux ensembles X et Y sera noté XY .

Exemple

PRODUIT \rightarrow PRIX

- ◇ PRODUIT détermine (fonctionnellement) PRIX ;
- ◇ PRIX dépend (fonctionnellement) de PRODUIT ;
- ◇ PRODUIT est le déterminant et PRIX est le déterminé de la dépendance fonctionnelle.

Exercices

Trouvez les dépendances fonctionnelles pour :

- 1 Un cours n'est donné que par un seul enseignant.
- 2 Un enseignant ne peut pas donner deux cours en même temps.
- 3 Un cours porte sur une seule matière.
- 4 Un numéro de plaque d'immatriculation est lié au propriétaire de la voiture.
- 5 Un modèle Lego est réalisé avec un certain nombre de pièces.
- 6 Suivant la moyenne obtenue par un étudiant pour son année, celui-ci recevra un grade. (12 satisfaisant, 14 distinction...).
- 7 Un code bic est associé à chacune des banques.

Exercices

Donnez les implications des DF suivantes :

◇ *Enseignant* \rightarrow *Cours*

- Un enseignant donne qu'un seul cours (Vrai/Faux)
- Un enseignant donne toujours le même cours (Vrai/Faux)
- Le cours est donné par un seul enseignant (Vrai/Faux)

◇ *code_postal* \rightarrow *ville*

- Une ville est associée à un et un seul code postal.
- Un code postal est associé à une et une seule ville.
- Autre.

La dépendance fonctionnelle *code_postal* \rightarrow *ville* représente-t-elle la situation des codes postaux en Belgique ?

Exercices

◇ *Version_Unicode* → *taille_du_charset*

- Le nombre de caractères Unicode est différent d'une version à l'autre.
- Je connais le nombre de caractère admis dans mon système, je connais donc la version Unicode prise en charge par celui-ci.
- Autre.

◇ $\{puissance, age\} \rightarrow taxe_circulation$

- Si je connais le montant de la taxe de circulation, je connais la puissance et l'âge d'une voiture.
- Si je connais l'âge d'une voiture, je connais le montant de la taxe.
- Si je connais la puissance et l'âge d'une voiture, je connais le montant de la taxe
- Autre.