

TD 6 : Bases de données distribuées
Fragmentation de données

Exercice 01

L'administrateur du portail web d'une entreprise spécialisés en vente par internet (e-commerce) dispose d'une table *Utilisateur* dont le schéma est le suivant :

Utilisateur (IDU, Nom, Prénom, email, mot_de_passe, Date_Enregistrement, Date_Dernier_Accès, Ville, Pays, Code_Postal, Téléphone, Age).

Soient les requêtes R1, R2 et R3 suivantes :

R1 : Select Nom, Prénom, email From Utilisateur where Date_Dernier_Accès < '01-03-2011'

R2 : Select Ville, Pays, Age From Utilisateur where Date_Enregistrement < '01-01-2011'

R3 : Select Nom, Prénom, Ville From Utilisateur where Age>18

Pour optimiser les requêtes R1 et R2, l'administrateur décide de fragmenter verticalement la table *Utilisateur* en trois fragments : User1, User2 et User3. Ces fragments sont définis pour que chaque requête charge uniquement les colonnes nécessaires à son exécution (R1 charge USER1, R2 Charge USER2).

1. Donnez les expressions algébriques permettant de représenter chaque fragment
2. Donnez la réécriture des requêtes R1 et R2 sur la table *Utilisateur* fragmentée.
3. Donnez une réécriture de la requête R3 sur la table *Utilisateur* fragmentée. Que concluez-vous ?

L'entreprise est répartie géographiquement en trois sites distants : Alger, Oran et Constantine. Une allocation simple de ces fragments consiste à mettre chacun sur un site : User1 à Alger, User2 à Oran et User3 à Constantine. Supposons que chaque attribut est codé sur 50 octets, que la table *Utilisateur* contient 1000 000 de tuples et que la taille d'une page système est de 6000 octets. Le nombre d'utilisateurs enregistrés avant le 01-01-2011 est de 50 000, le nombre d'utilisateurs dont le dernier accès a été fait avant le 01-03-2011 est de 5000, le nombre d'utilisateurs âgés de plus de 18 ans est de 600 000.

4. Calculez le coût d'exécution de R1, R2 et R3 avant fragmentation. Ce coût est exprimé en nombre de pages chargées de la mémoire secondaire pour exécuter chaque requête.
 - Exprimez ce coût en secondes sachant que le temps de chargement d'une page est estimé à 1 milliseconde.
5. Sachant que R1 et R2 ont été lancées à Oran et R3 à Constantine.
 - Donnez les stratégies d'exécution possibles pour R1 et R2.
 - Calculez le coût d'exécution global de R1 et R2 dans chaque stratégie. Rappelons que ce coût d'exécution regroupe le coût de chargement de données et le coût de communication. On suppose que le taux de transmission est de 10000 octets par seconde et que le délai d'attente avant de mettre les données sur le canal de transmission est de 1 seconde.
 - Que concluez-vous ?
 - Donnez toutes les stratégies d'exécution de la requête R3
6. Supposons qu'une table T composée de n attributs est fragmentée en m partitions ($m < n$) où chaque partition T_i contient n_i attributs. Donnez l'algorithme de réécriture d'une requête quelconque sur la table T fragmentée.
7. La fragmentation verticale n'est pas supportée sous Oracle. Proposez une implémentation de la fragmentation verticale en utilisant les vues.

Exercice 02

Soit le schéma BD suivant :

PRODUCTION (NUMSERIE, COMPOSANT, MODELE, QUANTITE, MACHINE)
VENTE (NUMSERIE, LOT, NOMCLIENT, NOMVENDEUR, MONTANT)
CLIENT (NOMCLIENT, RUECLIENT, VILLECLIENT, PAYSCLIENT)
VENDEUR (NOMVENDEUR, RUEVENDEUR, VILLEVENDEUR, PAYSVENDEUR)

L'entreprise possède quatre sites de production, tel que chaque site est responsable de la production d'un composant : Unité Centrale, clavier, écran et câble produits respectivement par les sites Tarf, Sétif, Annaba et Alger.

Les points de vente de l'entreprise sont à Sétif, Annaba et Alger. A l'exception des clients de Tarf qui s'adressent au point de vente de Annaba, tous les autres clients s'adressent aux points de vente existant de leurs villes respectives.

Questions :

1. Proposer une fragmentation des quatre relations Production, Vente, client et Vendeur, en justifiant vos choix et précisant le type de fragmentation utilisé.
2. Proposer un schéma d'allocation de tous les fragments.
3. Sachant que la fragmentation est transparente à l'utilisateur et que les requêtes suivantes sont émises à partir du site de Annaba:

R1 : Quelle est la quantité disponible du produit de numéro de série '77y6878'.

