

# ECOLE SUPERIEURE DE GESTION D'INFORMATIQUE ET DES SCIENCES

LOME : B.P. 13517 Tél. (228) 22 20 12 93 Tél/Fax 22 22 50 17

COTONOU : 06 B.P. 3233 Tél. 21 32 47 73 / 21 00 02 26 / 21 04 04 16

E-mail : esgis.togo@gmail.com / esgis.benin@gmail.com Site : <http://www.esgis.tg>

## SUPPORT DE COURS

# *INTRODUCTION AUX BASES DE DONNEES DISTRIBUEES*

**Niveau** : Licence 3 IRT (Informatique, Réseau, Télécommunication)

**Option** : AL

**Dispensé par** : N'SOUGAN Folly W.

**Année académique 2022-2023**

# PLAN

---

1. Besoins – Objectifs - Définitions
2. Conception des Bases de données Distribuées
3. Fragmentation et Allocation

# Thèmes des exposés

---

1. Réplication MYSQL
2. Réplication POSTGRESQL
3. Réplication SQL SERVER
4. Réplication entre 2 SGBD

# **Chapitre 1 :**

# **BESOINS OBJECTIFS DEFINITIONS**

# BESOINS OBJECTIFS DEFINITIONS

1. Définitions
2. Historique donnant naissance à la réplication
3. Problématique
4. Buts de la distribution des données
5. Avantages et inconvénients de BD distribuées
6. SGBD Distribuées
7. Objectifs

### 1. Définitions

#### a) Base de données

L'expression «base de données» est apparue vers 1964 pour désigner une collection d'informations partagées par différents utilisateurs d'un système d'informations militaire. En 1970, Edgar F. Codd apporte une contribution substantielle à cette définition en y introduisant les concepts de l'algèbre relationnelle donnant lieu aux «bases de données relationnelles». Il convient d'affirmer sans équivoque, que cette notion de bases de données relationnelles a été un apport significatif dans le domaine informatique, en permettant d'organiser efficacement l'information de l'entreprise.

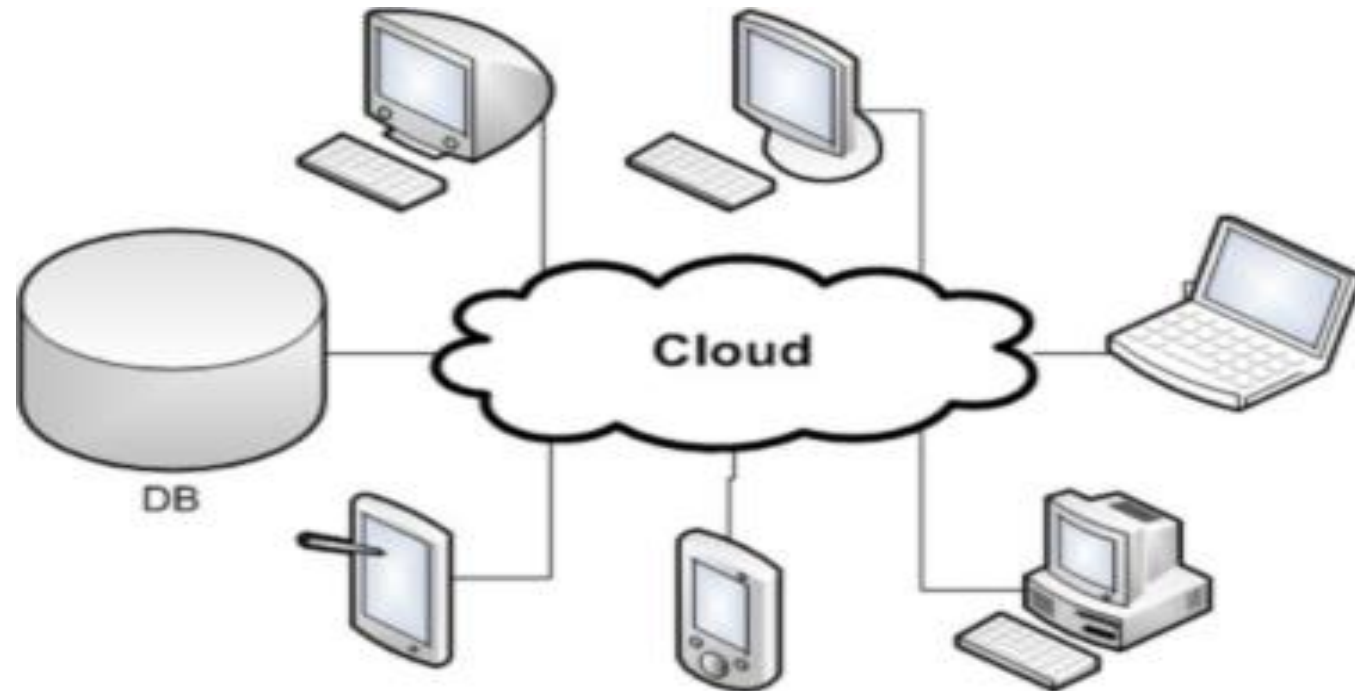
**D'une manière générale, une base de données** est un dépôt informatique qui permet de sauvegarder l'information de façon persistante en format tabulaire conformément au modèle de données basé sur les règles de l'algèbre relationnelle.

