

--	--	--

LI341 Bases de données

EXAMEN - 1ère session du 7 janvier 2015 – 2 heures – documents autorisés

Les téléphones mobiles doivent être éteints et rangés dans les sacs. Le barème sur 60 points (18 questions) n'a qu'une valeur indicative.

1 Requêtes (13 pts)

Considérez le schéma relationnel *BDPedia* pour stocker des extraits (incomplets) de Wikipedia dans les six tables suivantes :

LANGUE (CodeL, Langue)

ARTICLE (Aid, CodeL, Titre, Contenu)

CATEGORIE (Cid, CodeL, NomC)

REFERENCE (Aid, Cid)

PERSONNE (Pid, NomP, Prenom, Email)

ECRIT (Pid, Aid, CodeL)

Les attributs de la clé primaire de chaque table sont soulignés. Les attributs ayant le même nom dans les différentes tables portent sur les mêmes entités (articles, catégories, personnes, langues). Les différentes langues utilisées sont stockées dans la table **LANGUE** avec un code à deux lettres (par ex. 'fr') et le nom de la langue (par ex. 'français'). Chaque article (table **ARTICLE**) a un identifiant unique (Aid) et des contenus et titres dans plusieurs langues. Les articles sont organisés en catégories. Chaque catégorie (table **CATEGORIE**) a un identifiant unique (Cid) et des noms dans plusieurs langues. Un article appartient à une ou plusieurs catégories (table **REFERENCE**) indépendamment de sa langue. Le contenu d'un article dans une langue donnée peut être rédigé par un ou plusieurs auteurs (table **ECRIT**). Les identifiants (Pid), les noms, les prenom et les adresses mail des auteurs sont stockés dans la table **PERSONNE**.

Voici un extrait des trois tables **CATEGORIE**, **ARTICLE** et **REFERENCE** :

CATEGORIE		
<u>Cid</u>	<u>CodeL</u>	NomC
c12	en	Indian art
c12	fr	Art du monde indien
c45	fr	Monuments religieux

ARTICLE			
<u>Aid</u>	<u>CodeL</u>	Titre	Contenu
a23	en	The Taj Mahal	The Taj Mahal...
a23	fr	Le Taj Mahal	Le Taj Mahal...
a23	de	Das Taj Mahal	Das Taj Mahal...

REFERENCE	
<u>Aid</u>	<u>Cid</u>
a23	c12
a23	c45

Question 1 (3 points)

Formuler en **SQL** et en **algèbre relationnelle** la requête qui retourne *les noms d'auteurs d'articles en français (codeL='fr') dans la catégorie 'c12'*.

Réponse :
SQL

ARTICLE (Aid, CodeL, Titre, Contenu) **CATEGORIE** (Cid, CodeL, NomC) **REFERENCE** (Aid, Cid)
PERSONNE (Pid, NomP, Prenom, Email) **ECRIT** (Pid, Aid, CodeL) **LANGUE** (CodeL, Langue)

Réponse : ... les noms d'auteurs d'articles en français (codeL='fr') dans la catégorie 'c12'.

ALGEBRE

Question 2 (4 points)

Formuler en **calcul relationnel** et en **SQL** la requête qui retourne *le titre et la langue des articles qui n'ont qu'un seul auteur*.

Réponse :

CALCUL

SQL

ARTICLE (Aid, CodeL, Titre, Contenu) **CATEGORIE** (Cid, CodeL, NomC) **REFERENCE** (Aid, Cid)
PERSONNE (Pid, NomP, Prenom, Email) **ECRIT** (Pid, Aid, CodeL) **LANGUE** (CodeL, Langue)

Question 3 (3 points)

Formuler en **SQL** la requête qui trouve *les noms des auteurs qui ont rédigé des articles dans au moins deux langues (CodeL) et deux catégories (Cid) différentes.*

Réponse :**SQL****Question 4** (3 points)

Formuler en **SQL** la requête qui retourne *pour chaque catégorie et langue, le nombre d'articles dans cette langue* (on suppose que pour une langue et une catégorie il y a au moins un article de la même langue). Le résultat a la forme suivante :

CodeL	Cid	Categorie	NbArticles
en	67	American TV Series	1245
fr	67	Série TV Americaines	87
en	78	French Cuisine	145
fr	78	La Cuisine Française	745

Réponse : SQL

2 Dépendances fonctionnelles et normalisation (20 pts)

On considère un schéma de relation **R(A,B,C,D,E,H)** et les deux ensembles \mathcal{F} et \mathcal{G} de dépendances fonctionnelles :

$$\mathcal{F} = \{A \rightarrow C; AH \rightarrow D; BC \rightarrow D; AC \rightarrow E; DE \rightarrow A; H \rightarrow B\}$$

$$\mathcal{G} = \{A \rightarrow C; BC \rightarrow D; A \rightarrow E; DE \rightarrow A; H \rightarrow B\}$$

Question 5 (3 points)

On sait que \mathcal{G} est un ensemble minimal. Montrez que \mathcal{G} est une couverture minimale de \mathcal{F} .

