

Base de données réparties

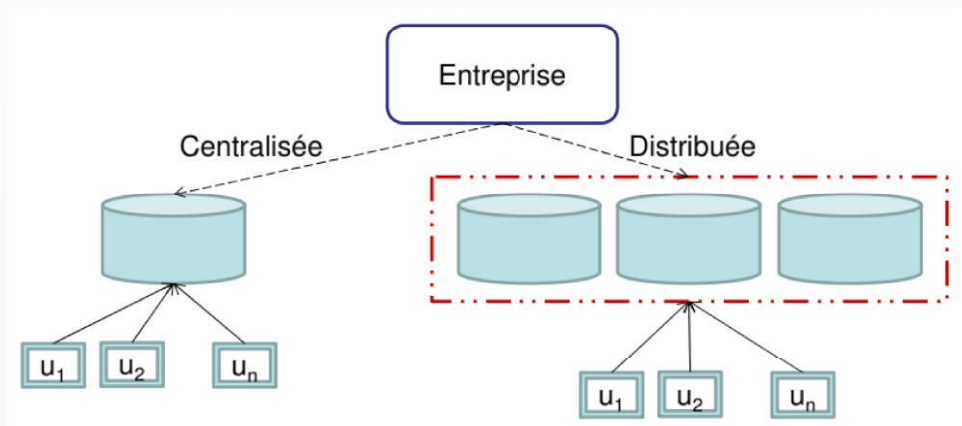
Pr. EL KABTANE Hamada

Plan du cours

- ☐ Définition d'une BDR
- ☐ Principe fondamental d'une BDR
- ☐ Architecture des schémas
- ☐ Conception de BDR
- ☐ Fragmentation répartie
- ☐ Duplication répartie (Synchrone et asynchrone)
- ☐ Transactions réparties
- ☐ Contrôle de concurrence réparti
- ☐ Optimisation de requête répartie

Introduction

Une entreprise, selon son besoin, peut choisir entre l'utilisation de base de données centralisée ou distribuée.



3

Introduction

Quelques définitions

- Une base de données distribuée est un ensemble de bases de données localisées et gérées par des sites différents et apparaissant à l'utilisateur comme une base unique.
- Une base de données distribuée est une collection de multiples bases de données distribuées sur un réseau informatique logiquement inter-reliées.
- Un client d'un système de gestion de base de données distribuée est une application qui accède aux informations distribuées par les interfaces du système.

4

Introduction

Quelques définitions (suite)

- Un système de gestion de base de données distribuées est un système qui gère des collections de BD logiquement reliées, distribuées sur un réseau, en fournissant un mécanisme d'accès qui rend la répartition transparente aux utilisateurs.
- Un système de gestion de bases de données distribuées (Distributed DBMS) est l'application qui permet la gestion de la base de données distribuée et rend la distribution transparente aux utilisateurs.

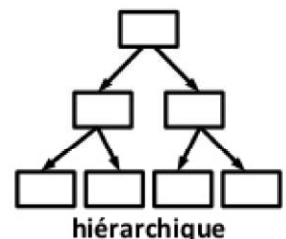
Pour référencer la BDD distribuée et le DBMS distribués nous utilisons l'abréviation DDBS (Distributed DataBase System).

5

Types des bases de données

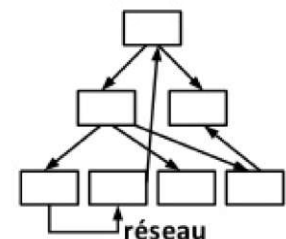
Base hiérarchique

Lie les enregistrements dans une structure arborescente où chaque enregistrement n'a qu'un seul possesseur.



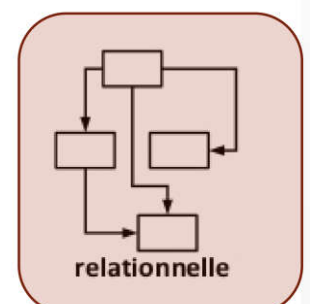
Base en réseau

Est une base hiérarchique mais permet en plus d'établir des relations transverses.



Base relationnelle

Stocke les informations décomposées et organisées dans des matrices appelées relations ou tables.



Base objet

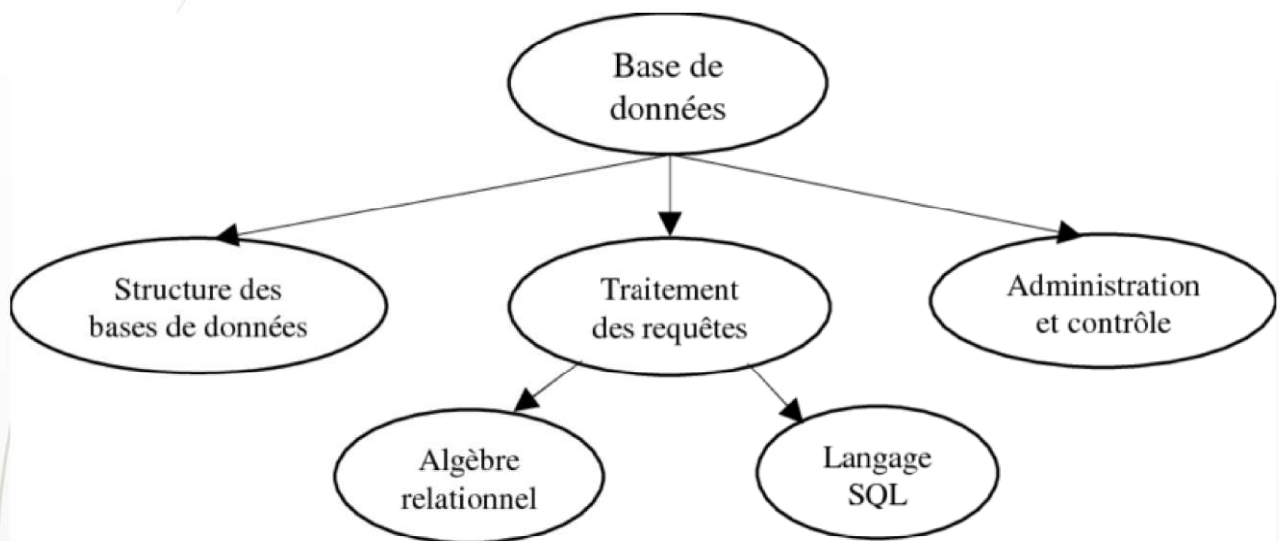
Stocke les informations groupées sous forme de collections d'objets persistants.

Base XML

S'appuie sur le modèle de données fourni par XML.

6

Bases de données relationnelles



7

Bases de données réparties (BDR)

➡ Différents niveaux de répartition

- **Données**
- **Schémas** de la base de données
- **SGBD**
- **Traitement (requêtes, transactions)**
- **Composants matériels:** mémoires, disques, ...

8

Terminologies

➤ BD répartie (**distributed database**) = BDR = BD + Réseau

- Ensemble de **BDs gérées par des sites différents** et qui apparaissent à l'utilisateur comme une base unique.

« To the user, a distributed system should look exactly like a non distributed system »

(Chris. Date, Introduction to Database Systems)

➤ SGBD Réparti (SGBDR)

- Logiciel qui gère une BDR et qui rend la répartition transparente

➤ Client de SGBDR

- Application qui accède aux informations distribuées par les interfaces du SGBDR

9

Applications

➤ Cas de **grosses entreprises** ou **organismes** ayant des agences géographiquement distribuées:

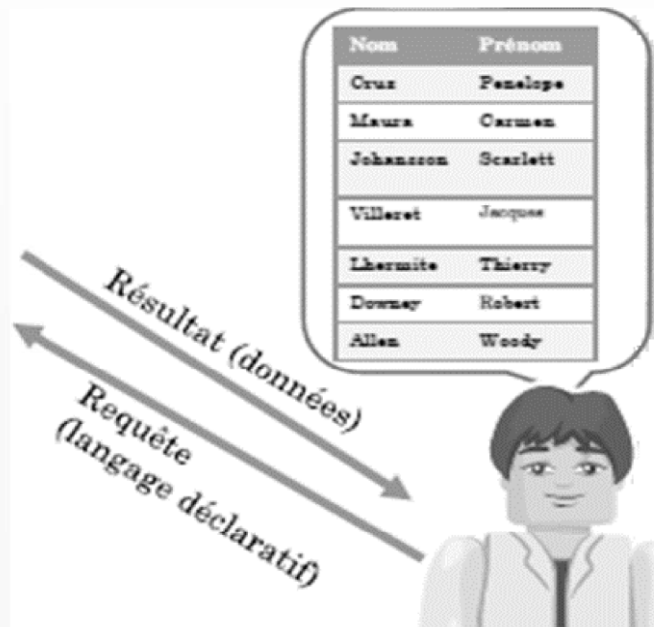
- Banques
- Fabrication
- Médicales (BD biologiques)
- Militaires
- Systèmes de réservation de compagnies aériennes
- WEB

10

JUSQU'À PRÉSENT ...

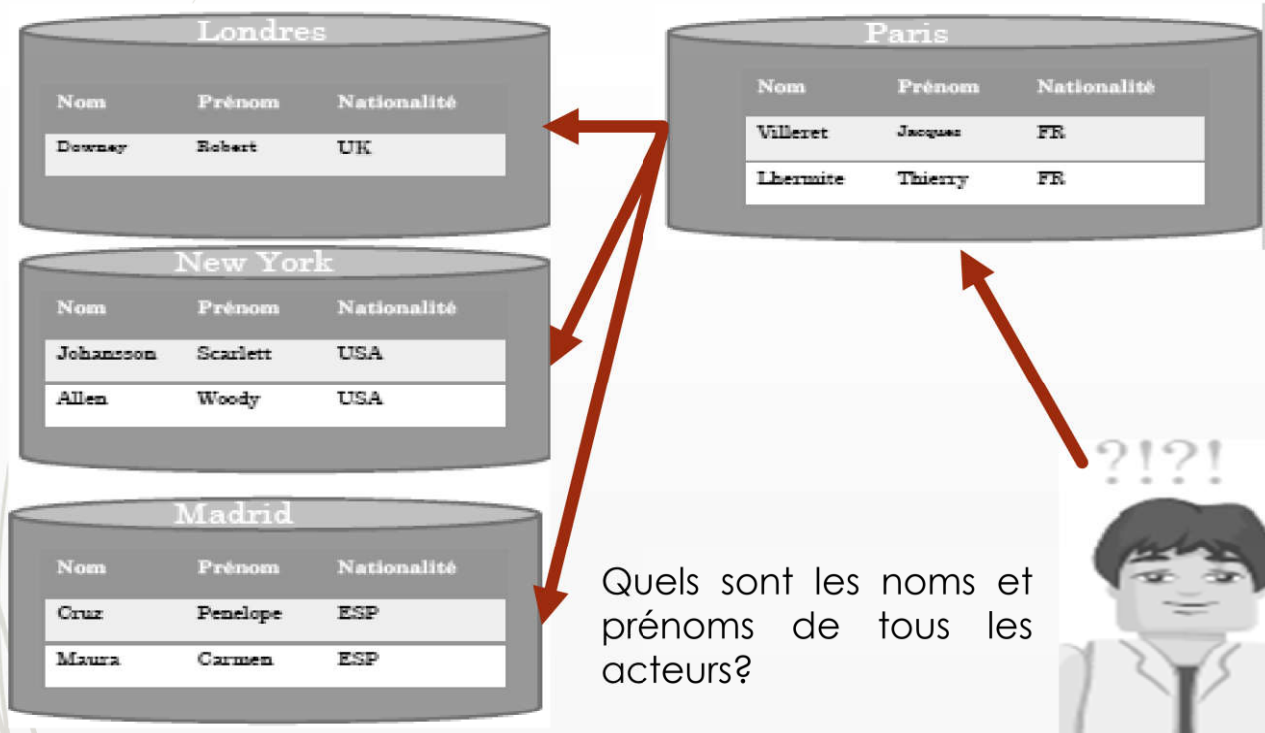
Quels sont les noms et
prénoms de tous les acteurs?

BD Relationnelle		
Acteur		
Nom	Prénom	Nationalité
Cruz	Penelope	ESP
Maura	Carmen	ESP
Johansson	Scarlett	USA
Villeret	Jacques	FR
Lhermite	Thierry	FR
Downey	Robert	UK
Allen	Woody	USA



11

ET MAINTENANT ?!?

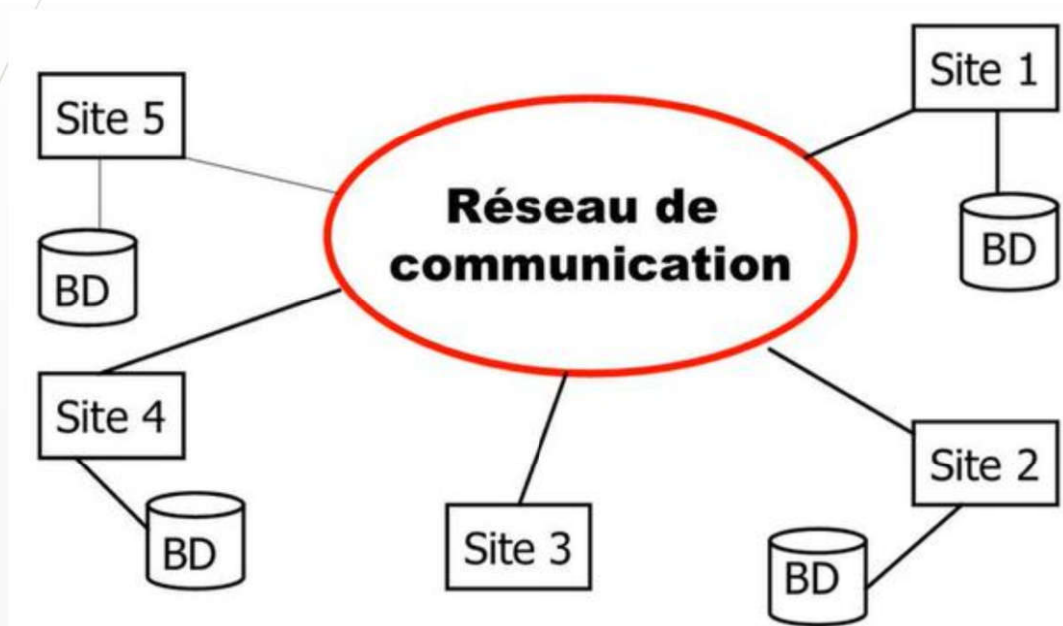


Quels sont les noms et
prénoms de tous les
acteurs?

12

DÉFINITION DU CONTEXTE

- Un ensemble de données stockées sur plusieurs sites (machines) communiquant *via un réseau*



13

Problématique et Objectifs

- **Exemple:**

- Soit une entreprise qui a des annexes dans plusieurs villes: Marrakech, Casablanca, Agadir Le gérant veut maintenir une base de données des employés et les projets sur lesquels ils travaillent.

Le gérant utilise 4 tables:

- Emp (idEmp, nomEmp, poste);
- Prj (idPrj, nomPrj, designation, budget, ville);
- Sal (poste, salaire);
- Aff (idEmp, idPrj, resp, duree) //Affectation d'un employé à un projet avec telle responsabilité et telle durée.

Si la base de données est centralisée, on peut trouver les noms des employés qui travaillent dans un projet pour une durée supérieure de 12 mois ainsi que leurs salaires avec la requête suivante:

```
SELECT nomEmp, salaire
FROM Emp, Aff, Sal
WHERE Aff.duree > 12
AND Emp.idEmp = Aff.idEmp
AND Sal.poste = Emp.poste
```

14

Problématique et Objectifs

- **Exemple (suite):**

Maintenant, si la base est distribuée sur les différents annexes, de telle sorte que les informations des employés et les projets de chaque annexe sont stockés dans une base de données localisée dans cette annexe. L'utilisateur doit poser la même question au système sans faire attention à la distribution des informations (transparence), et c'est le système qui s'occupe de la récupération des informations.

