Université de Sherbrooke

Théorie et modèles relationnels

Normalisation

TMR_06

Christina KHNAISSER (*christina.khnaisser@usherbrooke.ca*)

Luc LAVOIE (*luc.lavoie@usherbrooke.ca*)

Scriptorum/Scriptorum/TMR_06-Normalisation, version 3.1.0.a, en date du 2025-08-28

— document préliminaire —

Plan

Introduction

Le présent document a pour but de présenter une synthèse de la normalisation relationnelle proposée par Edgar F. Codd et développée par la suite grâce aux contributions théoriques et fondamentales de Hugh Darwen, Christopher J. Date, Raymond F. Boyce, Nikos A. Lorentzos, Ronald Fagin et Jeffrey D. Ullman.

La présentation repose sur des bases minimales :

- la logique classique du premier ordre,
- la théorie des ensembles,
- l'arithmétique,
- les langages rationnels.

1. Objectifs et principes de normalisation

1.1. Motivation

En résumé

La normalisation relationnelle a pour objectif de déterminer un ensemble de variables de relation :

- nécessaires et suffisantes pour couvrir un ensemble de prédicats issus du modèle du problème,
- telles que le nombre de redondances fonctionnelles soit minimal.

En outre, la normalisation relationnelle doit fournir l'information nécessaire à la détermination des contraintes garantissant la cohérence des redondances résiduelles.

Ainsi, la normalisation relationnelle contribue à rendre un modèle logique adéquat, plus particulièrement en regard de :

- la cohérence,
- l'efficacité,
- la complétude,
- l'évolutivité.

1.2. Définition

En théorie relationnelle, une forme normale désigne une représentation d'un ensemble de variables de relations permettant d'exclure toute redondance d'un certain type (chaque « forme normale » ciblant un certain type de redondance).

On distingue deux catégories de normalisation :

- **Normalisation des attributs** (1FN).
- **Normalisation des relations** (2FN, 3FN, FNBC, 4FN, FNPJ, FNS).

Note

La FNPJ est souvent désignée comme étant la 5FN et la FNS comme étant la 6FN.

En résumé

Sachant que *FNS* ⇒ *FNPJ* ⇒ *4FN* ⇒ *FNBC* ⇒ *3FN* ⇒ *2FN* ⇒ *1FN*
le module se concentre sur les formes essentielles suivantes :

- La FNS n'est impérative qu'en cas de relativisation (spatiale ou temporelle) des prédicats.
- La FNPJ est la forme ultime des dépendances fonctionnelles et de la décomposition de projection-jointure. Elle doit être ciblée par tout MLD adéquat.
- La FNBC qui permet de simplifier considérablement l'évaluation de la FNPJ ; elle peut aussi se substituer à la FNPJ en la complétant par des contraintes appropriées lorsque, exceptionnellement, la FNPJ est jugée trop contraignante.
- La 1FN peut (doit) être traitée en premier lieu, car les autres formes normales sont définies en la tenant pour acquise.

2. Première forme normale

2.1. Définition

Une relation R est en première forme normale (1FN) si et seulement si tous ses attributs sont atomiques.

Un attribut est atomique si et seulement s'il ne peut être associé, en tout temps, qu'à une et une seule valeur de son type de définition.

Soit R une relation $R(a_1{:}T_1,\ldots,\ a_n{:}T_n)$,

- R est normalisée en regard des attributs si et seulement si
  - $\forall\ t \in R$ la valeur de l'attribut a_i de t est de type $T_i \mid 1 \leq i \leq n$.

On désigne un modèle logique relationnel normalisé en regard des attributs comme étant en « première forme normale » (1FN).

Coloraires

- Un attribut ne peut être sans valeur
- La valeur d’un attribut est unique
- Toute relation est en 1FN par définition (du moins selon le modèle présenté dans le module TRM_02-Fondements).

Observations

- La théorie relationnelle permet l'existence d'attributs non scalaires (tuplets, relations) dans la mesure où les constructeurs de types correspondants sont pris en charge par le modèle relationnel de référence.
- En pratique, les types et les constructeurs de types acceptés dépendent du SGBDR.

