

Bases de données

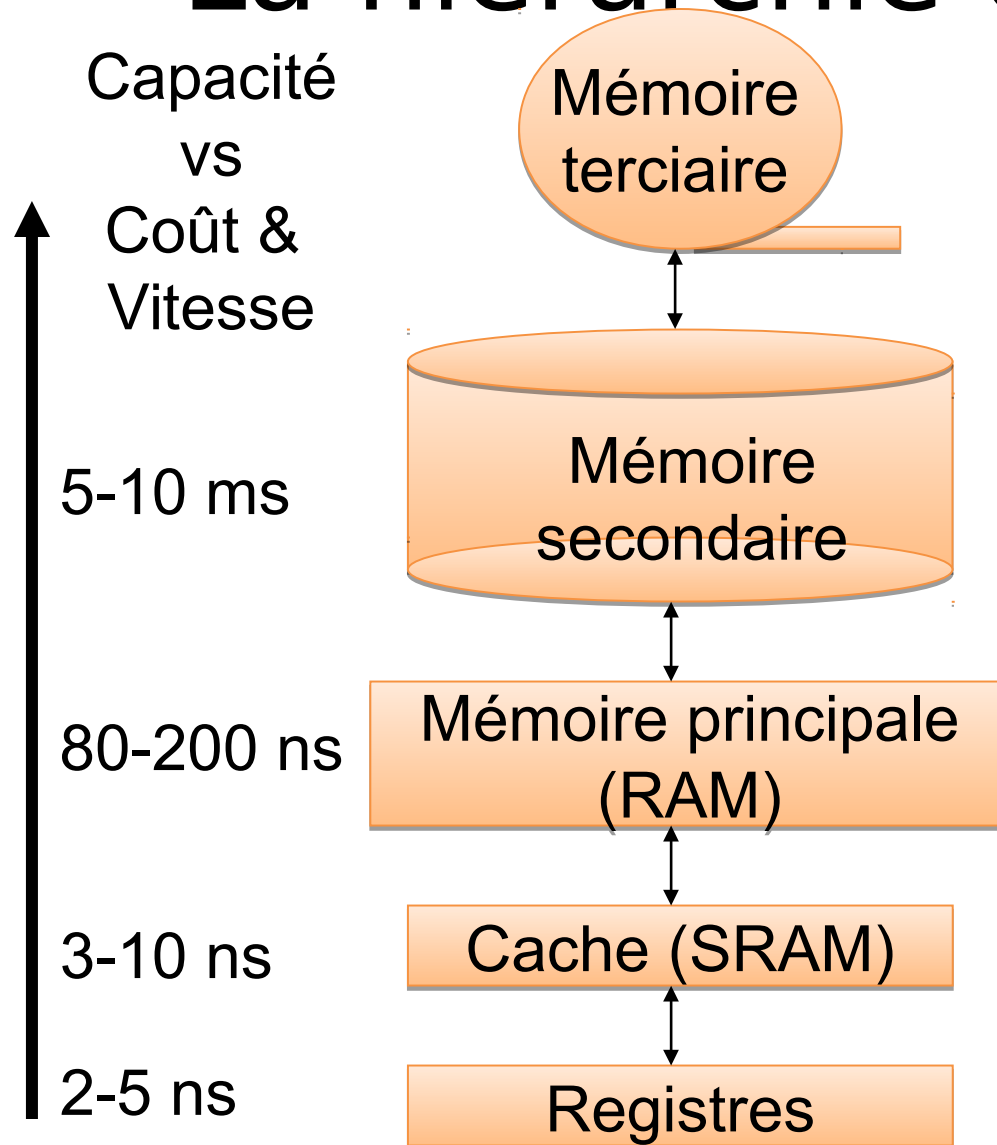
Introduction et

Objectifs

1. Introduction

- **Les entreprises gèrent des volumes de données très grands**
 - **Giga, Terra, Péta -octets**
 - **Numériques, Textuelles, Multi-média (images, films,...)**
- **Il faut pouvoir facilement**
 - **Archiver les données sur mémoires secondaires permanente**
 - **Retrouver les données pertinentes à un traitement**
 - **Mettre à jour les données variant dans le temps**
- **Les données sont structurées et identifiées**
 - **Données élémentaires ex: Votre salaire, Votre note en BD**
 - **Données composées ex: Votre CV, vos résultats de l'année**
 - **Identifiant humain ex: NSS ou machine: P26215**
- **Qu'est-ce qu'une BD ?**
 - **Collection de données structurées reliées par des relations**
 - **Interrogeable et modifiable par des langages de haut niveau**