## BESOINS OBJECTIFS DEFINITIONS

### Avantages de Base de données

- La capacité de *partager des données*.

Ceci signifie que plusieurs utilisateurs peuvent se servir simultanément des informations contenues dans la base de données en accédant tous à la même information et en l'utilisant chacun à des fins différentes.



### Avantages de Base de données

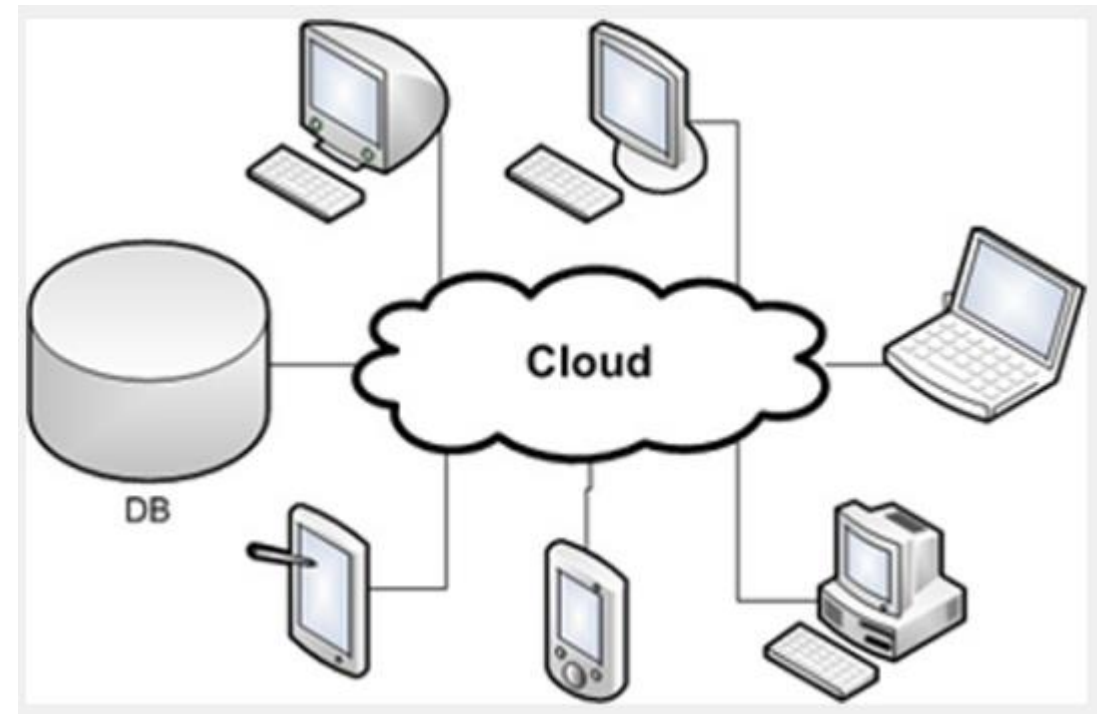
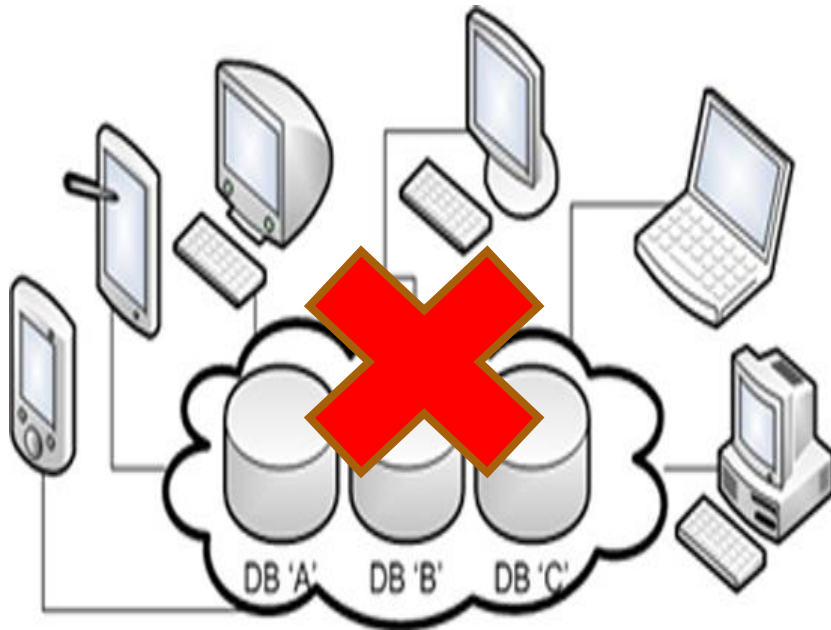
- Les bases de données permettent d'éviter les **redondances de données**. La redondance est une situation où les données sont répétées inutilement.
- On peut aussi éviter les **incohérences des données**. Le terme incohérence signifie que deux ou plusieurs données dans la base de données contiennent des valeurs différentes. D'habitude, cette incohérence est le résultat de la mise à jour de données redondantes.
- Disponibilité d'une **structuration des données**. La base de données de l'entreprise doit à tout moment refléter la réalité le plus fidèlement possible, car cela aide à comprendre sa structure et sa fonction. La conception de la base de données est indispensable pour cette raison, et il est évident que la base de données doit montrer la réalité, pas le contraire.
- **L'accès effectif aux données**: La base de données peut être parfaitement conçue et fonctionne de façon optimale, mais le vrai potentiel des bases de données est la possibilité de faire des requêtes qui peuvent combiner l'information de deux ou plusieurs tables.
- **La sécurité**, c'est-à-dire, la possibilité que l'administrateur applique des restrictions en matière de sécurité à différents aspects de la base de données. Par exemple, il peut définir des contrôles d'accès et des restrictions pour chaque utilisateur de l'information de la base de données.



