

UNIVERSITÉ DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE
FACULTE D'ELECTRONIQUE ET D'INFORMATIQUE
LABORATOIRE DES SYSTÈMES INFORMATIQUES



Les Bases de Données Distribuées

Pr. Z. Alimazighi

Dr. Kamel Boukhalfa

Boukhalk@gmail.com
<http://boukhalfa.jimdo.com>

PREMIERE PARTIE : Les concepts fondamentaux

1. Introduction
2. Les bases de données distribuées :
 - a) Définitions
 - b) Caractéristiques et avantages de la répartition
3. Le Système de Gestion de BD Réparties : SGBDR
4. Les techniques de répartition de données
5. Architecture de schémas d'une BDR
6. Architecture fonctionnelle Basée sur une approche client/serveur
7. Classification des approches de conception d'une BDR

Alimazighi & Boukhalfa

2

Deuxième PARTIE : SGBD distribués: concepts avancés

1. Gestion de transactions
 - A. Gestion de transactions : Définitions
 - B. Protocoles de gestion de la concurrence
 - C. Validation et reprise
2. Traitement de requêtes
 - A. Optimisation de requêtes
 - B. Stratégies d'évaluation de requêtes

Alimazighi & Boukhalfa

3

PREMIERE PARTIE

LES BASES DE DONNEES DISTRIBUEES
Concepts fondamentaux

INTRODUCTION

- Evolution de la technologie des bases de données
- Expérimentation de cette technologie sur un grand nombre d'applications
- Développement de la technologie réseaux :
- Connexion d'ordinateurs
- Échange de données
- Partage de ressources
- Intégration de la technologie réseau et bases de données



Bases de données distribuées

Alimazighi & Boukhalfa

5

INTRODUCTION

Différentes approches de la répartition

- Approche descendante
- Approche ascendante :
 - BD intégrée
 - BD fédérée

Alimazighi & Boukhalfa

6

INTRODUCTION

Notion de distribution

- **Un système de traitement distribué** : est une *unité d'exécution* de programmes *autonomes*, éventuellement *hétérogènes*, *reliés par un réseau* de communication et *coopérants* à la réalisation de tâches.
- **Une base de données distribuée** : est une collection de bases de données logiquement reliées et physiquement distribuées sur un réseau.
- **Un système de gestion de bases de données distribuées** : est un système logiciel qui fournit une interface logique unique aux différentes bases de données distribuées.

Alimazighi & Boukhalfa

7

INTRODUCTION

Première Génération de SGBD (1962)

- Modèles de données hiérarchiques et réseaux :
 - *représentation des liens 1-1, 1-N, M-N par pointeurs (réseaux) ou par 1-1, 1-N par duplication de données (hiérarchique)*
- Langage de Manipulation de données navigationnel :
 - *parcours de la hiérarchie (Get next)*
- Le programme doit spécifier les chemins d'accès aux données

Alimazighi & Boukhalfa

8

INTRODUCTION

Deuxième génération de SGBD (1970)

- Modèles de données relationnels
 - *Simplicité mais insuffisances de représentation de données complexes*
 - Définition, recherche et mise-à-jour de données grâce à la définition d'opérateurs de l'algèbre relationnelle, et de constructeurs du calcul prédicatif
 - Existence de langage de définition et de manipulation de données de haut niveau (SQL)
 - Prise en compte de la concurrence
 - Gestion de la cohérence

INTRODUCTION

Troisième génération de SGBD (1990)

- **Modèle objet**
 - *Objet = état, identifiant, comportement*
- Encapsulation : *Objet manipulé à travers son interface*
- **Hiérarchie** : *Simple ou multiple*
- **Polymorphisme** : *Existence de plusieurs codes (instances) pour une opération*

LES SYSTEMES REPARTIS

Définition : Système d'information distribué ou réparti

- Un SI réparti est une collection de données physiquement réparties sur plusieurs sites sur laquelle sont développées des applications
- Chaque site supporte sa propre application développée (**application locale**) sur une **BD locale** évoluant autour d'un **schéma conceptuel local**.
- Une **application distribuée** (ou répartie) nécessite l'accès aux multiples applications locales interprétée dans un contexte sémantique commun.
- *Exemple :*
- *Application commerciale distribuée d'une grande entreprise :*
 - *applications locales sur chacun des départements*
 - *coopération des divers départements à l'aide de réseau*

SYSTEMES REPARTIS

Définition : système informatique distribué

- Un système informatique distribué est un ensemble informatique constitué d'unités de traitement et/ou de stockage (pouvant être géographiquement réparties) ayant un large degré **d'autonomie** et interconnectées via un système de communication.
- **Un système d'information réparti est développé sur un système informatique grâce au système de gestion de BD réparti SGBDR**

CARACTERISTIQUES DE LA DISTRIBUTION

Ensemble de BD stockées sur plusieurs machines situées sur des sites différents reliés en réseau

- Les données sont manipulées par un système unique : le SGBDR
- Un utilisateur du système réparti ignore la répartition des données.