2.2. Transformations

Attribut non-clé

Si un attribut non-clé *x* d'une relation *R1* n'est pas en 1FN, car il représente une valeur appartenant plutôt à un *ensemble de T* :

- Retirer *x* de *R1*.
- Ajouter une nouvelle relation totale *R2* dont les attributs sont :
  - les attributs k_1, ..., k_n, d'une clé stricte de R1 (préférablement la clé primaire) ;
  - l'attribut *v:T*.
- Pour chaque valeur de la clé choisie,
  - pour chacune des valeurs *v* constitutives de *x*,
    - définir un tuplet $(k_1, ..., k_n, v)$ dans *R2*.

Attribut clé

Si un attribut clé *a* d'une relation *R1* n'est pas en 1FN, car il représente une valeur appartenant plutôt à un *ensemble de T* :

- Remplacer *a* par un attribut *k:S* (où *S* est un type scalaire de cardinalité suffisante).
- Ajouter une nouvelle relation totale *R2* dont les attributs sont :
    - l'attribut *k:S* ;
    - l'attribut *v:T*.
- Faire correspondre une (et une seule) valeur de *k* à chacune de valeurs de *a*.
- Pour chaque valeur de *a* répertoriée,
    - pour chacune des valeurs *v* constitutives de *a*,
        - définir un couple *(k, v)* dans *R2*,
            - où *k* est la valeur atomique retenue pour correspondre à celle de *a*.

2.3. Exemple

L'Université désire constituer un répertoire des activités proposées par les enseignants et consigner les inscriptions et les résultats (notes) par étudiant, par activité et par type d'évaluation.

Tableau 1. Enseignant, clé {matricule}

matricule	nom	compétences	adresse
324567	Christina K.	IFT187, IGE487, IGE791	Whitehorse, Yukon
465768	Luc L.	IFT187, IGE487, IFT723	Hearst, Ontario
234810	Jules J.	IFT159, IFT313, MAT115	Montréal, Québec
271841	Alice N.	null	Montréal, Québec

Une erreur fréquente consiste à définir le type d'un attribut comme un texte et d'y encoder plusieurs valeurs.

Tableau 2. Enseignant, clé {matricule}

matricule	nom
324567	Christina K.
465768	Luc L.
234810	Jules J.
271841	Alice N.

Tableau 3. Enseignant_Competence, clé {matricule,activite}

matricule	activite
324567	IFT187
465768	IFT187
324567	IGE487
465768	IGE487
324567	IGE791
465768	IFT723
234810	IFT159
234810	IFT313
234810	MAT115

Tableau 4. Enseignant_Adresse, clé {matricule}

matricule	ville	province
324567	Whitehorse	Yukon
465768	Hearst	Ontario
234810	Montréal	Québec
271841	Montréal	Québec

En fait, ce n’est pas la seule interprétation possible de la relation Enseignant_Adresse. D’autres interprétations seront proposées lors du traitement de la FNBC.

3. Forme normale de Boyce-Codd

3.1. Dépendance fonctionnelle

Un attribut A dépend fonctionnellement d'un ensemble d'attributs X si et seulement si (la valeur de) ces derniers permettent toujours de déterminer la valeur (unique) du premier.

Une dépendance fonctionnelle (DF) est notée $X \to A$ où X est le déterminant et A, le déterminé.

Une DF est **triviale** si le déterminé appartient au déterminant :

- $A \in X$

Une DF est **applicable** à une relation R si le déterminant appartient forme un sous-ensemble des attributs de R et si le déterminé est un attribut de R :

- $(X \subseteq R) \land (A \in R)$

La modélisation logique consiste notamment à déterminer (en fonction du problème) quelles sont les DF applicables devant contraindre chacune des relations.

3.2. Relation FNBC

Une relation R est en FNBC relativement à une DF applicable non triviale $X \rightarrow A$ si et seulement si X comprend une clé stricte de R.

Une relation R est dite en FNBC si et seulement si toutes les DF applicables non triviales à R sont issues d'une des clés de R.