R2 : La somme des quantités de produits par composant et modèle.

a- Pour chaque requête donner le script SQL associé, l'arbre canonique ainsi que l'arbre algébrique simplifié.

b- Donner les différentes stratégies d'exécution de la requête **R2**.

Exercice 03

Une école de conduite établie à Alger depuis plusieurs années n'a pas cessé de se développer sur le territoire national au point que ses dirigeants décident de créer des agences au niveau des villes de Constantine et d'Oran.

La base de données centralisée qui avait été développée initialement a la structure suivante :

AGENCE (NumAG, VilleAG, Num-gerant, Nbr-instructeurs)

PERSONNEL-ADMINISTRATIF (NumP, AdresseP, NomP, NumAG, Date-naissance, Fonction)

VEHICULE (NumV, Marque, Type, NumAG, Num-Instructeur)

CLIENT (NumCli, NomCli, AdresseCli, NumAG, Age)

COURS (Num-Instructeur, NumCli, Date-cours, NumV)

EXAMEN (NumCli, Date-examen, Resultat)

Remarques :

- Le gérant et les instructeurs font partie du personnel administratif
- Un véhicule est affecté à un instructeur
- L'école veut créer trois sites pour héberger les données relatives aux agences de chaque ville : Site d'Alger, Site d'Oran et Site de Constantine.

1. Fragmentation

- Proposer un schéma de fragmentation de cette base de données en tenant compte des hypothèses sur les sites. Donnez les expressions algébriques pour chaque table.
- Quelles sont les possibilités de fragmentation de la table COURS. Donner la meilleure solution.
- Donner les requêtes SQL correspondant à la fragmentation des deux relations AGENCE et CLIENT
- Oracle fragmente une table dès sa création. Donner une solution pour fragmenter une table contenant déjà des données. Appliquer la solution sur la table AGENCE.

2. Allocation

- Proposer un schéma d'allocation de tous les fragments.
- Les véhicules appartenant à des agences de la wilaya d'Alger, sont très souvent accédés à partir d'Oran. Proposer une amélioration du schéma d'allocation en prenant en compte cette contrainte. Donner les inconvénients de cette solution.

3. Evaluation des requêtes distribuées

Sachant que la fragmentation est transparente à l'utilisateur et que la requête **R** suivante est émise à partir du site d'Alger : « Donner les nom et numéro de tous les clients de moins de 25 ans, relevant des agences d'Alger ou d'Oran, qui ont eu un résultat positif à l'examen du 22/05/2019 »

- Donner le script SQL associé, l'arbre algébrique global, l'arbre canonique ainsi que l'arbre algébrique simplifié.
- Donner deux stratégies d'exécution de la requête **R**

Exercice 04

Trois établissements universitaires de la ville d'Alger (USTHB, ESI, Alger 01) ont décidé de mutualiser leurs bibliothèques et leur service de prêts, afin de permettre à l'ensemble des étudiants d'emprunter des ouvrages dans toutes les bibliothèques des universités participantes. Par exemple, un étudiant de l'USTHB pourra emprunter des ouvrages à la bibliothèque d'Alger 01.

La gestion commune des bibliothèques et des emprunts est effectuée par une base de données répartie, dont le schéma global est le suivant :

EMPLOYE (Id_pers, nom, adresse, statut, affectation)

*L'attribut **affectation** désigne ici la bibliothèque où travaille l'employé (Bib USTHB, Bib ESI, Bib Alger 01).*

ETUDIANT (Id_etu, nom, adresse, université, cursus, nb_emprunts)

*L'attribut **université** indique l'université où est inscrit l'étudiant*

OUVRAGES (Id_ouv, titre, éditeur, année, domaine, stock, site)

*L'attribut **site** indique la bibliothèque qui gère cet ouvrage (Bib USTHB, Bib ESI, Bib Alger 01).*

*L'attribut **domaine** permet de classer les ouvrages en catégories (physique, maths, informatique, médecine, etc.). L'attribut **stock** désigne le nombre d'ouvrages restant disponibles au prêt.*

PRETS (Id_ouv, Id_etu, date_emprunt, date_retour).

La gestion de cette application s'appuie sur les hypothèses suivantes :

- un employé est affecté à un seul site
- un étudiant est inscrit dans une seule université, mais peut emprunter dans toutes les bibliothèques.
- un ouvrage emprunté dans une bibliothèque est rendu dans la même bibliothèque.