Réponse :

Question 6 (5 points)

Donnez les trois clés de la table $R(A, B, C, D, E, H)$ par rapport à \mathcal{F} (ou \mathcal{G}) et montrez comment vous les obtenez (vous n'êtes pas obliges de montrer qu'il n'y a pas d'autres clés).

Réponse :

R(A,B,C,D,E)

$$\mathcal{G} = \{A \rightarrow C; BC \rightarrow D; A \rightarrow E; DE \rightarrow A; H \rightarrow B\}$$

Question 7 (3 points)

Est-ce que le schéma $R(A, B, C, D, E, H)$ est en 3e Forme Normale (3FN) par rapport à \mathcal{G} ? Est-ce qu'il est en Forme Normale Boyce-Codd (FNBC) ? Justifiez vos réponses.

Réponse :**Question 8** (6 points)

Est-ce que la décomposition $R_1(A, B, C)$ $R_2(A, B, E, H)$ $R_3(A, D, E)$ $R_4(C, D, E)$ de $R(A, B, C, D, E, H)$ est Sans Perte d'Information (SPI) ? Justifiez votre réponse avec la méthode du tableau en montrant le tableau initial et le tableau final avec les DF utilisées.

Réponse :

R(A,B,C,D,E)

$$\mathcal{G} = \{A \rightarrow C; BC \rightarrow D; A \rightarrow E; DE \rightarrow A; H \rightarrow B\}$$

Question 9 (3 points)

Utilisez la couverture minimal \mathcal{G} pour donner une décomposition de R qui est SPI, SPD et où chaque relation est en 3FN. Donnez les clés de chaque relation.

Réponse :

3 Optimisation (12 pts)

On considère à nouveau le schéma relationnel *BDPedia*. On suppose qu'il existe 5 langues différentes (attribut *CodeL*) et qu'il y a le même nombre de nuplets pour chaque langue dans la table **ARTICLE** (distribution uniforme). Il y a 400 articles distincts (attribut *Aid*). Tous les articles sont rédigés en exactement deux langues (le même nombre d'articles pour chaque couple distincts de langues). Il y a 100 personnes (auteurs) différentes et chaque personne a un nom différent (*nomP* est unique). La table **ECRIT** contient 5000 tuples et il y a le même nombre d'articles par auteur (distribution uniforme).

ARTICLE (*Aid*, *CodeL*, *Titre*, *Contenu*) **CATEGORIE** (*Cid*, *CodeL*, *NomC*) **REFERENCE** (*Aid*, *Cid*)
PERSONNE (*Pid*, *NomP*, *Prenom*, *Email*) **ECRIT** (*Pid*, *Aid*, *CodeL*) **LANGUE** (*CodeL*, *Langue*)

Justifiez vos réponses aux questions suivantes en précisant le calcul effectué.

Question 10 (1 point)

Soit D le nombre de couples distincts de deux langues différentes. Remarque : un couple ('fr', 'en') ou inversement ('en', 'fr') compte pour un seul couple. Combien vaut D ?

Réponse :

$D =$

Question 11 (1 point)

Combien d'articles (ayant des valeurs *Aid* distinctes) sont rédigés à la fois en '*fr*' et en '*en*' ? Répondre en fonction de D .

Réponse :

Question 12 (3 points)

- a) Quelle est la cardinalité de $\pi_{Aid}(ARTICLE)$?
- b) Combien y a-t-il de nuplets dans la table **ARTICLE** ($card(ARTICLE)$) ?
- c) Quelle est la cardinalité de $\pi_{Aid}(\sigma_{codeL='fr'}(ARTICLE))$?