Alimazighi & Boukhalfa

13

Avantages de la distribution

- **Reflète une structure organisationnelle:** nombre d'organisations sont aujourd'hui réparties sur plusieurs sites.
- **Amélioration du partage :** distribution des données sur les différents sites; un utilisateur peut accéder aux données de son site tout en pouvant voir les données sur les autres sites.
- **Disponibilité et fiabilité améliorée:** une panne de site ne bloque pas tout le système. Les données dupliquées impliquent la possibilité d'accès en cas de panne d'un site.
- **Performances améliorées:** La parallélisme inhérent à un SGBD distribué implique une vitesse des accès aux données meilleure que dans une BD centralisée.

Alimazighi & Boukhalfa

14

Avantages de la distribution

- **Flexibilité :** en ajoutant des sites au réseau le système réparti peut étendre la BD et enrichir les traitements pour de nouvelles applications
- **Autonomie :** la répartition permet aux différentes structures d'une même entreprise de conserver leurs spécificités (méthodes de travail, perceptions ...)
- **Economies:** La distribution est synonyme de petits équipements, donc la création de système informatique distribué est à coût inférieur si le même système était déployé sur un gros équipement.

Alimazighi & Boukhalfa

15

INCONVENIENTS MAJEURS DE LA DISTRIBUTION

- ❑ Le manque d'expérience des équipes de développement et des utilisateurs finaux.
- ❑ Coûts de déploiement (de communication, d'adaptation des applications et des hommes).
- ❑ Conception et développement complexe : fragmentation, allocation des fragments, réplication etc.
- ❑ Contrôle d'intégrité plus difficile
- ❑ Sécurité: données dupliquées, réseaux

Alimazighi & Boukhalfa

16

LE SYSTEME DE GESTION DE BD REPARTIES (SGBDR)

□ Le SGBDR

- Un SGBDR est un SGBD assurant l'accès coordonné à des données hétérogènes stockées sur des sites différents via des requêtes.
- Une requête interrogeant une BD répartie est dite **requête répartie**. Elle se décompose en un ensemble de **sous-requêtes** sur les BD locales.

□ Architecture fonctionnelle

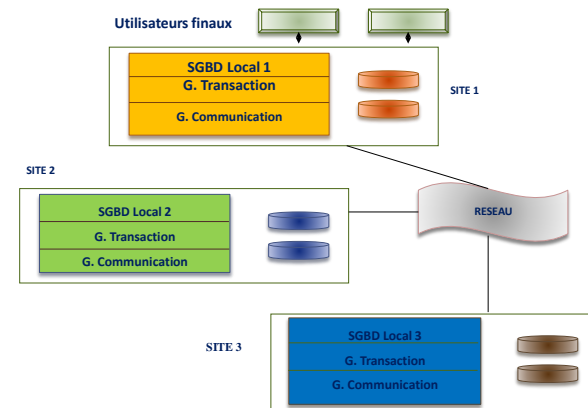
3 composants majeurs :

- SGBD local
- Gestionnaire de transactions
- Une composante de communication

Alimazighi & Boukhalfa

17

Architecture



Alimazighi & Boukhalfa

18

LE SGBDR: les composantes

□ SGBD Local : exécution d'une sous-transaction :

1. Lecture des données de la BD sur un espace de travail
2. Transfert de données entre l'espace de travail et les différents sites
3. Manipulation des données sur les espaces de travail
4. Mise-à-jour de la BD locale

□ Gestionnaire de transaction : coordination et exécution des transactions :

1. Translation des requêtes (en fonction des modèles)
2. Contrôle de concurrence
3. Contrôle de l'exécution de la requête répartie

□ Composante de communication : responsable de l'interconnexion de machines coopérantes.

Elle garantit la délivrance des messages entre les différentes machines sans perte d'information, surveille l'atomicité des transactions et permet certains mécanismes de synchronisation.

Alimazighi & Boukhalfa

19

LE SGBDR: Objectifs

1. Définition et manipulation des ensembles de données réparties sur différents sites comme un seul ensemble logique.
2. **Indépendance à la localisation.**
 - Les données ne résident pas nécessairement sur le site utilisateur. Celui-ci peut ignorer la localisation des données.
 - L'information concernant la localisation des données, est maintenue dans le *dictionnaire de données*, et consultée par le SGBD distribué pour déterminer la localisation des relations impliquées dans la requête des utilisateurs.

Alimazighi & Boukhalfa

20

LE SGBDR: Objectifs

2. Extensibilité:

- C'est la capacité d'augmentation incrémentale, par introduction de nouveaux sites dans le réseau, avec un impact minimal sur les bases de données locales et les programmes d'applications existants.
- Le système de gestion de bases de données distribué doit être capable d'intégrer les nouveaux nœuds ajoutés au réseau.
- L'interface de l'utilisateur global doit être assez évolutive, pour permettre d'intégrer de nouvelles fonctionnalités sans détruire la gestion de l'interface existante.

4. Fiabilité et performance

- Duplication transparente de certaines données, de façon à permettre au système global de router la requête utilisateur à une source alternative, quand la source principale ne fonctionne plus.