- Le champ *nb_emprunts* de la relation *ETUDIANT* est utilisé pour limiter le nombre d'ouvrages empruntés simultanément par un étudiant sur l'ensemble des bibliothèques. Il est mis à jour lors de chaque emprunt et chaque retour, quelle que soit la bibliothèque d'emprunt.

- Chaque université gère ses propres étudiants

- Chaque bibliothèque gère son personnel et les ouvrages qu'elle détient.

Les relations globales sont fragmentées et réparties sur les différents sites.

1. Proposer un schéma de fragmentation verticale de la table étudiant en deux fragments (Etudiant-Emprunts et Etudiant-Scolarité) pour séparer la gestion des emprunts des étudiants (*nb_emprunts*) de la gestion de scolarité. Donner les expressions algébriques.
2. Proposer un schéma d'allocation du fragment « Etudiant-Emprunts » pour que chaque bibliothèque puisse mettre à jour le nombre d'emprunts localement. Que faut-il faire pour que la BD reste cohérente.
3. Proposer un schéma de fragmentation horizontale du fragment « Etudiant-Scolarité » de la table étudiant. Donner les expressions algébriques et proposer un schéma d'allocation.
4. Proposer un schéma de fragmentation horizontale des tables employés et Ouvrages. Donner les expressions algébriques ainsi que le code SQL.
5. Proposer deux manières permettant de fragmenter horizontalement la table « PRETS ». Donner les expressions algébriques ainsi que le code SQL. Quelle est la plus intéressante ?
6. Dessiner dans une même figure le schéma d'allocation final de toutes les tables sur les trois sites.
7. Un utilisateur de la BD situé à l'université d'Alger 01 souhaite avoir le nombre total des livres du domaine « Informatique » dans les deux autres universités.
 - a. Ecrire la requête SQL répondant à ce besoin
 - b. Générer son arbre canonique.
 - c. Montrer l'arbre après simplification.
 - d. Donner les différentes stratégies d'exécution de cette requête.

Exercice 05

Une entreprise de vente d'appareils électrique dispose de la BD suivante :

Produit (NumP, NomP, Matière, Fournisseur).

Magasin (NumM, CodeWilaya*). Les magasins où sont vendus les produits.

Wilaya(CodeW, NomW). Les wilayas sont codées de 1 à 48.

Ventes (NumM*, NumP*, Date, Quantité, Prix Total). Le produit *NumP* a été vendu dans le Magasin *NumM* à la date mentionnée avec une certaine quantité.

Un utilisateur de la BD situé à Laghouat (Code Wilaya 3) souhaite avoir le **montant total** des ventes de produits à base de **Plastique** dans tous les magasins situés dans la wilaya d'**Alger**.

1. Ecrire la requête R1 correspondante (l'utilisateur ignore que le code de la wilaya d'Alger est 16).
2. Donner deux arbres algébriques de la requête R1.

Pour des besoins de gestion, l'administrateur de la BD souhaite répartir la BD sur quatre sites.

Site 1 : les données concernant les magasins se trouvant dans les wilayas codées de 1 à 15

Site 2 : les données concernant les magasins se trouvant dans les wilayas codées de 16 à 30

Site 3 : les données concernant les magasins se trouvant dans les wilayas codées de 30 à 48

3. Proposer un schéma de fragmentation permettant de réaliser cette répartition.

4. Pour une bonne optimisation, quel schéma de fragmentation proposez-vous pour la table Ventes ?
Donner la requête correspondante ?
5. Après la distribution des données sur les sites, on veut exécuter la requête R1.
6. Reprendre la requête R1 et générer son arbre canonique. Montrer l'arbre après simplification.
7. Sachant que la table Produit est sur le site 3. Donner les différentes stratégies d'exécution de R1.

Exercice 06

Soit une base de données relative à un concessionnaire automobile composée de deux tables : Véhicule et Ventes dont le schéma logique est le suivant.

Véhicule (N_série, Année_première_circulation, marque, type, nb_airbag, nb-vitesses, nb-cylindre, Couleur, Energie, Vitesse_max, Climatisation).

Ventes (N_Série, IDClient, Date_Ventes, Montant).

Le concessionnaire dispose de trois services dont la localisation est la suivante : après-ventes (Site A), marketing (Site B) et technique (Site C).

Le service après-ventes ne s'intéresse qu'aux véhicules vendus mis en service après 2009.