## BESOINS OBJECTIFS DEFINITIONS

### b) Bases de données centralisée

C'est une ensemble de données ayant un corps unifié, un seul SGBD, un seul stockage physique et une seule unité de traitement . Le SGBD se charge de stoker et de partager ces informations.



### b) Bases de données centralisée

#### **Avantages :**

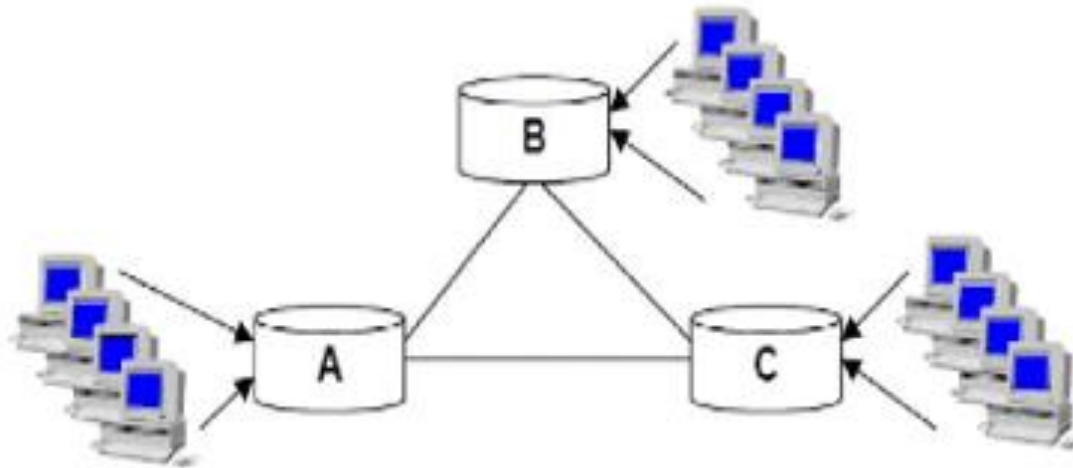
- Étant donné que toutes les données sont stockées à un seul endroit, il est donc plus facile d'accéder aux données et de les coordonner.
- La base de données centralisée a une redondance de données très minimale puisque toutes les données sont stockées en un seul endroit.
- C'est moins cher par rapport à toutes les autres bases de données disponibles.

#### **Inconvénients:**

- Le trafic de données dans le cas d'une base de données centralisée est élevé.
- Si une défaillance du système se produit dans le système centralisé, toutes les données seront détruites.

### c) Bases de données distribuée

*Une base de données distribuée* est un schéma global d'une collection de base de données logiquement reliés et réparties sur plusieurs sites en fournissant un moyen d'accès rendant la distribution transparente à l'utilisateur et donc visible comme un tout.