Alimazighi & Boukhalfa

21

LE SGBDR : Objectifs

5. Cohérence : la cohérence est le respect des contraintes d'intégrité

- Assurer la cohérence lors de la mise-à-jour des copies.
- Assurer la cohérence de la BD dans le cas de mise-à-jour concurrentes.
- Maintien de la cohérence de la BD en cas de panne.

La cohérence est assurée grâce au concept de transaction: dans les BDD distingue transaction globale, sous-transaction et transaction locale

6. Sécurité

Contrôle d'accès et protection des données confidentielles.

Alimazighi & Boukhalfa

22

LE SGBDR : Les règles de transparence de Stonebraker

1. **Transparence de consultation** : une opération de consultation doit donner le même résultat quelque soit le site
2. **Transparence de mise-à-jour** : émise sur n'importe quel site, une mise-à-jour impliquant plusieurs sites, donnera toujours le même résultat.
3. **Transparence de schémas** : les schémas de tous les sites peuvent être rendus visibles sur n'importe quel site.
4. **Transparence de performance** : les requêtes sont comparables en performance sans tenir compte du site d'émission.
5. **Transparence de transaction** : exécution des transactions réparties concurrentes en maintenant la cohérence de la BD répartie.
6. **Transparence de copies** (indépendance à la duplication) : des copies nécessaires de données peuvent être effectuées sur des sites sans que cela ne soit visible à l'utilisateur.

Alimazighi & Boukhalfa

23

CRITERES DE CLASSIFICATION

• Trois critères essentiels :

- ☐ AUTONOMIE
- ☐ DISTRIBUTION
- ☐ HETEROGENEITE

Alimazighi & Boukhalfa

24

CRITERES DE CLASSIFICATION

- **AUTONOMIE:** L'autonomie est la capacité des SGBDs locaux à traiter les requêtes locales de manière totalement indépendante du système global. Chaque site doit disposer localement de toutes les informations qui se trouvent dans le *dictionnaire de données* afin d'être autonome, sans avoir à compter sur un *dictionnaire de données global centralisé*.

La classification devra se baser sur :

- La mesure du degré d'indépendance
- Le volume des échanges intersites et la capacité d'exécution de transactions indépendamment des autres sites.
- La distribution du contrôle

Alimazighi & Boukhalfa

25

CRITERES DE CLASSIFICATION

□ DISTRIBUTION des données:

Elle mesure le degré de distribution des données. (distribué ou pas)

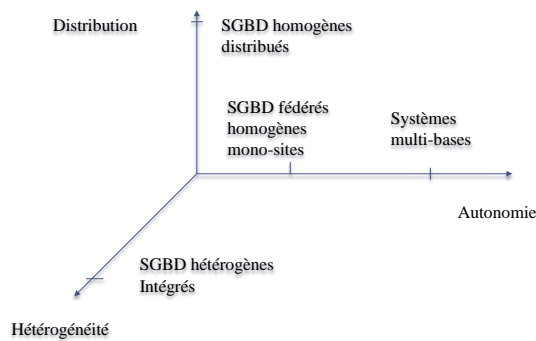
□ HETEROGENEITE: on distingue plusieurs type d'hétérogénéité

- Hardware, protocoles réseaux
- Modèles de données
- Langages

Alimazighi & Boukhalfa

26

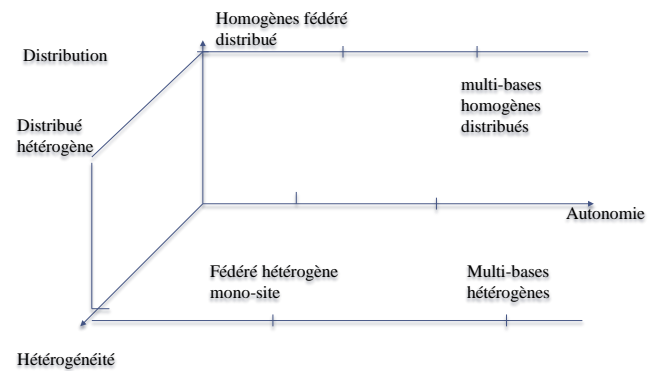
TPOLOGIE DE SGBD



Alimazighi & Boukhalfa

27

TPOLOGIE DE SGBD



Alimazighi & Boukhalfa

28

TPOLOGIE DE SGBD

- **Cas le plus général:** Les systèmes les plus utilisés
 - Systèmes fédérés distribués hétérogènes
 - Multi-bases de données hétérogènes distribués
- *Certains auteurs ne font pas la distinction entre multi-base et système fédéré.*
- Une MultiBase de données intègre les systèmes de gestions de bases de données (SGBD) locaux, pré existants.
- C'est un ensemble de bases de données faiblement couplées et autonomes qu'un utilisateur peut manipuler à l'aide d'un langage spécifique.
- **On peut dire que c'est le degré d'autonomie qui distingue les deux.**

2008/2009

Alimazighi & Boukhalfa

29

Allocation de données

Quatre stratégies

1. Allocation centralisée

Une seule base de données, un seul SGBD, sur un site et des utilisateurs répartis sur le réseau. Tous les sites sauf le site central doivent passer par le réseau pour tous les accès aux données.