3.3. Transformation

En décomposant la relation en projections choisies de telle sorte à faire disparaitre les redondances et dont la jointure est égale à la relation d'origine, c'est-à-dire :

Soit $E = \{a_1, ..., a_n\}$ l'entête de R, choisir une décomposition $D = \{d_1, ... , d_k\}$ telle que

- $d_1 \subseteq E, ..., d_k \subseteq E$
- $R = (R \pi d_1) \bowtie ... \bowtie (R \pi d_k)$
- le nombre de redondances dans D soit le plus petit possible
- s'il existe encore des redondances, les contraindre de façon à ce qu'elle soit cohérente, au mieux identique.

Toute la question se résume donc à choisir la bonne décomposition, c'est-à-dire le bon D.

3.4. Exemple

3.4.1. Enseignant

Tableau 5. Étudiant

matricule	nom
123456	Jeanne
124789	Jean
107834	Odette
203040	Tania
311211	Jean
234567	Mehri
312101	Tangui

$$DF1 : matricule \rightarrow nom$$

Transformation requise

Aucune… si ce n’est de déclarer la bonne clé déterminante !

rel Étudiant {matricule, nom} ***det*** {nom}

3.4.2. Activité

Tableau 6. Activité

code	no	titre	credit	département
IFT	187	Éléments de bases de données	3	Informatique
IMN	117	Acquisition des médias numériques	3	Informatique
IGL	487	Modélisation de bases de données	3	Informatique
IFT	723	Sujets approfondis en bases de données	3	Informatique
IFT	697	Projet d’intégration et de recherche	6	Informatique
BIO	101	Biométrie	3	Biologie
PHY	131	Optique	2	Physique

$$DF2 : code \rightarrow département$$

$$DF3 : code, no \rightarrow titre, credit$$

Transformation requise

Il faut donc décomposer *Activité* en deux, les attributs de l'activité elle-même et ceux de la discipline.

rel Discipline {code, departement} ***det*** {code}
rel Activité {code, no, titre, credit} ***det*** {code, no} ***ref*** {code} → Discipline

Tableau 7. Discipline

code	departement
IFT	Informatique
IMN	Informatique
IGL	Informatique
BIO	Biologie
PHY	Physique

Tableau 8. Activité

code	no	titre	credit
IFT	187	Éléments de bases de données	3
IMN	117	Acquisition des médias numériques	3
IGL	487	Modélisation de bases de données	3
IFT	723	Sujets approfondis en bases de données	3
IFT	697	Projet d’intégration et de recherche	6
BIO	101	Biométrie	3
PHY	131	Optique	2

4. Forme normale de projection-jointure

La FNBC (comme la 2FN et la 3FN) a été formulée de façon à respecter au mieux les dépendances fonctionnelles en recourant, au besoin à des décompositions binaires par projection-jointure.

La FNPJ a été élaborée à la suite de la constatation que certaines redondances ne pouvaient pas être éliminées par décomposition binaire et que des décompositions de degré supérieur pouvaient être requises (en fait, Codd l'avait noté dès 1969, mais la communauté informatique a longtemps négligé ce problème).

La FNPJ est fondée sur la notion de dépendance de jointure et couvre tous les cas de redondance solubles par décomposition de projection-jointure (pour plus de détails, voir les théorèmes de Fagin), la formalisation antérieure de la 4FN (ou forme normale multi-valuée) s’avérant au final un cas particulier de la FNPJ.

Plusieurs autres formes normales ont été développées par ailleurs en regard d’autres formes de redondance, dont, en particulier, la 6FN, qui garantit l’indépendance des prédicats élémentaires et que nous présenterons dans la section suivante.

4.1. Dépendance de jointure

La dépendance de jointure (DJ) est aussi appelée dépendance de projection-jointure (DPJ).