### c) Bases de données distribuée

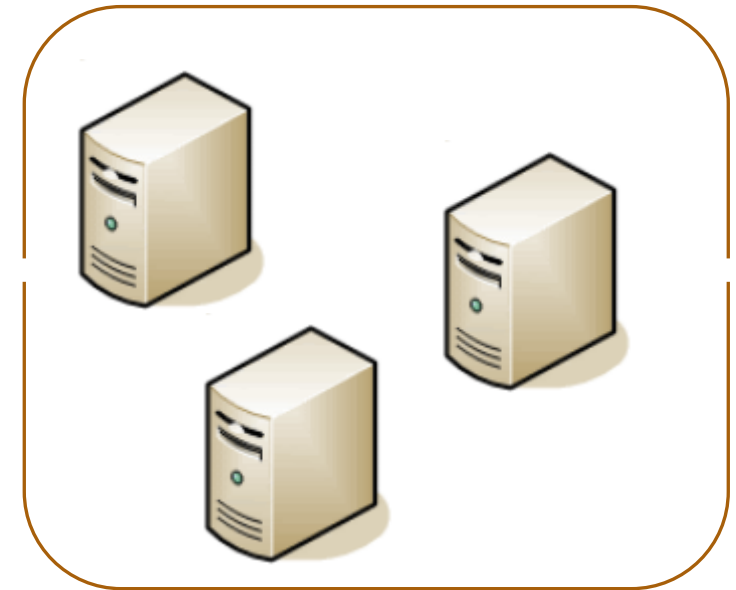
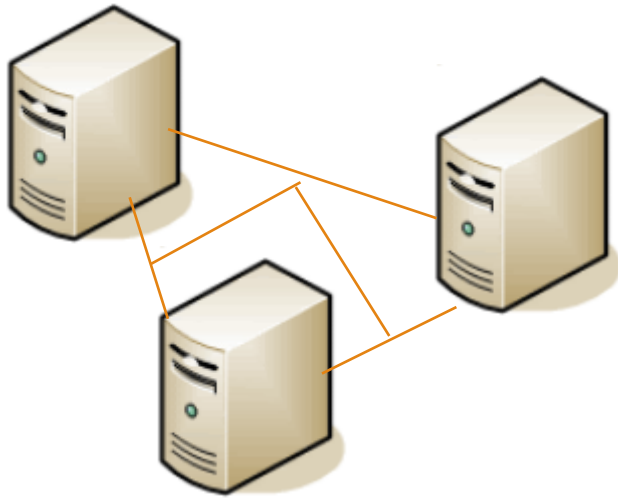
Tous les nœuds contiennent des informations. Les clients sont égaux et ont des droits égaux.

***Un système de gestion de base de données distribuée (SGBDD)*** est une application centrale qui administre une base de données distribuée comme si toutes les données étaient stockées sur le même ordinateur. Ce système synchronise toutes les données à des moments donnés et lorsque plusieurs utilisateurs souhaitent accéder à la même donnée, il s'assure que les mises à jour ( les insertions, modifications et suppressions) opérées sur la donnée à un endroit, sont automatiquement répliquées aux autres endroits de stockage.

**Base de données décentralisée :** Cela signifie qu'il n'y a pas de stockage central. Certains serveurs fournissent des informations aux clients. Les serveurs sont connectés les uns aux autres.

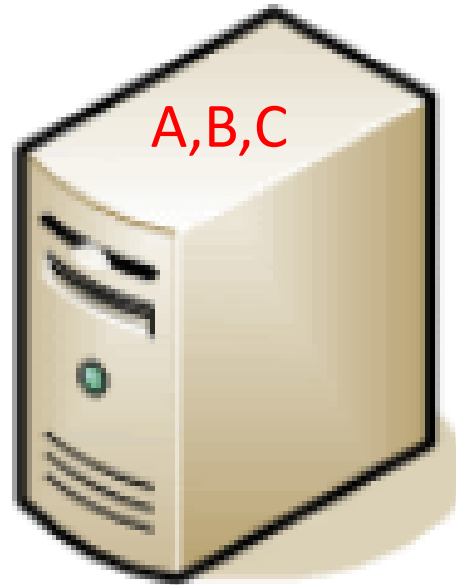
### N.B :

Une base de données distribuée est une base de données logique unique, installée sur un ensemble d'ordinateurs géographiquement situés à différents emplacements et *reliés par un réseau de communication* de données, tandis qu'une base de données décentralisée est une base de données installée sur des systèmes géographiquement situés à des emplacements différents, mais *non liés via* un réseau de communication de données.



### d) Homogénéité des Bases de données distribuée

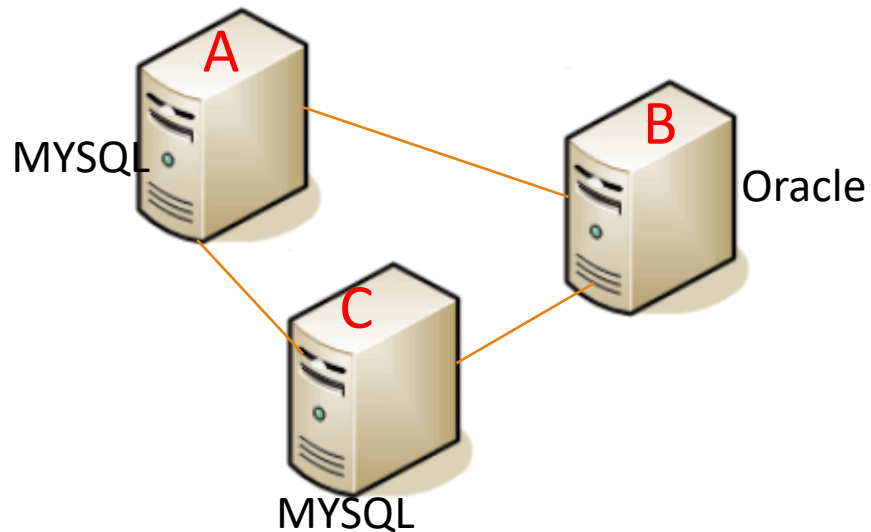
Une base de données distribuée **homogène** est caractérisée par le fait que toutes les instances ou nœuds démarrent sur un seul logiciel et partagent un seul matériel informatique (hardware); de plus, le système peut être affiché dans une interface unique comme s'il s'agissait d'une seule base de données.