La hiérarchie des mémoires



- *Un accès disque est environ 100,000 fois plus lent qu'un accès mémoire!*
- \Rightarrow
 - Eviter les accès disques
 - grande mémoire principale
 - Amortir les accès disques
 - placement des données
 - Minimiser le nombre d'accès disques
 - méthodes d'accès

Un peu d'histoire

- **Années 60:**
 - **Réceptacles logiques de données → fichiers sur disque**
 - **Accès séquentiel puis sur clé**
 - Lire (Nomf, Article), Ecrire (Nomf, Article)
 - Lire (Nomf, Article, Clé), Ecrire (Nomf, article, Clé)
- **Années 70:**
 - **Avènement des Bases de Données Réseaux (BD)**
 - **Ensemble de fichiers reliés par des pointeurs**
 - **Langage d'interrogation par navigation**
- **Années 80:**
 - **Avènement des Bases de Données Relationnelles (BDR)**
 - **Relations entre ensemble de données**
 - **Langage d'interrogation par assertion logique**

Systemes de fichiers



Comptabilité



Chirurgie

Consultations



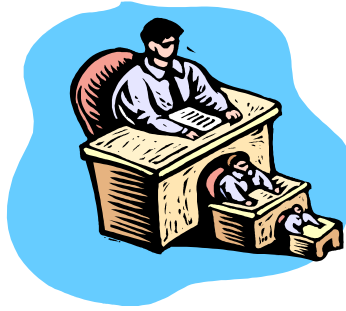
Psychiatrie



Caractéristiques

Problèmes

Format des fichiers



Talla
Symptomes : y
Turlututu : sqj
Symptomes : y
Turlututu : sdd
Analyses : xxx

Dika
Turlututusqjsk
Symptom: yyyy
Analyses xxxx
Turlututudhjsd
Analyses :xx



Achu
Symptomes : yy
Analyses : xxx
Symptomes : y

Dongo
Turlututu : sq
Symptomyyyy
Analysesxxxx
Turlututudhjsd



Caractéristiques

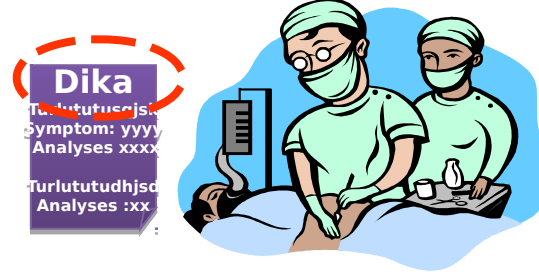
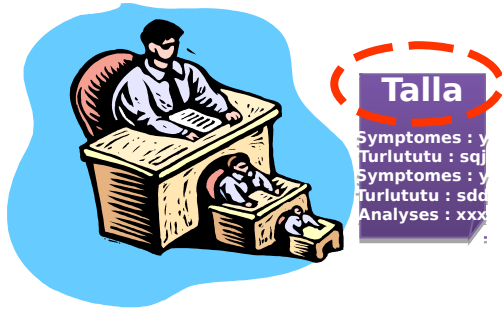
Plusieurs applications

- plusieurs formats
- plusieurs langages

Problèmes

- Difficultés de gestion

Redondance (données)