→ Coûts de communication très élevés, fiabilité et disponibilité de données réduites

2. Allocation fragmentée ou partitionnée: partitionnement de la BD en fragments disjoints distribués chacun sur un site.

→ Coûts de stockage et de communication réduits, mais fiabilité et disponibilité de données réduites

2008/2009

Alimazighi & Boukhalfa

30

Allocation de données

3. Réplication complète

Une copie de la totalité de la BD sur chaque site.,

Fiabilité, disponibilité optimales, mais coûts de stockage et communication élevés à cause des mises à jour.

4. Réplication sélective

Combinaison de la centralisation, fragmentation, et réplication.

Tous les avantages de chacune des approches tout en évitant leur inconvénients;

2008/2009

Alimazighi & Boukhalfa

31

Les Techniques de répartition de données: la réplication

□ Avantages :

- **Disponibilité des données** : une relation R peut être atteinte à partir de n'importe quel site, en cas d'avarie affectant un site donné.
- **Parallélisme des traitements** : lorsque la majorité des transactions relatives à une relation R ne mettent en jeu que des lectures, plusieurs sites peuvent travailler en parallèle sur cette relation.

□ Inconvénients :

- Alourdissement des procédures de mise-à-jour : toute mise-à-jour doit être appliquée à l'ensemble des sites dépositaires de la relation pour garantir la consistance des répliques des relations.
- Augmentation du volume de la BD

2008/2009

Alimazighi & Boukhalfa

32

Les Techniques de fragmentation de données

□ Avantages de la fragmentation

- **Utile**: en général, les applications travaillent sur des vues qui sont des sous-ensembles de relations, qui peuvent être des unités de répartition.
- **Efficacité**: stockage de données à proximité du lieu de leur utilisation fréquente.
- **Parallélisme**: une transaction globale peut être décomposée en sous-transactions travaillant chacune sous un fragment, donc accroissement du parallélisme.

□ Inconvénients de la fragmentation

- Performances réduites du fait de la répartition des fragments sur les différents sites.
- Intégrité difficile puisque du fait de la fragmentation, les dépendances fonctionnelles sont aussi réparties sur les différents sites.

2008/2009

Alimazighi & Boukhalifa

33

Les Techniques de répartition de données: la fragmentation

La fragmentation

- Fragmenter une relation R, c'est la diviser en un certain nombre de fragments r_1, r_2, \dots, r_n , dont chacun contient des données de R.

Règles de correction

- **Décomposition sans perte**: chaque élément de données appartenant à la relation globale appartient aussi à un ou plusieurs de ses fragments.
- **Reconstruction**: d'une relation globale à partir de ses fragments doit être toujours possible.
- **Non duplication**: impose que les fragments sont disjoints dans le cas de la fragmentation horizontale.

Il existe deux façons différentes de fragmentation :

- fragmentation horizontale
- fragmentation verticale

2008/2009

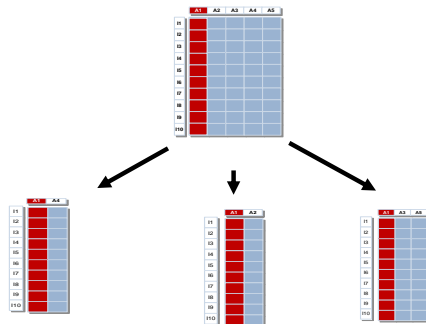
Alimazighi & Boukhalifa

34

Types de fragmentation

Fragmentation verticale

- Décomposer les objets en un ensemble de colonnes (attributs).



35

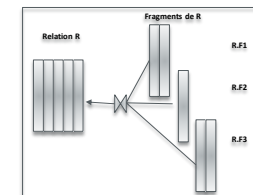
Les Techniques de répartition de données: la fragmentation

❖ La fragmentation verticale

La relation R est décomposée en plusieurs sous-ensembles r_1, r_2, \dots, r_n , tel que :

chaque fragment $r_i = \text{Project}(R / r_i)$

$R = r_1 \text{ join } r_2 \text{ join } r_3 \dots \text{join } r_n$



Remarque : il est nécessaire de rajouter un attribut spécial appelé tuple-identification à la relation R. Il correspond à l'adresse logique ou physique d'un tuple qui sera une clé primaire pour la relation augmentée. De manière générale cet attribut n'est pas visible à l'utilisateur.

2008/2009

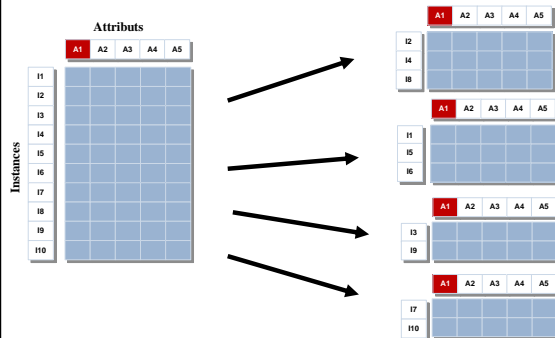
Alimazighi & Boukhalifa

36

Types de fragmentation

Fragmentation horizontale

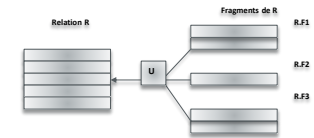
- Décomposer les objets en un ensemble de lignes (instances)



37

Les Techniques de répartition de données: la fragmentation

Fragmentation horizontale



- La relation R est subdivisée en un ensemble de sous relations r_1, r_2, \dots, r_m de même schéma que R.
- Chaque sous-relation se compose d'un certain nombre de tuples de R.
- Chaque tuple de R doit appartenir à l'un des fragments

- Un fragment peut être une sélection sur R : $r_i = \text{Restrict}(R / \pi_i)$

- La reconstruction de R : $R = \text{Union}(r_i) ; (i=1 \text{ à } n)$.

