Modélisation, normalisation et représentation

Amina, KACIMI

Valentin, DEHAINAULT

Minh-Hoang, DANG

December 14, 2018

Préambule

Lors de la phase précédente, nous avons nettoyé la base de données. Les données brutes ont été traité et transférées dans une table accueillante. Cette table, nommée Correction, est à l'image du fichier CSV, ne possède pas de clé et ses attributs sont nullables.

Dans cette phase, nous allons normaliser cette table. A la fin, la table Correction va laisser la place pour les sous-tables.

Déterminer les formes de normales

Avant de commencer, nous avons déterminé la clé primaire pour la table Correction . Nous avons remarqué que l'attribut Article est présent dans toutes les lignes. L'attribut Discriminant sert à compléter l'attribut Article. De ce logique, Article et Discriminant forment la clé primaire de Correction.

Correction				
1	ReferenceCindoc varchar	integer(10)	CO	
	Serie varchar	varchar(255)	N	
9	Article	integer(10)		
8	Discriminant	integer(10)		
I	Ville	varchar(255)	N	
I	CodePostal	varchar(5)	N	
1	Sujet	varchar(255)	N	
	DescDet	varchar(255)	N	
	Date	date	D)	
	NoteBP	varchar(255)	N	
	FicNum	varchar(22)	D3	
	ldx_lco	varchar(255)	00	
1	TailleCli	varchar(255)	N	
	N_V	varchar(3)	D)	
1	C_G	varchar(3)	N	
	Remarques	varchar(255)	N	
	CoordX	float(10)	N	
	CoordY Powered By Visual Paradic	float(10)	N 🐟	

Première forme normale

Correction n'est pas en 1FN. Plusieurs attributs restent divisibles.

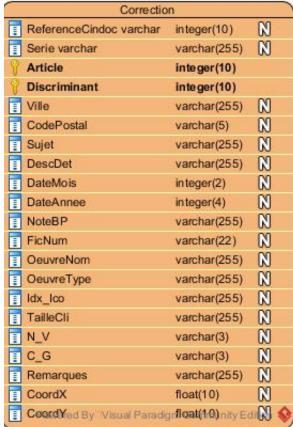
Attribut	Problème	Solution
Date	Les données sont présentes sous forme 'mois/année', 'année'. Nous souhaitons faire des requêtes concernants chaque élément (mois, année)	Un attribut pour mois (numeric), Un attribut pour année (numeric)
idx_per	Les données sont présentes sous forme "nom_oeuvre,type_oeuvre'.	Un attribut pour nom_personne et un attribut pour type_oeuvre.

Deuxième forme normale

La table Correction est maintenant en 1FN. Les dépendances fonctionnelles sont définies comme suivent:

Attribut maitre	Attributs auxiliaires
Ville	CoordX, CoordY
Date	DateMois, DateAnnee
Idx_Pers	OeuvreNom, OeuvreType
Cliche	N_V, C_G, TailleCli
ReferenceCindoc, Serie	Sujet, NoteBP, FicNum, DescDet, Remarques

Elle est également en 2FN car il n'y a pas de DF sur une partie de la clé. mais pas en 3FN, car il y a des DF hors de la clé.



Troisième forme normale

Pour obtenir la table Correction en 3FN, il faut dans un premier temps diviser chacune des DF en des tables distinctes. Ainsi on obtient les tables suivantes :

Nom de table	Colonnes de la table
Ville	idVille, nomVille, CoordX, CoordY
Date	idDate, DateMois, DateAnnee
Oeuvre	idOeuvre, OeuvreNom, OeuvreType
Cliche	idCliche, N_V, C_G, TailleCli
Photo	idPhoto, Sujet, NoteBP, FicNum, DescDet, Remarques
Article	idArticle, Discriminant, ReferenceCindoc, Serie,# idPhoto, #idCliche, #idOeuvre, #idDate, #idVille

Ainsi cet ensemble de données est en 3FN car on a transformé toutes les dépendances fonctionnelles existantes en tables qui sont référencées entre elles par des clés étrangères. Toutefois cette base peut encore être optimisée en terme d'espace car il reste des redondances d'informations au sein des nouvelles tables, en divisant à nouveau les informations nous obtenons les tables suivantes :

Nom de table	Colonnes de la table
Ville	idVille, nomVille, CoordX, CoordY
Date	idDate, DateMois, DateAnnee
IndexPersonne	idOeuvre, nomOeuvre, #idTypeOeuvre
Photo	Article, Discriminant,, NoteBP, FicNum, DescDet, NbreCli, Remarques, ReferenceCindoc, N_V, C_G, #idVille, #idDate, #idSerie.
Sujet	idSujet, nomSujet
ТуреОеиvге	idTypeOeuvre, nomTypeOeuvre
Serie	idSerie, nomSerie .
Cliche	idCliche, Taille
Cliche_Photo	idCliche, PhotoArticle, PhotoDiscriminant

IndexIconographique_Photo	idIco, PhotoArticle, PhotoDiscriminant
Photo_Sujet	PhotoArticle, PhotoDiscriminant, SujetIdSujet

Pour donner un exemple de ce à quoi ces tables servent, prenons le cas de Sujet. Comme le sujet est souvent composé de plusieurs informations dans la table Article, il est judicieux de créer une table Sujet contenant un id et le sujet afin d'éviter les répétitions d'un même sujet qui serait plus coûteuse que la répétition d'un id (entier).

Diagramme final:

