

Bases de Données TD1

Concepts Généraux

1. Soit **E**, **EC** deux fichiers plats d'un Système de Gestion de Fichier manipulés par un langage de programmation, par ex. C++, à structure (schéma) définie comme suit :

E (E#, Nom, Prénom, Née, Email) : EC (E#, C#, Note)

Egalement, soit **E**, **EC** deux tables relationnelles d'une BD relationnelle de MsAccess avec les mêmes schémas que les fichiers.

Indiquez quelques différences entre l'utilisation pratique de ces fichiers et des tables.

2. Soit ETUD une BD avec les tables relationnelles définies comme suit :

E (E#, Nom, Prénom, Sex, NomJF, Statut_marital, Née, Email) : /* Etudiants
EC (E#, C#, Cours_Nom, Note) /* Notes
C (C#, Nom, P#, Vol_C, Vol_TD, Vol_TP) /* Cours
P (P#, Nom_Prenoms) /* Profs

Enrichissez ce schéma par quelques contraintes d'intégrité de votre choix :

mono-attribut

multi-attribut au sein d'une même table

Est-ce que ce schéma vous semble intégré ? Sinon, comment le rendre ainsi ?

- Rappelez la différence entre un lien sémantique et une contrainte d'intégrité référentielle.
- Que veut dire « le langage de manipulation assertionnel » ?
- Pourquoi on appelle les SGBD relationnels ceux de 3-ème génération ?
- Qui a inventé le modèle relationnel, où et à quelle date ?
- Est-ce que la définition ci-dessus de la base ETUD est son CS, son IS, ou son ES ?
- Voici trois commandes principales de définition de données en SQL : Create Table, Create View, Create Index. Elles concernent le CS, le ES ou le IS ?
- On veut construire les ESs suivantes de la BD ETUD et d'une autre BD PERSO avec la table E (E#, Photo). Lesquels sont :
 - conformes à l'architecture ANSI-SPARC ?
 - impossibles de créer avec tout SGBD existant à l'heure actuelle ?

E (E#, Email) :

E (E#, Nom, Prénom, Age)

E (E#, Nom, Prénom, Tel)

E (E#, Nom_Prenom, Photo)

EC (E#, C#, E_Nom, C_Nom, Note) * E_Nom : nom d'étudiant ; C_Nom...

C (C#, Nom, Nbre_Etudiants)

- Indiquez les raisons de la popularité de l'architecture client-serveur par rapport à celle centralisée.
- L'assureur FIAM veut une BD disponible 24/24. Quelles architectures du cours sont envisageables pour le DBA de FIAM ?

12. La FIAM a 1 million d'assurés dans la table

Assuré (A#, Nom, Région, Département, Age, photo...)

Chaque assuré constitue *un tuple* de cette table. Certaines requêtes, telles que $Q = \ll \text{Age moyen d'un assuré} \gg$ demandent à examiner tout tuple de **Assuré**. La table est en totalité sur un disque typique. L'exécution de Q fait venir les tuples un à un, l'accès étant aléatoire, du disque en RAM. A combien de temps estimez vous l'attente d'un usager de FIAM pour avoir sa réponse ?

13. Si le temps estimé vous semble excessif, quelle solution du cours proposez-vous pour que l'attente soit de l'ordre d'une minute max ?

14. Le DBA a appris qu'il peut acheter 32 GO de mémoire flash pour 50€ avec le temps de réponse de l'ordre de 1 ms par accès. Il sait aussi que le dossier d'un assuré sans photo fait 10 KO environ, chaque photo occupant par contre 250 KO. Comment peut-il appliquer utilement cette connaissance au traitement de Q ?

15. La FIAM aimerait remplacer **Assuré** par une collection multibase de tables des assurés, afin d'avoir moins de données par base et répartir les bases entre plusieurs serveurs. Sur quel critère le DBA pourrait définir ces BDs, et comment pourrait être la table des assurés dans chaque base ?

16. La FIAM est un assureur en expansion, ajoutant 20 % de plus d'assurés chaque année. Quel serait votre conseil éclairé au DBA de la FIAM afin qu'il offre le temps de réponse à Q constant ?

17. Expliquez les pourquoi des architectures : (i) SGBD scalable, (ii) dans le nuage, (iii) EWS. Donnez des exemples concrets

1. Soit **E**, **EC** deux fichiers plats d'un Système de Gestion de Fichier manipulés par un langage de programmation, par ex. C++, à structure (schéma) définie comme suit :

E (E#, Nom, Prénom, Née, Email) : EC (E#, C#, Note)

Egalement, soit **E**, **EC** deux tables relationnelles d'une BD relationnelle de MsAccess avec les mêmes schémas que les fichiers.

Indiquez quelques différences entre l'utilisation pratique de ces fichiers et des tables.

On ne peut pas formuler sous forme assertionnelle de requêtes d'usager aux fichiers. Notamment comportant la jointure. Idem pour la formulation de contraintes d'intégrité. Il faut tout programmer. Un travail de longue haleine comparativement.