2008/2009

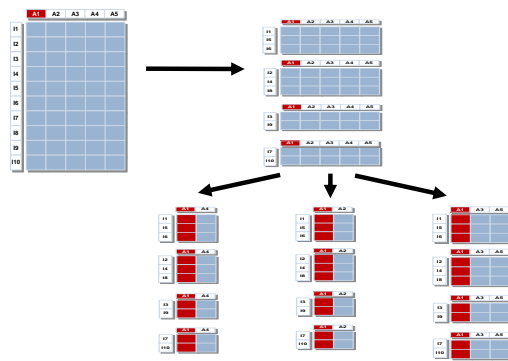
Alimazighi & Boukhalfa

38

Types de fragmentation

Fragmentation mixte

- Horizontale suivie d'une verticale.

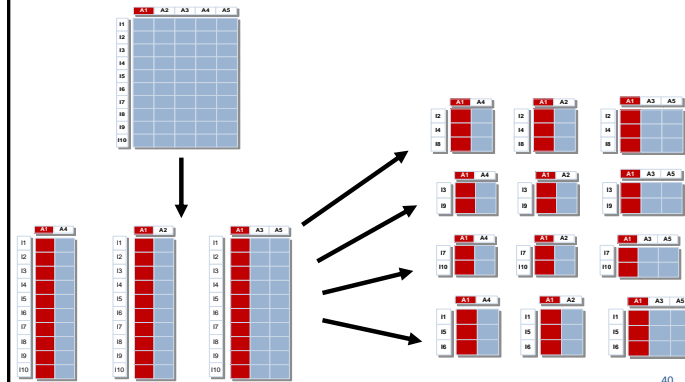


39

Types de fragmentation

Fragmentation mixte

- Verticale suivie d'une horizontale.



40

Fragmentation Horizontale Primaire et dérivée (I)

Fragmentation horizontale primaire (FHP)

- Fragmenter une table en utilisant les prédicats de sélection définis sur cette table

Prédicat : Attribut θ Valeur, $\theta \in \{=, <, \leq, >, \geq\}$ et valeur \in Domaine(Attribut).

Exemple: Client (Client_id, Nom, Ville)

- Client₁ : $\sigma_{\text{Ville}='Alger'}(\text{Client})$
- Client₂ : $\sigma_{\text{Ville}='Béchar'}(\text{Client})$

Impact de la FHP sur les requêtes

- Optimisation des sélections

Select Nom, Age
From Client
Where
Ville=Béchar



Select Nom, Age
From Client2

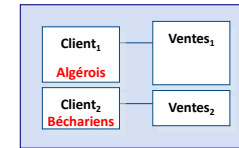
41

Fragmentation Horizontale Primaire et dérivée (II)

Fragmentation horizontale dérivée (FHD)

- Fragmenter une table (S) selon des attributs d'une autre table (T) : (existence de lien entre S et T)

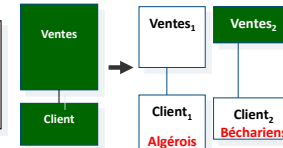
- Ventes(Client_id, Produit_id, Date, Montant)
- Ventes₁=Ventes \Leftrightarrow Client₁
- Ventes₂=Ventes \Leftrightarrow Client₂



Impact de la FHD sur les requêtes

- Optimisation de la jointure entre S et T

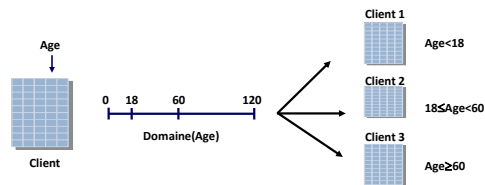
Select sum(Montant)
From Client C, Ventes
V
Where C.CID=V.CID
and Ville=Béchar



Select sum(Montant)
From Ventes2

42

Fragmentation par intervalle « Range »

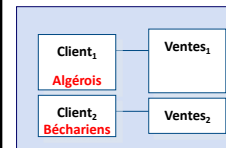


```
CREATE TABLE Client
(Client_id NUMBER(5),
Nom Varchar2(20),
Ville Varchar2(20),
Age Number(3),
Genre Varchar2(1))
PARTITION BY RANGE (Age)
(
PARTITION Client_Moins_18 VALUES LESS THAN (18) TABLESPACE TBSPMoins27,
PARTITION Client_18_59 VALUES LESS THAN (60) TABLESPACE TBSP27-59,
PARTITION Client_60_Et_Plus VALUES LESS THAN (MAXVALUES) TABLESPACE TBSPPlus60
);
```