- **Problématique:**

- Comment offrir aux utilisateurs un accès à l'ensemble des données en **faisant abstraction de leur localisation physique** tout en garantissant un traitement efficace des requêtes et le respect de contraintes (physiques) géographiques ?

15

Problématique et Objectifs

- **Objectifs:**

- Autonomie locale
- Transparence
- Performance améliorée
- Fiabilité et disponibilité accrues
- Partage accru de données et ressources
- Expansion graduelle

16

Problématique et Objectifs

- Pour cela, il a été défini plusieurs types de transparence:

1/ Transparence de données (Indépendance)

La définition des données dans une base de données se fait à deux niveaux (structure logique et structure physique), se qui permet de définir une indépendance de données logique et indépendance de données physique.

L'indépendance de données logique signifie que les applications de l'utilisateur ne sont pas autorisées à modifier la structure logique des données (le schéma).

L'indépendance de données physique signifie que les détails de la structure de stockage des données seront cachés aux applications de l'utilisateur.

17

Problématique et Objectifs

2/ Transparence du réseau (ou de distribution)

Ce type de transparence fait référence à ce que l'existence du réseau est caché à l'utilisateur. L'accès aux données de la base doit se faire de la même manière soit on mode centralisé ou distribué.

18

Problématique et Objectifs

3/ Transparence de réplication (redondance)

Pour des raisons de performance et disponibilité des données, il est préférable de faire des copies de quelques données de la base sur plusieurs sites.

Par exemple, les données qui se trouvent sur un autre site et qui sont réclamées souvent par un utilisateur, sera préférable d'en faire une copie sur la base de données local de cet utilisateur.

En plus, si une machine s'arrête, on peut trouver les données sur une autre machine.

La décision de réplication des données, et de combien de copies doit-on faire, dépend des applications des utilisateurs. L'utilisateur n'a pas besoin de savoir si les données sont répliquées ou non. Il envoie ses requêtes comme s'il existe une seule copie des données.

19

Problématique et Objectifs

3/ Transparence de fragmentation

Dans certains cas, et pour des raisons de performance, on aura besoin de fragmenter une table de la base de données et placer chaque fragment sur un site différent.

La fragmentation permet aussi de réduire les inconvénients de la réplication.

Cette opération de fragmentation doit être faite d'une manière transparente à l'utilisateur. Il utilise la table originale comme si elle n'est pas fragmentée.

Il existe deux types de fragmentation:

- Fragmentation horizontale: La table est partitionnée en plusieurs sous-tables contenant chacune un sous-ensemble de lignes de la table originale.
- Fragmentation verticale: La table est partitionnée en plusieurs sous-tables qui sont définies sur un sous-ensemble d'attributs (de colonnes) de la table originale.

20

Défis des BDR

- **Conception d'une BDR:**

- Fragmentation
- Allocation
- Réplication (totale ou partielle)

- **Transparence à la répartition:**

- Extension de la notion d'indépendance logique et physique des données.
- Location (réplication, fragmentation): Aucune spécification de la localisation des données.

21

Défis des BDR

- **Optimisation de requêtes réparties:**

- Choix de la copie en lecture
- Mise à jour de toutes les copies
- Plan d'exécution répartie

- **Transaction réparties:**

- Utilisateur aura à formuler ses transactions de la même manière que dans un environnement centralisé.
- Maintien des propriétés ACID (atomicité, cohérence, isolation et durabilité) des transactions.

22

ACID

- **Atomicité:**

- La propriété d'atomicité assure qu'une transaction se fait au complet ou pas du tout : si une partie d'une transaction ne peut être faite, il faut effacer toute trace de la transaction et remettre les données dans l'état où elles étaient avant la transaction.

- **Cohérence:**

- La propriété de cohérence assure que chaque transaction amènera le système d'un état valide à un autre état valide.

23

ACID

- **Isolation:**

- Toute transaction doit s'exécuter comme si elle était la seule sur le système.

- **Durabilité:**

- La propriété de durabilité assure que lorsqu'une transaction a été confirmée, elle demeure enregistrée même à la suite d'une panne d'électricité, d'une panne de l'ordinateur ou d'un autre problème.

24

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.