A: BD Maritime  
B: BD PLATEAUX  
C: BD SAVANES

### d) Homogénéité des Bases de données distribuée

Une base de données distribuée **hétérogène** peut avoir différents matériels informatiques (hardware), systèmes d'exploitation, SGBD, et différents schémas de base de données (la structure décrite dans un langage formel qui soutient le modèle géré par le SGBD).



A: BD Maritime  
B: BD PLATEAUX  
C: BD SAVANES

### e) Autres Types de base de données

Les bases de données distribuées, également appelées « bases de données réparties », peuvent être confondues avec d'autres types de bases de données comme:

- les *bases de données fédérées* ou
- les *bases de données parallèles*.

#### ➤ Les bases de données fédérées :

- Ensemble de bases de données intégrées via un modèle commun;
- Vue commune des données via un modèle global ;
- Chaque site a son schéma local, pas forcément inclus entièrement dans le schéma global (il y a un site central).

#### ➤ Les bases de données parallèles :

- Exploitent le parallélisme à l'intérieur d'un même site ;
- Peuvent utiliser plusieurs processeurs et/ou disques.



### 2. Historique donnant naissance à la réplication

A la fin des années 80 et au début des années 90, la plupart des fournisseurs de SGBD avaient des projets voire des produits de versions distribuées de leur SGBD et des capacités à fédérer des BD (homogènes et hétérogènes).

Les différentes expériences faites, par les utilisateurs des applications, avec ces SGBD ont montrés leurs limites, en particulier dans un contexte de mise à jour.

Au début des années 2000, les fournisseurs ne mettent plus en avant les aspects distribués.

Les fournisseurs ont évolués vers des solutions à base de réplication des bases de données et des moyens d'accès à des bases de données distantes.

### 3. Problématique

La mise en place des bases de données a subi une évolution avec de nouvelles versions de SGBD qui prennent en compte certaines contraintes comme les types de données à accueillir. D'autres pressions en demeurent :

- Augmentation du volume de l'information et des transactions
- Impératif de décentralisation de l'information

Ces défis conduisent à une lenteur applicative (temps de réponse plus grand ou pas d'accès à l'information). D'où la nécessité de serveurs de BD qui fournissent un bon temps de réponse sur des gros volumes de données.

- ✓ Quel type d'architecture faut-il adoptée pour répondre à ces attentes ?
- ✓ Quelles sont les aspects techniques à prendre en compte pour faciliter la mise en œuvre ?

### 3. Problématique

- ✓ Faut-il répercuter les modifications de manière synchrone ou asynchrone ?
- ✓ Comment gérer les conflits d'écriture si deux opérations différentes ont lieu sur la même donnée sur des nœuds différents, et ce en même temps ?
- ✓ Comment gérer les limites inhérentes des systèmes distribués, notamment les latences réseau pour la réplication des données (certains nœuds peuvent recevoir les données avant les autres).
- ✓ Comment gérer l'échec d'un nœud : re-routage des requêtes, mise à jour lors de sa réintégration.

### 4. Buts de la distribution des données

Les bases de données distribuées ont une architecture plus adaptée à l'organisation des entreprises décentralisées.

Les raisons qui président à la création des bases de données réparties sont multiples :

- Limiter le transfert d'informations
- Augmenter la fiabilité et disponibilité : les bases de données distribuées ont souvent des données répliquées.
- Meilleures performances : la réduction du trafic sur le réseau, la répartition de charge de travail entre plusieurs unités de traitement ou de stockage permettent l'accroissement des performances. Le but de la répartition des données est de les rapprocher de l'endroit où elles sont accédées. Répartir une base de données sur plusieurs sites permet de répartir la charge sur les processeurs et sur les entrées/sorties.
- Accroître la performance: l'accroissement se fait par l'ajout de machines sur le réseau. Les données se trouvent près du site de la plus grande demande/ la concurrence pour les services de l'unité central et des entrées sorties se trouvent nettement réduite par rapport à un SGBD centralisé
- Economie: l'ajout des stations de travail à un réseau est nettement moins coûteux que l'extension d'un gros système centralisé

### 5. Avantages et inconvénients de BD distribuées

#### a) Avantages :

- **Amélioration du partage et de l'autonomie locale ;**
- **Amélioration de la disponibilité :** une panne d'un SGBD dans un site ou une rupture de ligne de communication isolant un ou quelques sites n'immobilise pas l'ensemble du système ;
- **Performances améliorées :** Les données se trouvent près du site de la plus grande demande/la concurrence pour les services de l'unité centrale et des entrées sorties se trouve nettement réduite par rapport à un SGBD centralisé ;
- **Economie :** l'ajout de stations de travail à un réseau est nettement moins coûteux que l'extension d'un gros système centralisé ;
- **Facilité d'accroissement** (scalabilité).