43

Fragmentation dérivée par le mode Référence

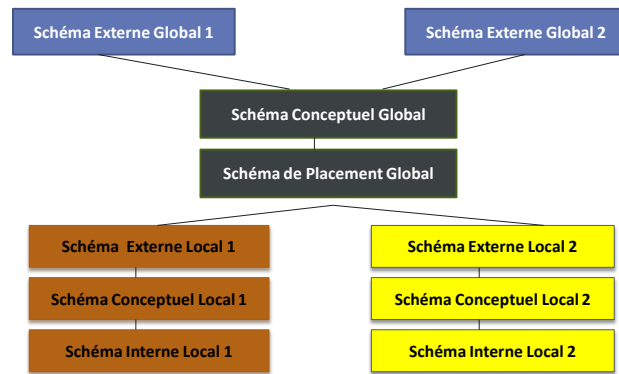
- Fragmenter une table selon le schéma de fragmentation d'une autre table en utilisant le lien par clé étrangère.



```
CREATE TABLE Ventes
(Client_id NUMBER(5),
Time_id NUMBER(5)),
Montant NUMBER(20),
CONSTRAINT order_items_fk
FOREIGN KEY(Client_id) REFERENCES Client(Client_id)
)
PARTITION BY REFERENCE(order_items_fk);
```

44

Architecture de référence



Alimazighi & Boukhalfa

45

Architecture de référence

□ Schéma conceptuel global

Il contient la description unifiée de la base de données répartie indépendamment de l'environnement réparti. Il est défini pour être la synthèse des besoins de l'ensemble des applications.

□ Schéma de Placement

- Nécessité de la définition d'un **dictionnaire de données** qui contient les informations concernant la localisation provenant de la répartition (fragmentation et duplication), on l'appelle aussi **schéma de placement**.
- Un dictionnaire est dit **global** lorsqu'il contient des informations sur tous les sites.
- Un dictionnaire est dit **local** lorsqu'il contient uniquement des informations d'un site. Dans ce cas, le SGBD devra s'appuyer sur les dictionnaires des autres sites pour localiser une information.

Alimazighi & Boukhalfa

46

Architecture de référence

- **Schéma externe global** : vue d'un groupe d'utilisateurs sur la BD répartie.
- **Schéma interne local et schéma conceptuel local** : identiques à ceux d'une base de données centralisées.
- **Schéma local externe** : vue de la base de données locales. (ensemble de fragments de relations)

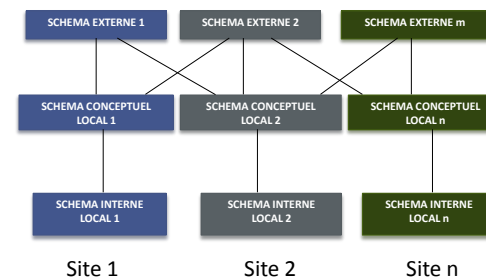
Alimazighi & Boukhalfa

47

Architecture de référence

- L'existence de schéma conceptuel global permet de regrouper toutes les structures de données afin de les maintenir de façon centralisée. Lorsque des BD existent déjà et qu'il va falloir mettre en relation, il n'est pas nécessaire de regrouper toutes les structures de données.

• Architecture sans schéma global



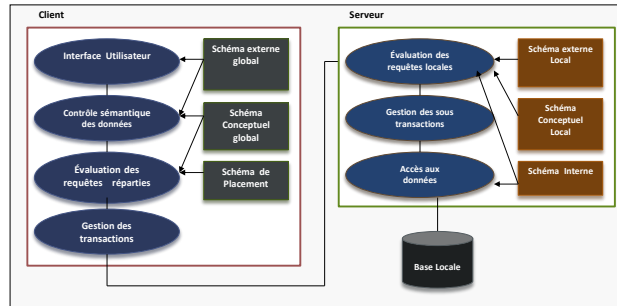
Alimazighi & Boukhalfa

48

ARCHITECTURE FONCTIONNELLE BASEE SUR UNE APPROCHE CLIENT SERVEUR

Deux composants systèmes :

- ❑ Gérant d'applications sur le site (Client)
- ❑ Gérant de données sur le site (Serveur)



Alimazighi & Boukhalfa

49

ARCHITECTURE FONCTIONNELLE BASEE SUR UNE APPROCHE CLIENT SERVEUR

Site Client

❑ Interface utilisateur

- vérification de la conformité de la requête répartie avec le schéma externe.
- Restitution des résultats après exécution de la requête

❑ Contrôle sémantique des données

- vérification des contraintes d'intégrité sémantique
- contrôle d'autorisation

Alimazighi & Boukhalfa

50

ARCHITECTURE FONCTIONNELLE BASEE SUR UNE APPROCHE CLIENT SERVEUR

❑ Évaluation de requêtes réparties

- A chaque requête répartie est construit un plan d'exécution réparti.
- Un plan d'exécution réparti est un programme découvrant la stratégie d'exécution optimisée d'une requête répartie. Une requête globale est décomposée en un ensemble de sous-requêtes locales, (génération d'un arbre syntaxique optimisé).
- **Optimisation** : choisir la meilleure localisation des copies de fragments dans le dictionnaire de données et ordonnancer les requêtes locales de façon optimisée.

