

Résumé du chapitre

- Une **base de données distribuée** est un ensemble logiquement interdépendant de données partagées (et une description de ces données), physiquement distribuées sur un réseau informatique. Le **DDBMS** est le logiciel qui gère de manière transparente la base de données distribuée.
- Un DDBMS est distinct du **traitement distribué**, où un SGBD centralisé est accessible sur un réseau. Il est également distinct d'un **SGBD parallèle**, qui est un SGBD fonctionnant sur plusieurs processeurs et disques et qui a été conçu pour évaluer les opérations en parallèle, dans la mesure du possible, afin d'améliorer les performances.
- Les avantages d'un DDBMS sont qu'il reflète la structure organisationnelle ; cela rend les données distantes plus partageables ; il améliore la fiabilité, la disponibilité et les performances ; cela peut être plus économique ; il permet une croissance modulaire, facilite l'intégration et aide les organisations à rester compétitives. Les principaux inconvénients sont le coût, la complexité, le manque de normes et l'expérience.
- Un DDBMS peut être classé comme homogène ou hétérogène. Dans un système **homogène**, tous les sites utilisent le même produit SGBD. Dans un système **hétérogène**, les sites peuvent exécuter différents produits SGBD, qui n'ont pas besoin d'être basés sur le même modèle de données sous-jacent, et ainsi le système peut être composé de SGBD relationnels, réseau, hiérarchiques et orientés objet.
- Un **système multibase de données (MDBS)** est un SGBD distribué dans lequel chaque site conserve une autonomie complète. Un MDBS réside de manière transparente au-dessus des bases de données et des systèmes de fichiers existants et présente une base de données unique à ses utilisateurs. Il maintient un schéma global par rapport auquel les utilisateurs émettent des requêtes et des mises à jour ; un MDBS ne gère que le schéma global et les SGBD locaux eux-mêmes conservent toutes les données utilisateur.
- La communication s'effectue sur un réseau, qui peut être un réseau local (LAN) ou un réseau étendu (BLÈME). Les LAN sont destinés à de courtes distances et fournissent une communication plus rapide que les WAN. Un cas particulier du WAN est un réseau métropolitain (MAN), qui couvre généralement une ville ou une banlieue.
- En plus d'avoir la fonctionnalité standard attendue d'un SGBD centralisé, un DDBMS aura besoin de services de communication étendus, d'un catalogue système étendu, d'un traitement distribué des requêtes et de services étendus de sécurité, d'échange de devises et de récupération.
- Une relation peut être divisée en un certain nombre de sous-relations appelées **fragments, qui sont attribuées à un ou plusieurs sites**. Les fragments peuvent **être répliqués** pour améliorer la disponibilité et les performances.
- Il existe deux principaux types de fragmentation : **horizontale** et **verticale**. Les fragments horizontaux sont des sous-ensembles de tuples et les fragments verticaux sont des sous-ensembles d'attributs. Il existe également deux autres types de fragmentation : **mixte** et **dérivée, un type** de fragmentation horizontale où la fragmentation d'une relation repose sur la fragmentation d'une autre relation.
- La définition et l'allocation des fragments sont effectuées de manière stratégique pour obtenir une localité de référence, une fiabilité et une disponibilité améliorées, des performances acceptables, des capacités et des coûts de stockage équilibrés et des coûts de communication minimaux. Les trois règles d'exactitude de la fragmentation sont l'exhaustivité, la reconstruction et la disjonction.
- Il existe quatre stratégies d'allocation concernant le placement des données **centralisées** (une seule base de données centralisée), **fragmenté** (fragments affectés à l'un des), **réplication** complète (copie complète de la base de données conservée sur chaque site) et **réplication sélective** (combinaison des trois premières).
- Le DDBMS doit apparaître comme un SGBD centralisé en offrant une série de transparences. Avec **la transparence de la distribution**, les utilisateurs ne doivent pas savoir que les données ont été fragmentées/répliquées. Avec **la transparence des transactions**, la cohérence de la base de données globale doit être maintenue lorsque plusieurs utilisateurs accèdent simultanément à la base de données et lorsque des échecs se produisent. Avec **la transparence des performances**, le système doit être capable de gérer efficacement les requêtes qui référencent des données sur plusieurs sites. Avec **la transparence du SGBD**, il devrait être possible d'avoir différents SGBD dans le système.