Définition

Une dépendance de jointure (DJ), notée $\bowtie\{D_1, \ldots, D_k\}$ où $D_1, \ldots, D_k$ sont des ensembles d'identifiants d'attributs, est vérifiée par R si et seulement si :

- $D_1 \subseteq id(R), \ldots, D_k \subseteq id(R)$ (applicabilité)
- $R = (R \pi D_1) \bowtie \ldots \bowtie (R \pi D_k)$ (assertion)

Autrement dit, une DJ est vérifiée par une relation R si elle peut être décomposée **sans perte** selon les D_i.

Remarque 1 — corolaire

Une DJ vérifiée par R implique que l'union des D_i est égale à $id(R)$ (l'ensemble des attributs de l'entête de R).

Remarque 2 — DJ triviale

Une dépendance de jointure $\bowtie\{D_1, \ldots, D_k\}$ vérifiée par R est **triviale** si et seulement si au moins un des D_i est égal à $id(R)$.

Une DF vérifiée par R est **strictement triviale** si et seulement si elle est composée du seul élément $id(R)$.

Remarque 3 — induction par les clés

Une DJ est **induite par les clés** de R si chaque D_i est une clé de R (c'est-à-dire, contient une clé stricte de R).

4.2. Relation FNPJ

Une relation R est en forme normale de projection-jointure (FNPJ ou 5FN) si et seulement si chacune des dépendances de jointure non triviales qui lui sont applicables est induite par les clés de R.

Théorème 1

Toute relation en 3FN (*a fortiori* en FNBC) dont au moins un attribut ne participe à aucune clé stricte de R est en FNPJ.

Théorème 2

Toute relation en FNBC dont tous les attributs forment ensemble une clé stricte est en FNPJ.

4.3. Transformations

Si une relation R en FNBC n'est pas en FNPJ, deux transformations peuvent être envisagées pour corriger la situation :

- Remplacer la relation R par une vue calculant la jointure sur la base des sous-relations correspondant aux clés strictes de R.
- Garder la relation R et ajouter une contrainte pour vérifier que la jointure des relations D_i est égale à la relation (cette contrainte pouvant souvent être mise en oeuvre par l'ajout des clés référentielles appropriées).

4.4. Exemple

Soit le modèle conceptuel de données suivant [1] :

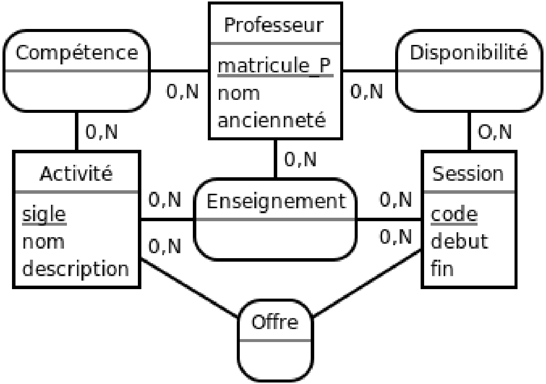

Figure 1. MCD — Université : exemple 5FN

Liste des prédicats

Professeur {matriculeP, nom, ancienneté} clé {matriculeP}

Le professeur dont le matricule est « matriculeP », le nom est « nom » enseigne à l'UdS depuis « ancienneté » an.

Activité {sigle, nom, description} clé {sigle}

L'activité « sigle », décrite par le nom « nom » et la description « description », est offerte par l'UdeS.

Session {code, debut, fin} clé {code}

La session « code » débute le « debut » et se termine le « fin ».

Compétence {matriculeP, activite} clé {matriculeP, activite}

Le professeur « matriculeP » est apte à enseigner l'activité « activite ».

Disponibilité {matriculeP, session} clé {matriculeP, session}

Le professeur « matriculeP » est disponible pour enseigner à la session « session ».

Enseignement {matriculeP, activite, session} clé {matriculeP, activite, session}

Le professeur « matriculeP » enseigne l'activité « activite » à la session « session ».

Offre {activite, session} clé {activite, session}

L'activité « sigle » est programmée à la session « session ».

Liste des contraintes

- Un professeur ne peut pas enseigner une activité qu'à condition qu'il soit apte à le faire.
- Un professeur ne peut pas enseigner une activité à une session donnée qu'à condition qu'elle soit offerte à cette session.
- Un professeur ne peut pas enseigner une activité à une session donnée qu'à condition qu'il soit disponible à cette session.