### 5. Avantages et inconvénients de BD distribuées

#### b) Inconvénients :

- ✓ **Complexité:** un SGBDD masque sa nature répartie aux yeux des utilisateurs et fournit un niveau acceptable de performances, de fiabilité et de disponibilité
- ✓ **Sécurité:** Dans un système centralisé, l'accès aux données est d'un contrôle facile, tandis que dans un système distribué, non seulement il faut contrôler l'accès aux données dupliquées dans des emplacements multiples, mais le réseau lui-même doit être sécurisé.
- ✓ **Contrôle d'intégrité de données plus difficile:** L'intégrité de base de données fait référence à la validité et à la cohérence des données stockées.
- ✓ **Complexité plus grande du design de base de données:** le design d'une base de données distribuée doit prendre en considération la fragmentation des données, l'allocation des fragments à des sites spécifiques et la réplication des données,
- ✓ **Coût:** la distribution entraîne des coûts supplémentaires en terme de communication, et de gestion des communications (hardware et software à installer pour gérer les communications et la distribution)

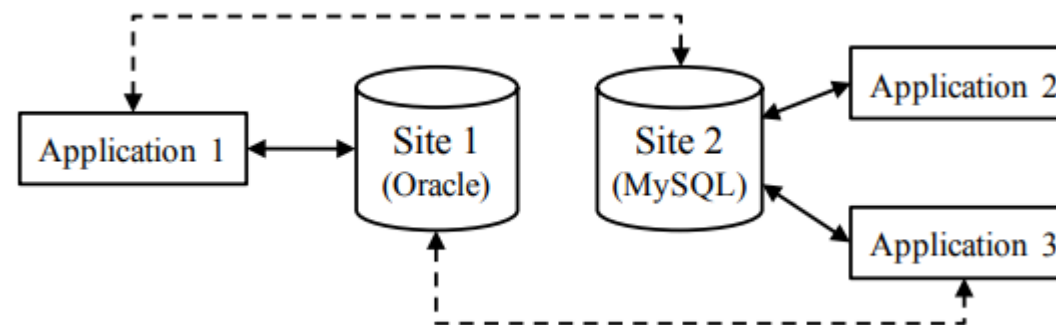
### 6. SGBD Distribuées

#### a) Description et spécificités

Une base de données centralisée est gérée par un seul SGBD, est stockée dans sa totalité à un emplacement physique unique et ses divers traitements sont confiés à une seule et même unité de traitement. Par opposition, une base de données distribuée gérée par plusieurs processeurs, sites ou SGBD.

Un système de bases de données distribuées ne doit donc en aucun cas être confondu avec un système dans lequel les bases de données sont accessibles à distance. Il ne doit non plus être confondu avec une multi-base.

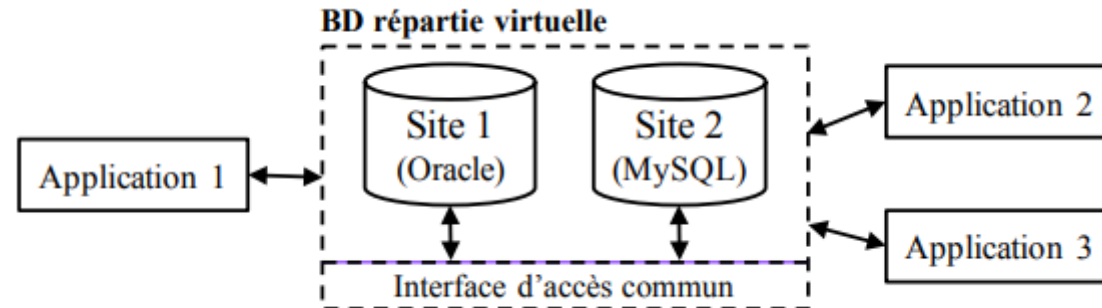
Dans une multi-base, plusieurs BDs inter opèrent avec une application via un langage commun et sans modèle commun.



### 6. SGBD Distribuées

#### a) Description et spécificités

Dans une BD fédérée, plusieurs BDs hétérogènes sont accédées comme une seule via une vue commune





### 6. SGBD Distribuées

#### a) Description et spécificités

En général, un SGBD réparti doit être capable d'offrir les mêmes services qu'un SGBD centralisé, en déchargeant les utilisateurs de tous les problèmes de concurrence, de fiabilité et d'optimisation de requêtes.