Caractéristiques

Plusieurs applications

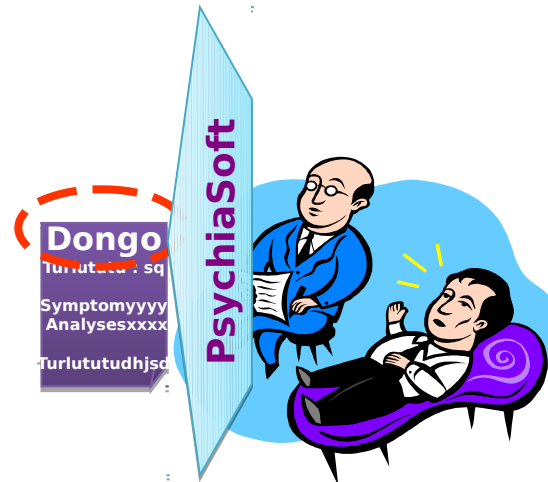
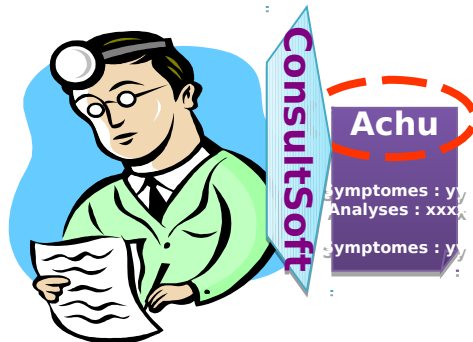
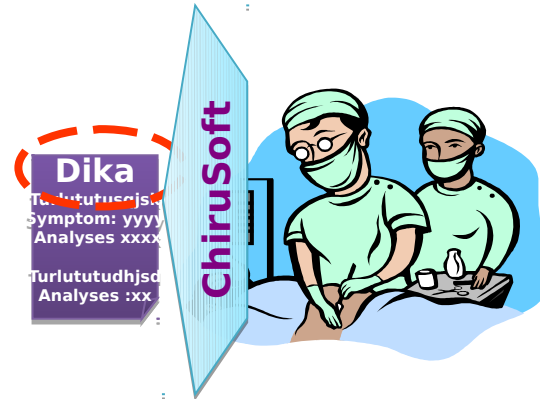
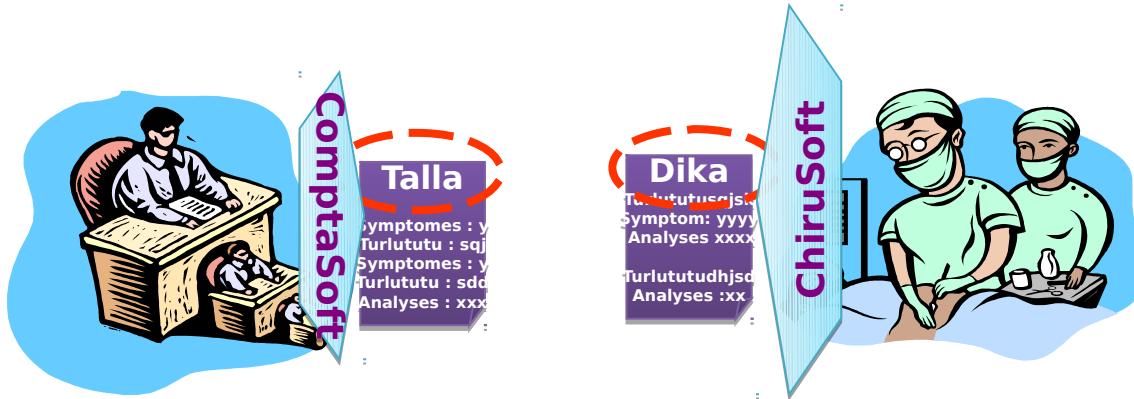
- plusieurs formats
- plusieurs langages