Questions de révision

- 24.1 Expliquez ce que l'on entend par un DDBMS et discutez de la motivation à fournir un tel système.
- 24.2 Comparer et contraster un DDBMS avec un traitement distribué. Dans quelles circonstances choisiriez-vous un DDBMS sur traitement distribué?
- 24.3 Expliquez pourquoi les processus utilisés pour concevoir une base de données relationnelle centralisée ne sont pas les mêmes que ceux d'une base de données relationnelle distribuée.
- 24.4 Discuter des avantages et des inconvénients d'un DDBMS.
- 24.5 Quelle est la différence entre un DDBMS homogène et un DDBMS hétérogène ? Dans quelles circonstances Dans quelles positions de tels systèmes surgiraient-ils généralement ?
- 24.6 Que sont les protocoles réseau et pourquoi sont-ils importants ?
- 24.7 Quelles fonctionnalités attendez-vous d'un DDBMS?
- 24.8 Qu'est-ce qu'un système multibase de données? Décrire une architecture de référence pour un tel système.
- 24.9 L'un des problèmes des DDBMS est celui de la conception de bases de données distribuées. Discutez des problèmes qui doivent être résolus avec la conception de bases de données distribuées. Expliquez comment ces problèmes s'appliquent au catalogue système global.
- 24.10 Quelles sont les différences entre les schémas de fragmentation horizontale et verticale?
- 24.11 Décrire des schémas alternatifs pour fragmenter une relation globale. Indiquez comment vous vérifieriez l'exactitude de s'assurer que la base de données ne subit pas de changement sémantique lors de la fragmentation.
- 24.12 Quelles sont les différences entre les processus d'optimisation des requêtes dans les SGBD centralisés et distribués. Pourquoi ce dernier est-il considéré comme complexe ?
- 24.13 Un DDBMS doit s'assurer que deux sites ne créent pas un objet de base de données avec le même nom. Une solution à ce problème consiste à créer un serveur de noms central. Quels sont les inconvénients de cette approche ? Proposer une approche alternative qui pallie ces inconvénients.
- 24.14 Que sont les systèmes de gestion de bases de données parallèles? Décrire leur architecture et la façon dont ils sont appliqués dans organisations.

Des exercices

Une société d'ingénierie multinationale a décidé de diffuser ses informations de gestion de projet au niveau régional en Grande-Bretagne continentale. Le schéma relationnel centralisé actuel est le suivant:

Employé	(NIN, fName, lName, adresse, DOB, sexe, salaire, taxCode, deptNo)
département	(deptNo, deptName, managerNIN, businessAreaNo, regionNo)
Projet	(projNo, projName, contractPrice, projectManagerNIN, deptNo)
Travaille sur	(NIN, projNo, heures travaillées)
Affaires	(businessAreaNo, businessAreaName)
Région	(regionNo, regionName)

où Employé contient les détails de l'employé et le numéro d'assurance nationale NIN est la clé.

Department contient les détails du département et deptNo est la clé. managerNIN identifie l'employé qui est

le responsable du département Il n'y a qu'un seul responsable pour chaque département.

Projet contient les détails des projets de l'entreprise et la clé est projNo. Le chef de projet est identifié par le projectManagerNIN, et le département responsable du projet par deptNo.

Travaille sur contient le détail des heures travaillées par les employés sur chaque projet et les formulaires (NIN, projNo) la clé.

Affaires contient les noms des domaines d'activité et la clé est businessAreallo.

Les contient les noms des régions et la clé est regionNo.

départements régionaux sont regroupés par région comme suit:

Région 1 : Écosse Région 2 : Pays de Galles Région 3 : Angleterre

Les informations sont requises par domaine d'activité, qui couvre: Génie logiciel, génie mécanique et Ingénierie électrique. Il n'y a pas de génie logiciel au Pays de Galles et tous les départements de génie électrique sont en Angleterre. Les projets sont gérés par les bureaux départementaux locaux.

En plus de la distribution des données au niveau régional, il existe une exigence supplémentaire pour accéder aux données des employés soit par des informations personnelles (par les RH) soit par des informations liées au travail (par la paie).

24.15 Dessinez un diagramme Entité-Relation (ER) pour représenter ce système.

24.16 À l'aide du diagramme ER de l'exercice 24.15, produisez une conception de base de données distribuée pour ce système et incluez:

- (a) un schéma de fragmentation approprié pour le système;
- (b) dans le cas de la fragmentation horizontale primaire, un ensemble minimal de prédicats ;
- (c) la reconstruction de relations globales à partir de fragments.

Énoncez toutes les hypothèses nécessaires pour appuyer votre conception.

24.17 Répétez l'exercice 24.16 pour l'étude de cas *DreamHome* documentée à l'annexe A.

24.18 Répétez l'exercice 24.16 pour l'étude de cas *EasyDrive School of Motoring* documentée à l'annexe B.2.

24.19 Préparer un plan d'action pour le processus de distribution de la base de données des études de cas *Wellmeadows* documentée dans Annexe B.3.

24.20 À l'article 24.5.1, lors de la discussion sur la transparence de la dénomination, nous avons proposé l'utilisation d'alias pour identifier de manière unique chaque réplique de chaque fragment. Fournissez un plan de conception pour la mise en œuvre de cette approche de la transparence des noms.

24.21 Comparez un SGBD distribué auquel vous avez accès aux douze règles de Date pour un DDBMS. Pour chaque règle à laquelle le système n'est pas conforme, donnez les raisons pour lesquelles vous pensez qu'il n'y a pas de conformité à cette règle.