Ainsi, un SGBD réparti doit disposer de :

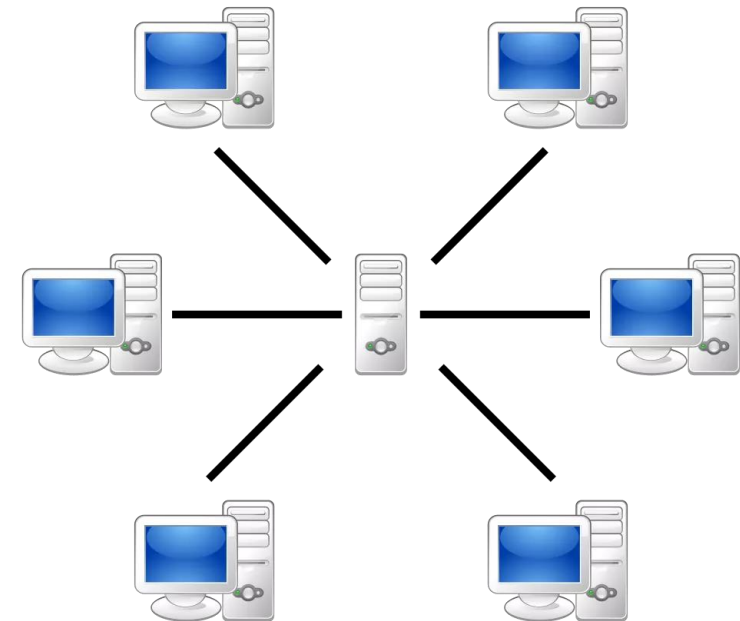
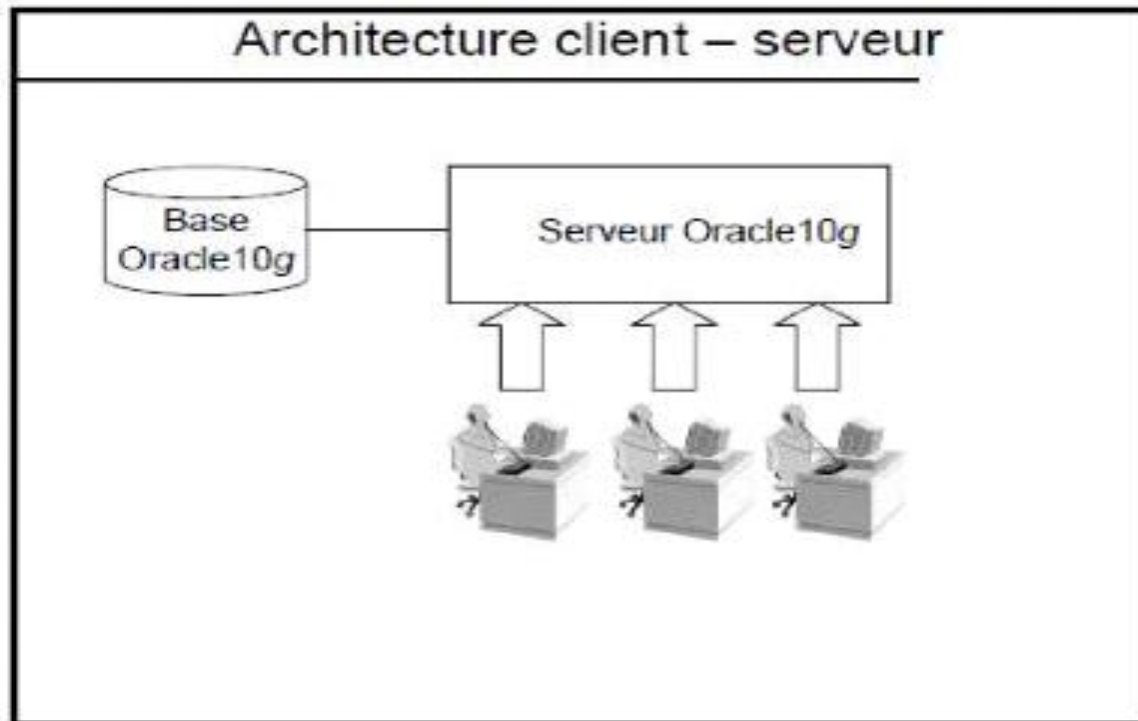
- Dictionnaire de données réparties ;
- Traitement de requêtes réparties ;
- Gestion des transactions réparties ;
- Communication de données inter-site ;
- Gestion de la cohérence et de la sécurité.

## 6. SGBD Distribuées

### b) Architectures des SGBD Distribuées

Du point de vue organisationnel nous distinguons deux architectures :