1. Donner le schéma de fragmentation le plus intéressant pour satisfaire les besoins du service après-ventes (Soit Vehicule 1 et Vehicule 2 les deux fragments obtenus).
2. Donner le schéma de fragmentation de la table Ventes en fonction de celui de la table Véhicule (Soit Ventes1 et Ventes 2 les deux fragments obtenus).
3. Proposer une allocation intéressante des fragments (Véhicule 1, Véhicule 2, Ventes 1 et Ventes 2) sur les sites A et B.
4. Donnez les commandes SQL permettant d'effectuer la fragmentation des deux tables.
5. Donner la réécriture de la requête R1 suivante sur le schéma résultat:

R1 : Select Nb_Vitesses, Moy(Montant)
From Vehicule V, Ventes T
where T.N_Série=V.N_Série and V.Année_première_circulation<2011.

3. Donner les stratégies d'exécution de la requête R1 :
 - a. si elle a été lancée au site A
 - b. si elle a été lancée au site C

Le service marketing s'intéresse le plus souvent aux caractéristiques commerciales du véhicule : Année_première_circulation , Marque, Type, Couleur, énergie et Climatisation. Par contre, le service Technique s'intéresse souvent aux caractéristiques techniques : Nb_airbag, nb_Vitesses, nb_Cylindre, Vitesse_max.

1. Donner la nouvelle fragmentation de la table Véhicule pour satisfaire les besoins exprimés par les trois services (sachant que les tables ont été déjà fragmentées comme dans les questions 1 et 2).
2. Comment appelle-t-on ce type de fragmentation ?
3. Donner la réécriture de la requête R1 suite au nouveau schéma de fragmentation.
4. Donnez les stratégies d'exécution de R1 si elle est lancée sur le site C

Exercice 07 (remettre comme devoir)

Soit la base de données suivante qui gère la réservation de vols assurés par différentes compagnies aériennes.

Compagnie (**Nom-Compagnie**, adresse- Siège, Raison-social, Pays, Capital)

Vol Générique (**Num-VG**, Jour, Heure, Durée, Aeroport-Dep*, Aeroport-Arr*, Nom-Compagnie*, type, Escale)

Vol (**Num-Vol**, Date-vol, etat, Num-VG*, Id-Avion*)

Avion (**Id-Avion**, nb-siège, type, constructeur, Pays-Constructeur).

Réservation (**Nres**, Date-Res, Nb-Place, Montant, Num-Vol*, NumCl*)

Client (**NumCl**, nomCl, AdrCl, Ntel-Cl, Pays)

Aéroport (**NomAéroport**, Ville, Pays, Type)

Escale (**Num-Esc**, JourEsc, HeureEsc, Durée)

Fait Escale (**Num-VG**, **NomAéroport**, Num-ESC*)

- L'attribut « Escale » indique si la durée du vol nécessite de faire escale (s'arrêter) dans un aéroport. Il peut prendre l'une des deux valeurs « Oui » ou « Non ».

- L'attribut « Etat » peut prendre les valeurs : Ouvert à la réservation, confirmé, annulé, Fermé à la réservation.

Pour une gestion efficace des réservations, on veut créer trois sites : un site en Algérie pour les pays d'Afrique et de l'Europe, un site au Canada pour les pays d'Amérique (nord et sud) et un site en Chine pour les pays d'Asie et le reste du monde. On veut que chaque site gère les données relatives aux pays qui le concerne.

1. Donner les tables qu'il serait intéressant de fragmenter horizontalement.
 - a. Donnez les expressions algébriques.
 - b. Le code SQL pour les tables Compagnie et Vol-Générique.
 - c. Le script permettant de remplir les tables fragmentées.
2. Identifiez les tables à ne pas fragmenter et les tables où plusieurs fragmentations sont possibles. Justifier les choix.
3. Proposer un schéma d'allocation des fragments sur les sites. Que faire pour les tables non fragmentées pour améliorer les performances.
4. L'administrateur veut savoir le nombre total d'heures de vol pour les avions Airbus effectuant un vol entre Alger et Rome entre le 01/01/2018 et 31/12/2018.
 - a. Donner la requête correspondante en SQL
 - b. Donner l'arbre global, l'arbre canonique et l'arbre simplifié.
5. Pour des besoins spécifiques sur la table Vol-Générique, on veut mettre les informations sur les aéroports de départs et d'arrivée ainsi que la compagnie dans une table séparée.
 - a. Proposer une solution en donnant les expressions algébriques
 - b. Proposer une implémentation de la solution sur un SGBD Oracle. Donner les requêtes correspondantes
 - c. Proposer un nouveau schéma d'allocation
 - d. Proposer un nouvel arbre simplifié.