Redondance de données

Problèmes

- Difficultés de gestion
- Incohérence des données

Interrogations



Caractéristiques

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

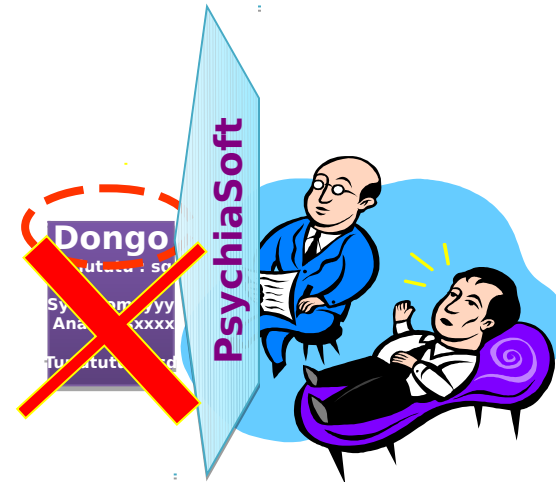
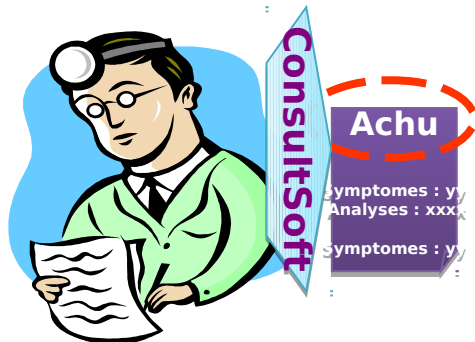
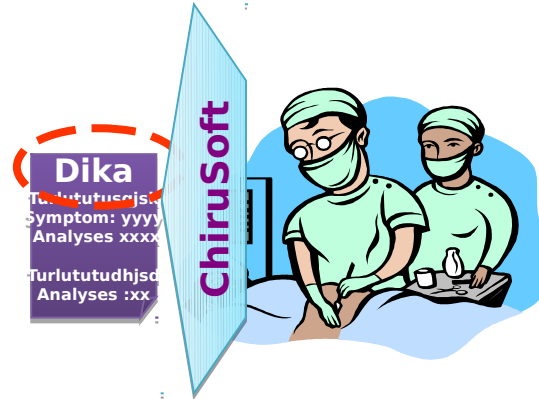
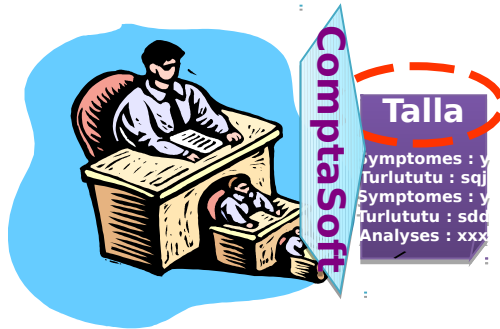
Pas de facilité d'interrogation

- Question ⇒ développement

Problèmes

- Difficultés de gestion
- Incohérence des données
- Coûts élevés
- Maintenance difficile

Pannes ???



Caractéristiques

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

Pas de facilité d'interrogation

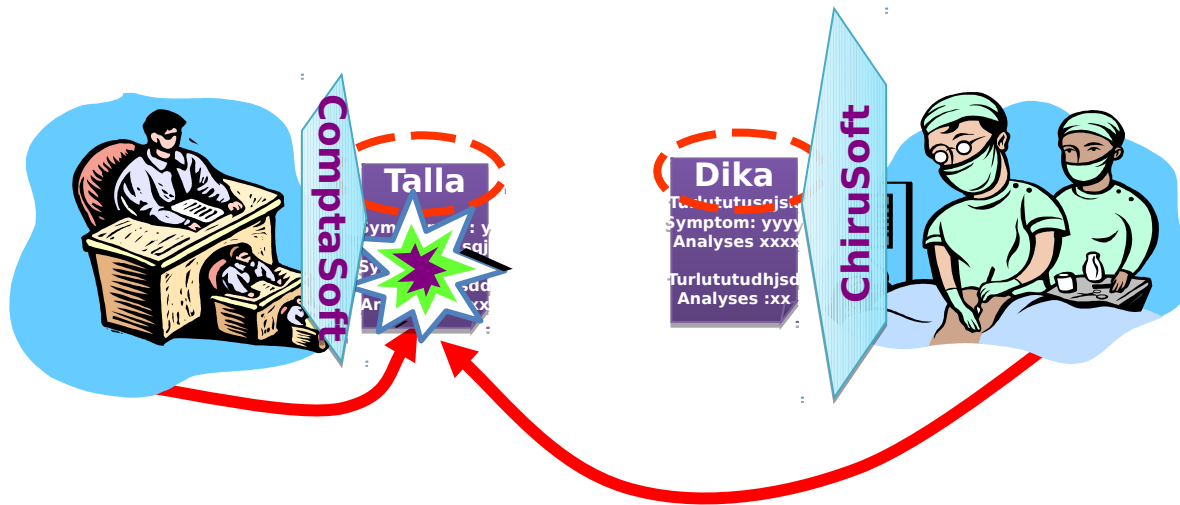
- Question ⇒ développement

Redondance de code

Problèmes

- Difficultés de gestion
- Incohérence des données
- Coûts élevés
- Maintenance difficile
- Gestion de pannes ???