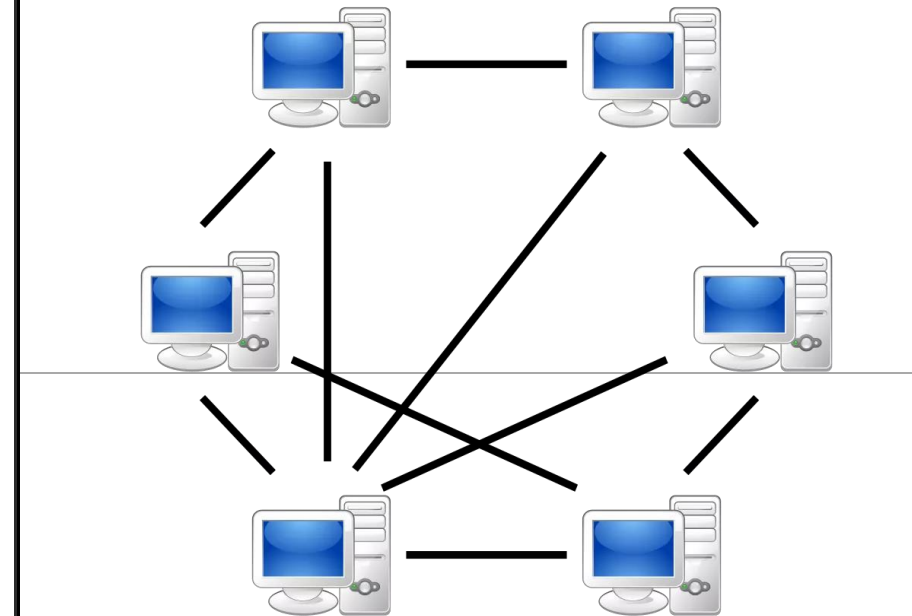
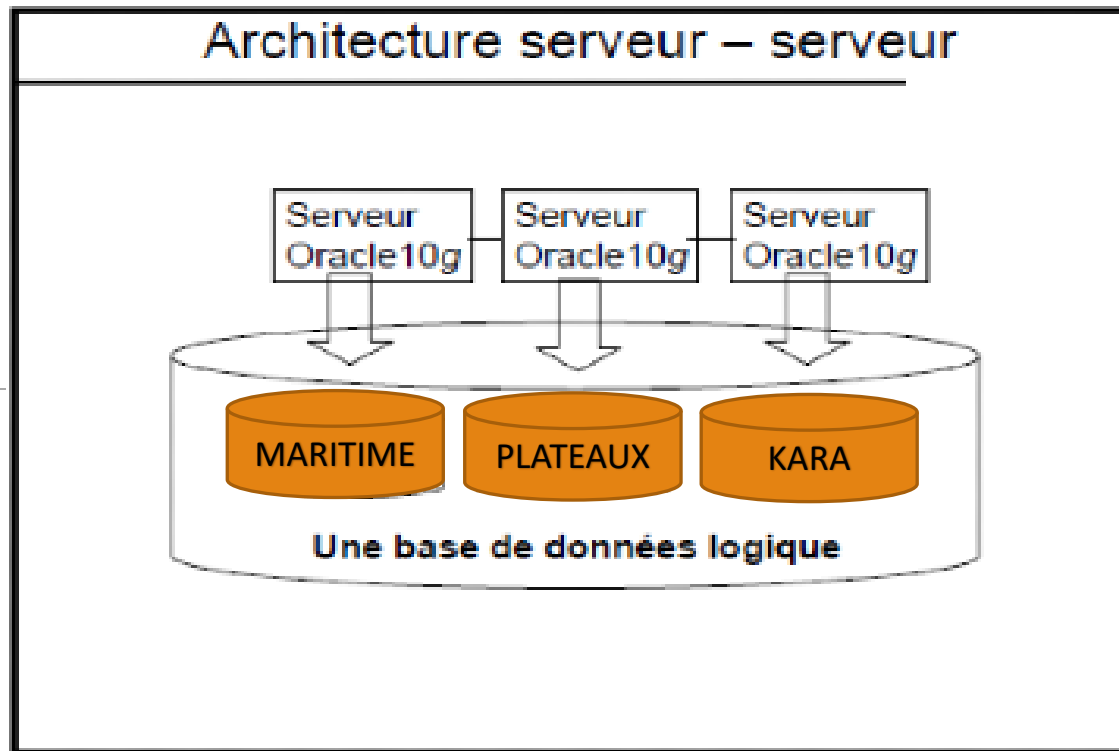
- ❑ **Architecture Client-Serveur** : les serveurs, ont pour rôle de réaliser une tâche demandée par le client.



## 6. SGBD Distribuées

### b) Architectures des SGBD Distribuées

- ❑ **Architecture Pair-à-Pair ou serveur-serveur (Peer-to-Peer, P2P)** : par ce terme on désigne un type de communication pour lequel toutes les machines ont une importance équivalente.



### 7. Objectifs

Les principaux objectifs visés par une répartition des données avec ces architectures sont :

- La transparence pour l'utilisateur,
- Autonomie de chaque site,
- Pas de site privilégié,
- Continuité de service,
- Traitement des requêtes distribuées,
- Transparence vis-à-vis de :
  - **la localisation des données** : L'utilisateur accède au schéma conceptuel via des vues ; il ne sait pas sur quel site se trouvent physiquement les données
  - **la fragmentation ou partitionnement** l'utilisateur ne sait pas comment la base est partitionnée
  - **la réplication** l'utilisateur ne sait pas s'il existe des copies des informations, ni où elles se trouvent si elles existent
- Indépendance au niveau du:
  - matériel
  - système d'exploitation
  - réseau
  - SGBD