Réponse :

a) $card(\pi_{Aid}(ARTICLE)) =$

b) $card(ARTICLE) =$

c) $card(\pi_{Aid}(\sigma_{codeL='fr'}(ARTICLE))) =$

ARTICLE (Aid, CodeL, Titre, Contenu) **CATEGORIE** (Cid, CodeL, NomC) **REFERENCE** (Aid, Cid)
PERSONNE (Pid, NomP, Prenom, Email) **ECRIT** (Pid, Aid, CodeL) **LANGUE** (CodeL, Langue)

Question 13 (3 points)

Donnez la cardinalité des requêtes suivantes :

- $Q_1 = ECRIT \bowtie (\sigma_{NomP='Dupont'}(PERSONNE))$
- $Q_2 = (\sigma_{codeL='fr'}(ECRIT)) \bowtie PERSONNE$
- $Q_3 = \sigma_{codeL='fr'}(ECRIT) \bowtie \sigma_{NomP='Dupont'}(PERSONNE)$

Réponse :

$card(Q_1) =$

$card(Q_2) =$

$card(Q_3) =$

Question 14 (1 point)

On considère les deux requêtes :

- $Q_4 = \pi_{Aid, codeL}(\sigma_{codeL='fr' \vee NomP='Dupont'}(ECRIT \bowtie PERSONNE))$
- $Q_5 = \pi_{Aid, codeL}(\sigma_{codeL='fr'}(ECRIT) \bowtie \sigma_{NomP='Dupont'}(PERSONNE))$

Est ce que Q_4 est équivalente à Q_5 ?

Réponse : entourer **oui** ou **non**

Question 15 (3 points)

Transformez Q_4 pour traiter les sélections le plus tôt possible.

Réponse :

4 Transactions (15 pts)

Les deux parties suivantes sont indépendantes.

Partie 1 : Soit une base de données contenant la relation $R(A,B,C)$ avec les trois tuples suivants :

R		
A	B	C
x	y	m
z	t	n
u	v	r

Le gestionnaire de concurrence utilise le verrouillage à deux phases strict. Le verrouillage se fait à la granularité «attribut de n-uplet», ce qui définit 9 *granules* : x, y, z, t, u, v, m, n et r qui sont verrouillées (quand nécessaire) indépendamment des autres granules à chaque opération de lecture et d'écriture dans la base. Les valeurs des granules sont supposées toutes différentes les unes des autres.

Question 16 (4 points)

Quels sont les *granules* verrouillés en lecture (VP) et en écriture (VX) pour les requêtes suivantes :

select A from R where B=y ;

Réponse :

VP :

VX :

update R set B=var where A=z ;

Réponse :

VP :

VX :

select r1.C from R r1 where not exists (select * from R r2 where r1.A=r2.B) ;

Réponse :

VP :

VX :

select r1.C from R r1 where exists (select * from R r2 where r1.A=r2.B) ;

Réponse :

VP :

VX :

Partie 2 : Soient T_1, T_2, T_3, T_4 et T_5 cinq transactions et x, y, z, t quatre granules d'une base de données. On note respectivement $L_i(g)$ et $E_i(g)$ la lecture et l'écriture par la transaction T_i du granule g . Soit la séquence d'opérations (exécution) S suivante :

$S = L4(z), L1(x), L3(y), L4(y), E1(x), E3(y), L4(t), L3(y), L3(z), L4(z), \mathbf{E4(t)}, \mathbf{L1(t)}, E2(t), \mathbf{E2(x)}, L5(y), E5(z), \mathbf{L3(x)}, \mathbf{E5(t)}$

On suppose qu'une transaction se termine dès que la dernière opération de cette transaction est effectuée (ces opérations sont notées en gras).

Question 17 (5 points)

En supposant que les opérations sont exécutées dans l'ordre d'arrivée (celui de la séquence S), dire si l'exécution correspondant à la séquence S est sérialisable ou non, en justifiant votre réponse. Si l'exécution est sérialisable, donner un ordre en série équivalent.

Réponse :

Question 18 (6 points)

On suppose que les opérations arrivent dans l'ordre d'arrivée (de la séquence S) dans un gestionnaire de concurrence contrôle par verrouillage en deux phases strict.

- Montrez le *nouvel ordre d'exécution* produit jusqu'à la fin de la séquence S ou jusqu'à l'apparition d'un interblocage.
- Pendant l'exécution, une *table de verrous* contient pour chaque granule les transactions T_i avec un verrou en lecture (VP) ou un verrou en écriture (VX) en cours. Représenter la table des verrous lorsque la première transaction qui termine a exécuté sa dernière opération *mais n'a pas encore validé* ou, si un interblocage apparaît avant, au moment de l'interblocage.

Réponse :

Ordre d'exécution produit :

Table des verrous :

Granule	x	y	z	t
VX				
VP				