Partage de données



Caractéristiques

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

Pas de facilité d'interrogation

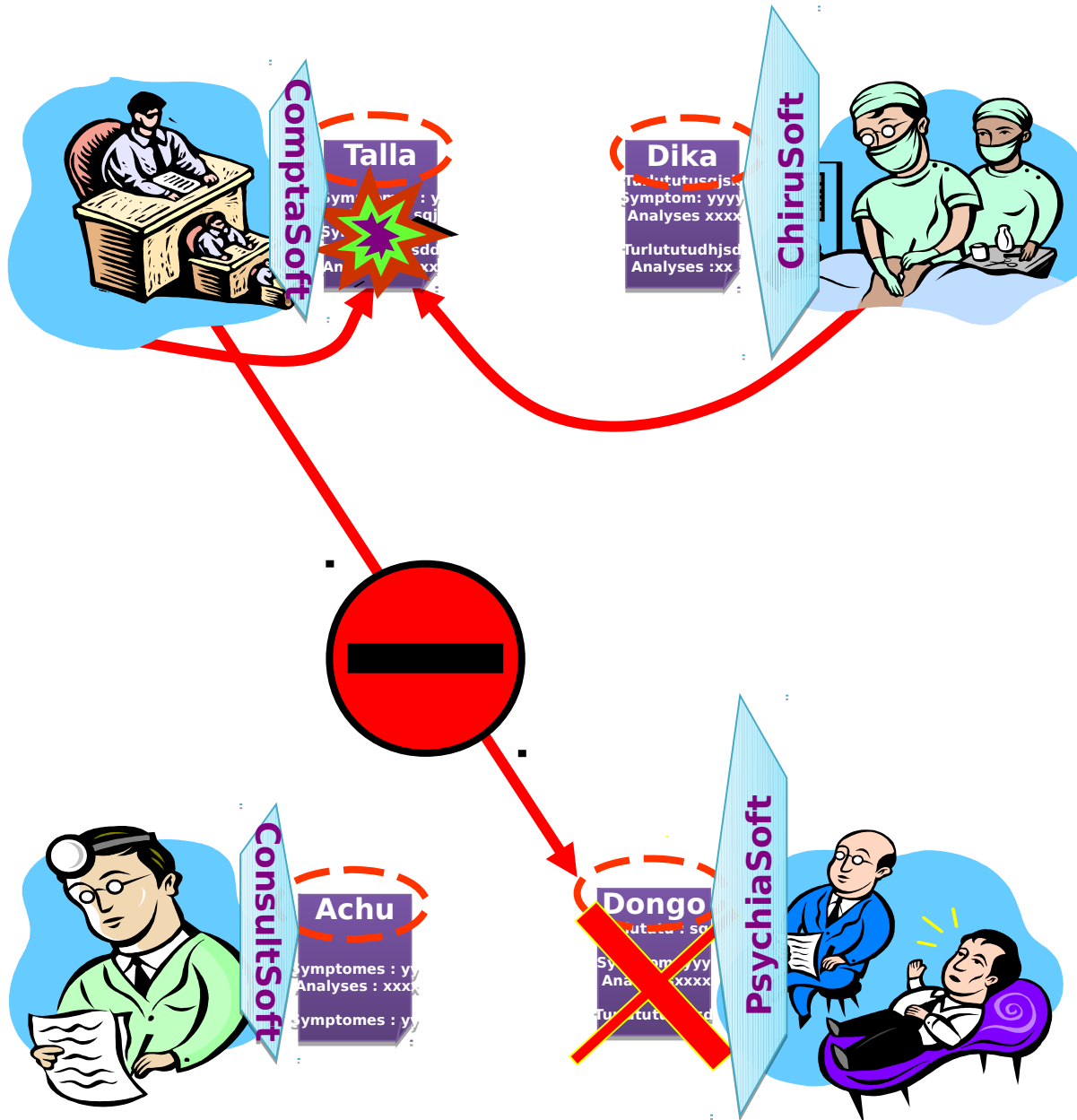
- Question ⇒ développement

Redondance de code

Problèmes

- Difficultés de gestion
- Incohérence des données
- Coûts élevés
- Maintenance difficile
- Gestion de pannes ???
- Partage des données ???

Confidentialité



Caractéristiques

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

Pas de facilité d'interrogation

- Question ⇒ développement

Redondance de code

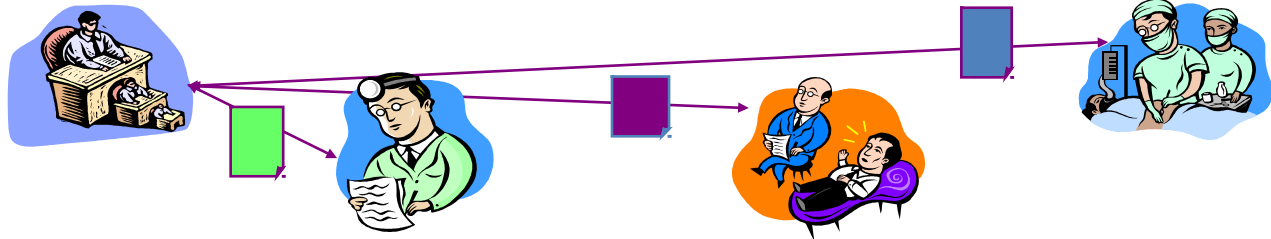
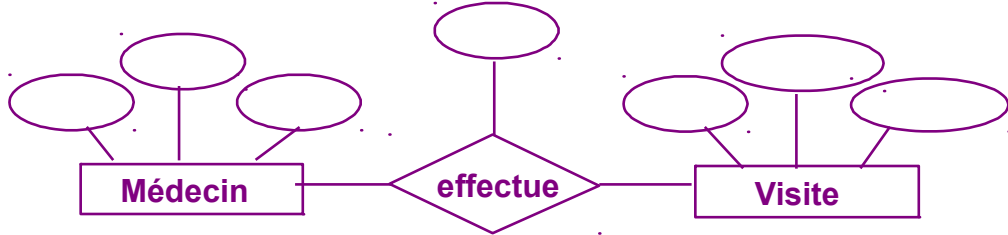
Problèmes

- Difficultés de gestion
- Incohérence des données
- Coûts élevés
- Maintenance difficile
- Gestion de pannes ???
- Partage des données ???
- Confidentialité ???

L'approche “Bases de données”

- **Modélisation des données**
 - ➔ **Éliminer la redondance de données**
 - ➔ **Centraliser et organiser correctement les données**
 - ➔ **Plusieurs niveaux de modélisation**
 - ➔ **Outils de conception**
- **Logiciel «Système de Gestion de Bases de Données»**
 - ➔ **Factorisation des modules de contrôle des applications**
 - Interrogation, cohérence, partage, gestion de pannes, etc...
 - ➔ **Administration facilitées des données**

Modélisation du réel

Réel					
Modèle conceptuel	<ul style="list-style-type: none">• Indépendant du modèle de données• Indépendant du SGBD				
Modèle logique	<ul style="list-style-type: none">• Dépendant du modèle de données• Indépendant du SGBD	CodasyI	Relationnel	Objet	XML
Modèle Physique	<ul style="list-style-type: none">• Dépendant du modèle de données• Dépendant du SGBD	<ul style="list-style-type: none">• Organisation physique des données• Structures de stockage des données• Structures accélératrices (index)			

12

Modélisation Relationnelle

(1)