2. Soit ETUD une BD avec les tables relationnelles définies comme suit :

E (E#, Nom, Prénom, Sex, NomJF, Statut_marital, Née, Email) : /* Etudiants
EC (E#, C#, Cours_Nom, Note) /* Notes
C (C#, Nom, P#, Vol_Cours, Vol_TD, Vol_TP) /* Cours
P (P#, Nom_Prenoms) /* Profs

Enrichissez ce schéma par les contraintes d'intégrité de votre choix (*MsAccess 2007*):

mono-attribut : *Sex = M ou Sex = F ; Note ≥ 0 et Note ≤ 20.*

multi-attribut au sein d'une même table : *Sex = M Imp NomJF is nul*

Est-ce que ce schéma vous semble intégré ? Sinon, comment le rendre ainsi ?

Changer Nom_Prenom en (Nom, Prenom) ou vice versa

3. Rappelez la différence entre un lien sémantique et une contrainte d'intégrité référentielle. *Le lien entre deux attributs A et B signifie qu'une même valeur identifie un même objet. La contrainte d'intégrité référentielle A -> B spécifie que A est la clé primaire et B celle étrangère. Auquel cas, pour toute valeur de B il faut qu'il existe dans la table de un tuple avec une même valeur.*
4. Que veut dire « le langage de manipulation assertionnel » ? *On définit ce que l'on veut obtenir, pas le programme pour le faire.*
5. Pourquoi on appelle les SGBD relationnels ceux de 3-ème génération ? *Voir le cours.*
6. Qui a inventé le modèle relationnel, où et à quelle date ? *Idem*
7. Est-ce que la définition ci-dessus de la base ETUD est son CS, son IS, ou son ES ? *Idem*
8. Voici trois commandes principales de définition de données en SQL : Create Table, Create View, Create Index. Elles concernent le CS, le ES ou le IS ? Successivement : CS, ES et IS.
9. On veut construire les ESs suivantes de la BD ETUD et d'une autre BD PERSO avec la table E (E#, Photo). Lesquels sont :
 - a. conformes à l'architecture ANSI-SPARC ?
 - b. impossibles de créer avec tout SGBD existant à l'heure actuelle ?

E (E#, Email) : *Oui et Non*

E (E#, Nom, Prénom, Age) *Oui et Non*

E (E#, Nom, Prénom, Tel) *Oui et Oui*

E (E#, Nom_Prénom, Photo) *Non et Non*
EC (E#, C#, E_Nom, C_Nom, Note) *Oui et Non*
C (C#, Nom, Nbre_Etudiants) *Oui et Non*

10. Indiquez les raisons de la popularité de l'architecture client-serveur par rapport à celle centralisée. *Performance / Prix*

11. Supposons que l'assureur FIAM a une base qui doit être disponible 24/24. Quelles architectures du cours sont envisageables pour le DBA de FIAM ? *Client Serveurs (Répliquée et/ou BDP) et Distribuée Scalable*

12. Supposons que FIAM a 1 million d'assurés dans la table

Assuré (A#, Nom, Région, Département, Age, photo...)

Chaque assuré constitue *un tuple* de cette table. Certaines requêtes, telles que $Q = \ll \text{Age moyen d'un assuré} \gg$ demandent à examiner tout tuple de **Assuré**. La table est en totalité sur un disque typique. L'exécution de Q fait venir les tuples un à un du disque en RAM. A combien de temps estimez vous l'attente d'un usager de FIAM pour avoir sa réponse ?

On peut compter 10 ms par accès disque aléatoire. Il y en aurait pour 3 heures d'attente.

13. Si le temps estimé vous semble excessif, quelle solution du cours proposez-vous pour que l'attente soit de l'ordre d'une minute max ? *Paralléliser la FIAM sur 200 serveurs.*

14. Le DBA a appris que la mémoire flash coûte 40\$ pour 4 GO que son temps de réponse est de l'ordre de 1 ms par accès et que le dossier d'un assuré sans photo fait 10 KO environ. Comment peut-il appliquer utilement cette connaissance au traitement de Q ? *Acheter cette mémoire. Puis, décomposer Assuré en deux tables soit : AssuréAge (A#, Nom, Région, Dép, Age) et AssuréPhoto (A#, Photo). On voit intuitivement que c'est une décomposition sans perte d'information. Ce qui veut dire que toute valeur dans Assuré pour un assuré identifié par la valeur $A\# = X$ sera aussi dans l'une de deux tables nouvelles dans un tuple avec un même X . La table **AssuréAge** peut entrer dans la mémoire flash. Le temps de réponse diminuera 10 fois au moins. Il reste un peu long néanmoins.*

15. La FIAM aimerait remplacer **Assuré** par une collection multibase de tables des assurés, afin d'avoir moins de données par base et répartir les bases entre plusieurs serveurs. Sur quel critère le DBA pourrait définir ces BDs, et comment pourrait être la table des assurés dans chaque base ?

Par exemple une base par département.

16. La FIAM est un assureur en expansion, ajoutant 20 % de plus d'assurés chaque année. Quel serait votre conseil éclairé au DBA de la FIAM afin qu'il offre le temps de réponse à Q constant ?

Une base scalable sur SD-SQL Server.

18. Expliquez les pourquoi des architectures : (i) SGBD scalable, (ii) dans le nuage, (iii) EWS. Donnez des exemples concrets

Voir le cours. SD-SQL Server, SQL Azure