• Gestion de transactions

- assure le contrôle de concurrence réparti
- validation des transactions
- reprise de transactions

Alimazighi & Boukhalfa

51

ARCHITECTURE FONCTIONNELLE BASEE SUR UNE APPROCHE CLIENT SERVEUR

Site serveur : Gérant des données

❑ Évaluation de requêtes locales

- Vérification que les requêtes locales sont conformes au schéma externe local.
- Génération du plan d'exécution de la requête locale

❑ Gestionnaire de sous-transactions

- Synchronise les sous-transactions d'une même transaction globale
- Validation des sous-transactions

❑ Accès aux données

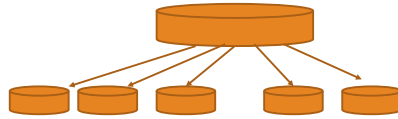
- fonctions normales de gestion de données : gestion de la mémoire et des méthodes d'accès aux données du site.

Alimazighi & Boukhalfa

52

CLASSIFICATION DES APPROCHES DE CONCEPTION D'UNE BDR

❑ Approche descendante



- ❑ Répartition d'une BD centralisée.
- ❑ Approche de décomposition descendante d'une BD centralisée en une collection de données appartenant **logiquement** au même système mais physiquement étalée sur plusieurs sites. La BDR est dite **homogène**.
- ❑ Définition d'une stratégie de duplication et fragmentation
- ❑ La description globale est nécessaire
- ❑ **Exemple** : R* de IBM (1979,1984). Ce SGBD réparti consiste dans plusieurs sites autonomes dotés du même SGBD.

Alimazighi & Boukhalfa

53

CLASSIFICATION DES APPROCHES DE CONCEPTION D'UNE BDR

La démarche descendante comprend trois étapes

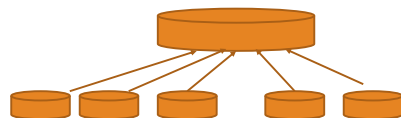
- la conception
- la fragmentation
- l'allocation
- **Avantages**
 - favorise la croissance incrémentale en facilitant l'ajout d'un nouveau site dans le réseau et convient à une nouvelle BD
- **Inconvénient**
 - une modification du schéma global peut entraîner la reprise tout le processus de conception

Alimazighi & Boukhalfa

54

CLASSIFICATION DES APPROCHES DE CONCEPTION D'UNE BDR

❑ Approche ascendante



BD virtuelle

- ❑ Prise en compte d'un ensemble de BD existantes pour construire la BDR : intégration de BD en vue d'une seule dite **multibase** de données.
- ❑ La multibase est dite homogène si les SGBD gérant les BD locales sont les mêmes.
- ❑ La multibase est dite hétérogène dans le cas contraire.

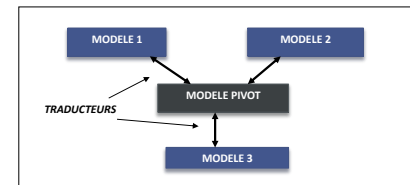
Alimazighi & Boukhalfa

55

CLASSIFICATION DES APPROCHES DE CONCEPTION D'UNE BDR

La démarche ascendante comprend les étapes suivantes :

- sélection d'un modèle de données commun pour la description des schémas au niveau global appelé **modèle pivot** ou canonique.
- Transformation de chaque schéma local dans le modèle pivot.
- Intégration de l'ensemble des schémas locaux.



Alimazighi & Boukhalfa

56

LES ETAPES DE DEVELOPPEMENT D'UNE FEDERATION

ETAPE 1 : Translation

- Construction du schéma composant à partir du schéma local en utilisant le modèle pivot.

ETAPE 2 : Définition

- Définition des schémas d'exportation à partir du schéma composant. C'est l'étape qui permet la construction des schémas d'exportation à partir des données que le gestionnaire de la fédération souhaite faire partager.

ETAPE 3 : Intégration

- Intégration des schémas d'exportation en un ou plusieurs schémas fédérés. C'est cette étape qui réalise la coopération entre les SGBD locaux.

ETAPE 4 : Attribution

- Définition des schémas externes pour chaque utilisateur ou chaque classe d'utilisateurs.

Alimazighi & Boukhalfa

57

Synthèse

- Un système fédéré ne saurait s'envisager sans résoudre les problèmes de translation qui se posent lors du transfert de données d'un contexte local à un contexte global.
- Le principe général d'une telle translation repose sur la définition des caractéristiques intrinsèques des modèles et de développer ensuite des règles de correspondance permettant de générer un schéma composant dans le modèle pivot à partir d'un schéma source exprimé dans le modèle local.
-

Alimazighi & Boukhalfa

58

MODELE COMMUN ET TRANSFORMATION DE SCHEMAS

Une BD fédérée supporte deux classes d'utilisateurs

- Les utilisateurs de la fédération qui manipulent les données par l'intermédiaire des schémas fédérés
- Les utilisateurs locaux qui accèdent uniquement aux données locales
- La conception de BDR fédérés hétérogènes est basée sur la notion de translation de schémas pour permettre la transformation depuis les structures de données depuis un modèle vers les structures de données d'un autre modèle appelé modèle commun.

