Licence Sciences et Technologies
Mention Informatique

Conception de Bases de Données

Etude de cas

La base de données utilisé dans cette étude de cas est un sous-ensemble d'une base de données réélle, issue des données d'IMDb (imdb.com) et MovieLens (movielens.org)

Exercice 1 Exercice 1 de l'examen 2012-2013/s1

Soit une base de données contenant informations sur des films, des utilisateurs et des evaluations (ratings) des films faites par des utilisateurs. Elle comporte 5 relations :

Movies (MovieId, TitleIMDB, CompanyName, Director)

De chaque film on stocke son identifiant, son titre (tel qu'il apparait dans l'Internet Movie DataBase, i.e., il inclut l'année de production du film), les compagnies de production (il peut avoir plusieurs quand il s'agit d'une coproduction), et son directeur (on stocke un seul directeur par film).

Genres (MovieId, Genre)

Pour chaque film on stocke ses genres (en générale, un film est décrit par plusieurs genres). L'attribut MovieId reference la table Movies.

Countries (MovieId, Country, LongName, DomainCode, IdContinent, Continent, Area, Inhabitants)

Pour chaque film on stocke les pays de production (il peut avoir plusieurs quand il s'agit d'une coproduction), en indiquant, le nom du pays, son nom long, son code de domaine (ex. FR pour la France), son continent (identifiant et description ; si un pays fait partie de deux continents on stocke seulement le continent principale), sa superficie et sa population. L'attribut MovieId reference la table Movies.

Users (UserId, Gender, AgeId, AgeGroup, OccupationId, Occupation, ZipCode)

Pour chaque utilisateur on stocke son identifiant, son sexe, sa tranche d'âge (identifiant et description de la tranche), son occupation (identifiant et description) et code postale.

Evaluations (UserId, MovieId, Rating, Timestamp)

Les utilisateurs réalisent des évaluations des films qu'ils ont regardés. Pour chaque évaluation on stocke l'utilisateur qui l'a réalisée, le film évalué, le rating donnée (entre 1 et 5 étoiles) et le moment de l'évaluation (compté en millisecondes depuis 1/1/1970). L'attribut MovieId reference la table Movies et l'attribut UserId reference la table Users.

La BD compte un million d'évaluation, réalisés par 6040 utilisateurs sur 1881 films. Quelques exemples de tuples sont donnés en annexe.

Questions :

a) Indiquez quelles relations ne sont pas en 1NF en expliquant les violations. Argumentez pourquoi cela pose un problème.
b) Mettez la BD en 1NF. Pour chaque relation obtenue, expliquez la sémantique des nouveaux attributs. Il n'est pas demandé de donner des scripts de migration ni de recopier les exemples de tuples.
c) La BD obtenue est en 1NF mais elle souffre de problèmes de redondance. Donnez deux exemples d'anomalies de mise à jour.
d) Indiquez les dépendances fonctionnelles et les clés de chaque relation.
e) Mettez la BD en 3NF en justifiant la méthode suivie. Indiquez les clés, clés étrangères et les DF de chaque relation obtenue.

Annexe :

genres

MovieId	Genre
1	Adventure
1	Animation
1	Comedy
1	Family
1	Fantasy
2	Action
2	Adventure
2	Comedy
2	Drama
2	Family
2	Fantasy
3	Comedy
3	Romance
4	Comedy
4	Drama
5	Comedy
5	Family

movies

MovieId	TitleIMDB	CompanyName	Director
1	Toy Story (1995)	Pixar Animation Studios	Lasseter, John
1	Toy Story (1995)	Walt Disney Pictures	Lasseter, John
2	Jumanji (1995)	Interscope Communications	Johnston, Joe (I)
2	Jumanji (1995)	Teitler Film	Johnston, Joe (I)
3	Grumpier Old Men (1995)	Lancaster Gate	Deutch, Howard
3	Grumpier Old Men (1995)	Warner Bros. Pictures	Deutch, Howard
4	Waiting to Exhale (1995)	20th Century Fox	Whitaker, Forest
5	Father of the Bride Part II (1995)	Sandollar Productions	Shyer, Charles
5	Father of the Bride Part II (1995)	Taylor-Made Productions	Shyer, Charles
5	Father of the Bride Part II (1995)	The Meyers/Shyer Company	Shyer, Charles
5	Father of the Bride Part II (1995)	Touchstone Pictures	Shyer, Charles

countries

MovieId	Country	LongName	DomainCode	IdContinent	Continent	Area	Inhabitants
1	United States	United States of America	us	1	North America	9631418	299102661
2	United States	United States of America	us	1	North America	9631418	299102661
3	United States	United States of America	us	1	North America	9631418	299102661
4	United States	United States of America	us	1	North America	9631418	299102661
5	United States	United States of America	us	1	North America	9631418	299102661
30	France	French Republic	fr	3	Europe	551695	63587700
30	China, People's Republic of	People's Republic of China	cn	4	Asia	9596960	1315844000
49	Canada	Canada	ca	1	North America	9984670	32547200
53	Italy	Italian Republic	it	3	Europe	301318	58751711
53	France	French Republic	fr	3	Europe	551695	63587700

users

UserId	Gender	AgeId	AgeGroup	OccupationId	Occupation	ZipCode
1	F	1	0-17	10	K-12 student	48067
2	M	56	56+	16	self-employed	70072
3	M	25	25-34	15	scientist	55117
4	M	45	45-49	7	executive/managerial	2460
5	M	25	25-34	20	writer	55455
6	F	50	50-55	9	homemaker	55117
7	M	35	35-44	1	academic/educator	6810
8	M	25	25-34	12	programmer	11413
9	M	25	25-34	17	technician/engineer	61614
10	F	35	35-44	1	academic/educator	95370

evaluations

UserId	MovieId	Rating	Timestamp
1	1	5	978824268
1	48	5	978824351
2	21	1	978299839
5	6	2	978245916
5	16	3	978245645
5	24	1	978242934
5	29	5	978245065
5	32	4	978244962
5	34	4	978244603
5	36	3	978244808
5	39	3	978245037
5	41	4	978244692
5	47	3	978245334