Voici une instance de base de données valide :

Tableau 9. Compétence

matriculeP	activite
324567	IFT187
324567	IGE487
465768	IFT187
465768	IGE487

Tableau 10. Disponibilité

matriculeP	session
324567	2024-3
465768	2025-1

Tableau 11. Offre

activite	session
IFT187	2024-3
IFT187	2025-1
IGE487	2024-3

Tableau 12. Enseignement

matriculeP	activite	session
324567	IGE487	2024-3
465768	IFT187	2025-1

Soit

- DJ : $\bowtie$ ({matriculeP, session}, {matriculeP, activite}, {activite, session})

Qui correspond à la contrainte suivante :

- Si un professeur est apte à enseigner une activité offerte pour une session, alors le professeur enseigne cette activité à cette session.

Est-ce que $\mathit{Enseignement} = \mathit{Disponibilit\acute{e}} \bowtie \mathit{Comp\acute{e}tence} \bowtie \mathit{Offre}$?

Non, l’instance ne respecte pas la contrainte :

- L’activité "IFT187" est offerte à la session "2024-3".
- Le professeur "324567"
  - est apte à enseigner "IFT187" et
  - est disponible la session "2024-3",
  - mais **il n’enseigne pas "IFT187" la session "2024-3"**.

Alors, la relation Enseignement n’est pas en FNPJ.

5. Forme normale stricte

La forme normale stricte (FNS) est souvent désignée comme sixième forme normale (6FN).

Une relation R est en FNS si et seulement si chacune des dépendances de jointure applicables est triviale.

Ceci a pour effet de placer chaque DF non triviale dans une relation propre. Une telle contrainte est nécessaire lorsqu'une relation doit être relativisée localement (en particulier, dans le cas des chronologies).

Ceci doit être généralisé dans le cadre de la relativisation (spatiale, temporelle ou spatio-temporelle) complète d'une BD. Ceci est traité dans le module `TMR_09-Historicisation`.

Conclusion

- Un modèle relationnel est normalisé en regard d'une forme X si et seulement si toutes ses relations sont normalisées en regard de la forme X.
- Une relation R respectant ses contraintes de type est en 1FN.

- Une relation R est en FNBC si et seulement si chacune des dépendances fonctionnelles qui lui sont applicables est induite par une de ses clés.

  Corolaire

  Une relation R en FNBC est également en 1FN.

- Une relation R est en FNPJ si et seulement si chacune des dépendances de jointure qui lui sont applicables est induite par une de ses clés.

  Corolaire

  Une relation R en FNPJ est également en FNBC.

  Corolaire

  Une relation R en FNBC qui comporte au moins un attribut non-clé est en FNPJ.

  Corolaire

  Une relation totale R en FNBC ne comportant aucune clé composite est en FNPJ.

- Une relation R est en FNS si et seulement si chacune des dépendances de jointure applicables est triviale.

  Corolaire

  Une relation R en FNS est également en FNPJ.

Références

[Elmasri2016]
Ramez ELMASRI et Shamkant B. NAVATHE;
Fundamentals of database systems ;
7th Edition, Pearson, Hoboken (NJ, US), 2016 ;
ISBN 978-0-13-397077-7.

[Date2014a]
Chris J. DATE, Hugh DARWEN, Nikos A. LORENTZOS ;
Time and Relational Theory: Temporal Databases in the Relational Model and SQL ;
Morgan Kaufmann, Waltham (MA, US), 2014 ;
ISBN 978-0-12-800631-3.

[DeSainteMarie2012]
François de SAINTE MARIE;
Bases de données relationnelles et normalisation : de la première à la sixième forme normale ;
Site Developpez.com, 2012 ;
http://fsmrel.developpez.com/basesrelationnelles/normalisation/

Produit le 2025-09-10 12:01:55 UTC

Université de Sherbrooke

[1] (adaptation libre de DeSainteMarie2012)