Alimazighi & Boukhalfa

59

CRITERES DE CHOIX DU MODELE COMMUN DE DONNEES

Objectifs d'un modèle commun

- Capturer les capacités structurelles et sémantiques des modèles locaux.
- Faciliter la communication entre utilisateurs des systèmes divers et parfois incompatibles.

Critères de choix

- **Complétude** : le modèle pivot doit permettre de décrire toutes les notions conceptuelles des autres modèles de données.
- **Facilité de traduction** : le modèle pivot doit pouvoir faciliter la translation entre le schéma local et le schéma composant.
- **Évolution du schéma** : Le modèle de données commun doit pouvoir assurer l'évolution d'un schéma avec un minimum de remise en cause de l'existant.
- **Expression de la répartition** : le modèle pivot doit permettre de décrire la répartition de données

Alimazighi & Boukhalfa

60

TRANSFORMATION DE SCHEMAS

- **Principe général** : développement de langages de description permettant de définir les opérations à effectuer pour passer des valeurs d'une caractéristique à une autre valeur.

Premier modèle pivot : le modèle relationnel

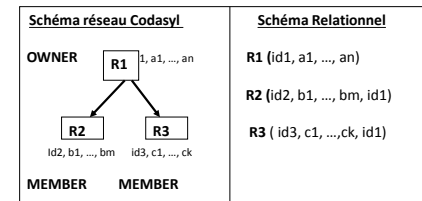
- Les justificatifs étant les suivants :
- facilité de composition et de recomposition de relations (fragmentation facile)
- la représentation uniforme des attributs permet de retirer des informations sur la base de valeurs d'attributs communs.
- Les translations peuvent être exprimées dans les langages de manipulation rendant la décomposition de requêtes assez directe.

Alimazighi & Boukhalifa

61

TRANSFORMATION DE SCHEMAS

Application



Alimazighi & Boukhalifa

62

IMPACT DE L'INTRODUCTION DES CONCEPTS OBJET DANS LE DEVELOPPEMENT DES BDR HETEROGENES

Pourquoi un modèle objet comme modèle commun ?

□ Un modèle de données commun doit :

- avoir une puissance de représentation au moins égale à celle des modèles natifs des schémas locaux,
- offrir une formalisation puissante qui doit permettre d'enrichir les schémas locaux.

Comme les sources d'information sont dispersées dans un système fédéré, et dans l'optique actuelle, les informations sont complexes (multimédias), il devient intéressant d'augmenter les possibilités d'expression et de capturer la dynamique des applications distribuées.

Alimazighi & Boukhalifa

63

IMPACT DE L'INTRODUCTION DES CONCEPTS OBJET DANS LE DEVELOPPEMENT DES BDR HETEROGENES

□ Le modèle relationnel

- Le modèle relationnel ne prévoit aucune extension pour représenter les hiérarchies de types (données simples).
- Dans les systèmes relationnels, les tuples ne peuvent être distingués que sur la base des valeurs de leurs attributs (états).
- Manque de perception globale : définition de beaucoup de relations pour représenter les entités et leurs liens
- La manipulation des entités nécessitent beaucoup de jointures
- Les modèles sémantiques ont été utilisés avec succès (modèle fonctionnel ou modèle E/A et E/A étendu). Étant des modèles statiques, leur insuffisance a été leur limite dans l'expression du comportement.



Le modèle OBJET : encapsulation du comportement dans des méthodes accompagnant les données

Alimazighi & Boukhalifa

64

IMPACT DE L'INTRODUCTION DES CONCEPTS OBJET DANS LE DEVELOPPEMENT DES BDR HETEROGENES

Règles de correspondance avec le modèle objet

- **Règle 1** : Transformer chaque entité de base qui peut être soit une relation, soit un segment, soit un record en une classe d'objets portant le même nom.
- **Règle 2** : A chaque constituant simple d'une entité de base correspond une propriété de composition simple du modèle objet.
- **Règle 3** : A chaque propriété de composition multiple, définir une classe objet séparée reliée à la classe supérieure par un lien de composition.
- **Règle 4** : A une association spécifiant un lien binaire non porteur de propriétés :
 - pour une cardinalité 1 :N : créer un lien de référence multiple au niveau de la classe d'objet participante référençante.
 - pour une cardinalité M :N créer un lien de référence multiple au niveau de chaque classe d'objets participante.

Alimazighi & Boukhalifa

65

IMPACT DE L'INTRODUCTION DES CONCEPTS OBJET DANS LE DEVELOPPEMENT DES BDR HETEROGENES

- **Règle 5** : Pour un lien d'association autre que celui spécifié dans la règle précédente, créer une nouvelle classe d'objet ayant comme propriétés les propriétés du lien et des liens de référence référençant toutes les classes d'objets participantes.
- **Règle 6** : Transformer un lien de sous-typage par un lien de généralisation au niveau de la classe spécialisée.

Alimazighi & Boukhalifa

66