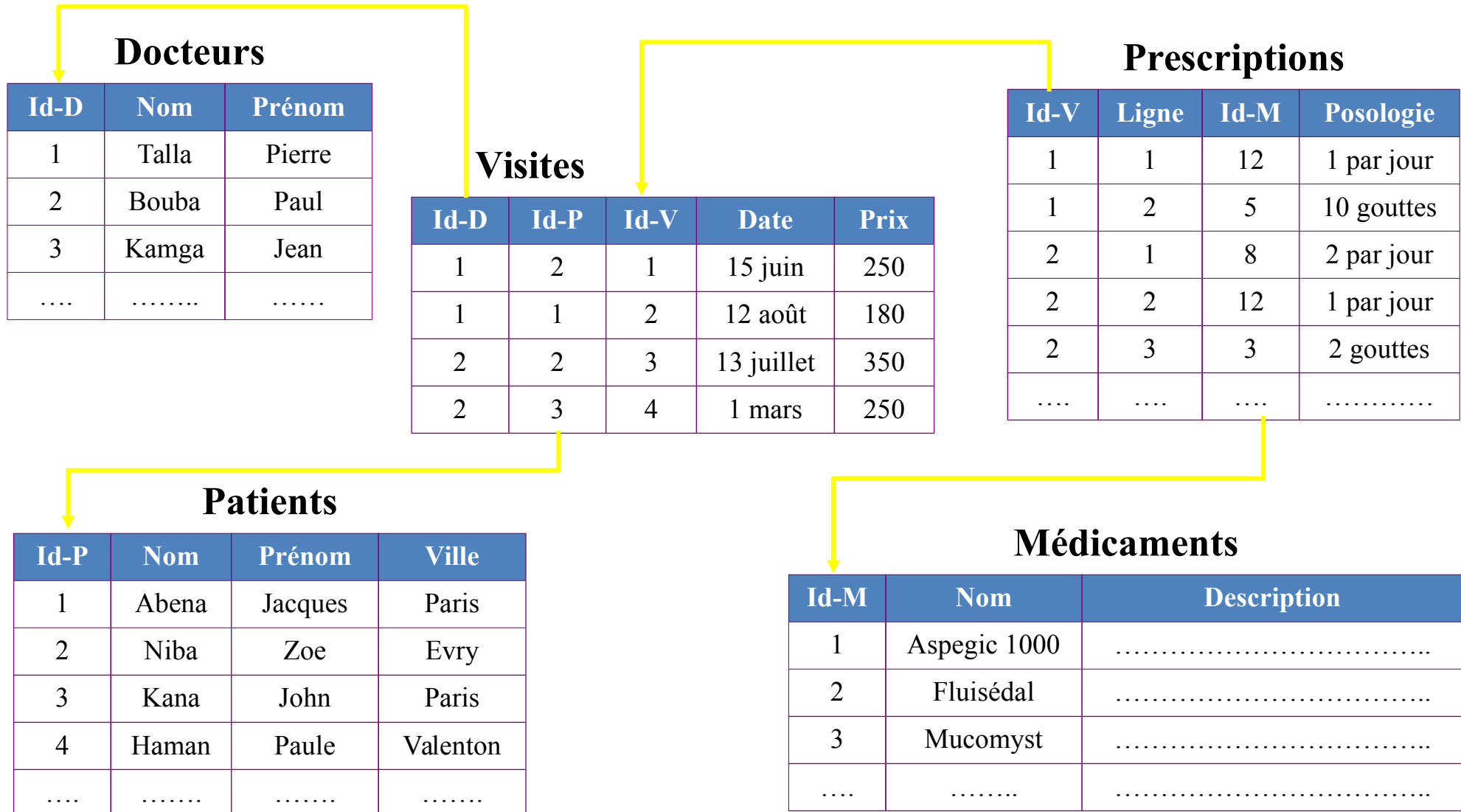
Relation ou
table

Champs,
attributs,
colonnes

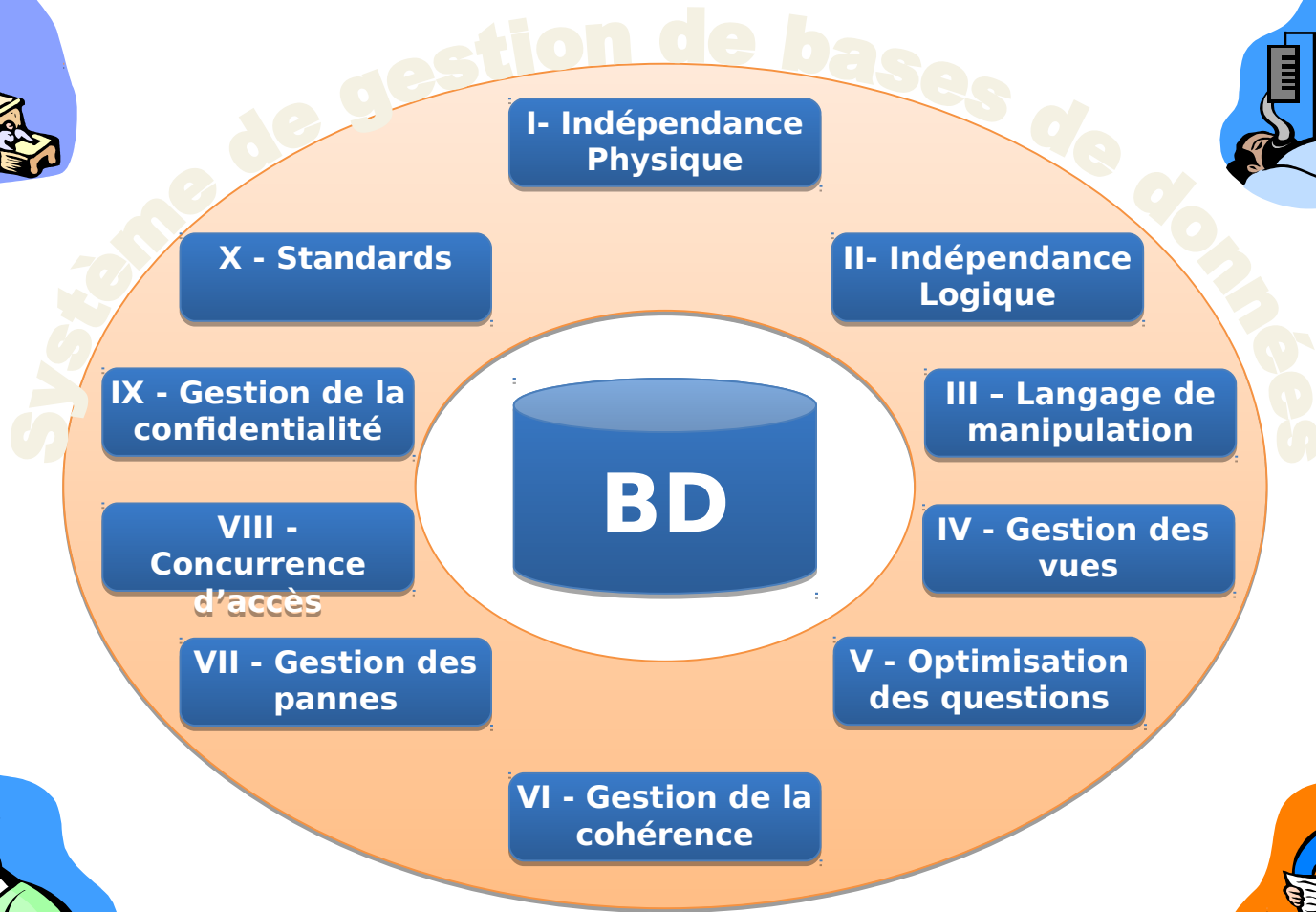
Id-D	Nom	Prénom
1	Talla	Pierre
2	Bouba	Paul
3	Kamga	Jean
....

Tuples, lignes ou
n-uplets

Modélisation Relationnelle (2)



2. Objectifs des SGBD



I - Indépendance Physique

- **Indépendance des programmes d'applications vis à vis du modèle physique :**
 - **Possibilité de modifier les structures de stockage (fichiers, index, chemins d'accès, ...) sans modifier les programmes;**
 - **Ecriture des applications par des non-spécialistes des fichiers et des structures de stockage;**
 - **Meilleure portabilité des applications et indépendance vis à vis du matériel.**

II - Indépendance Logique

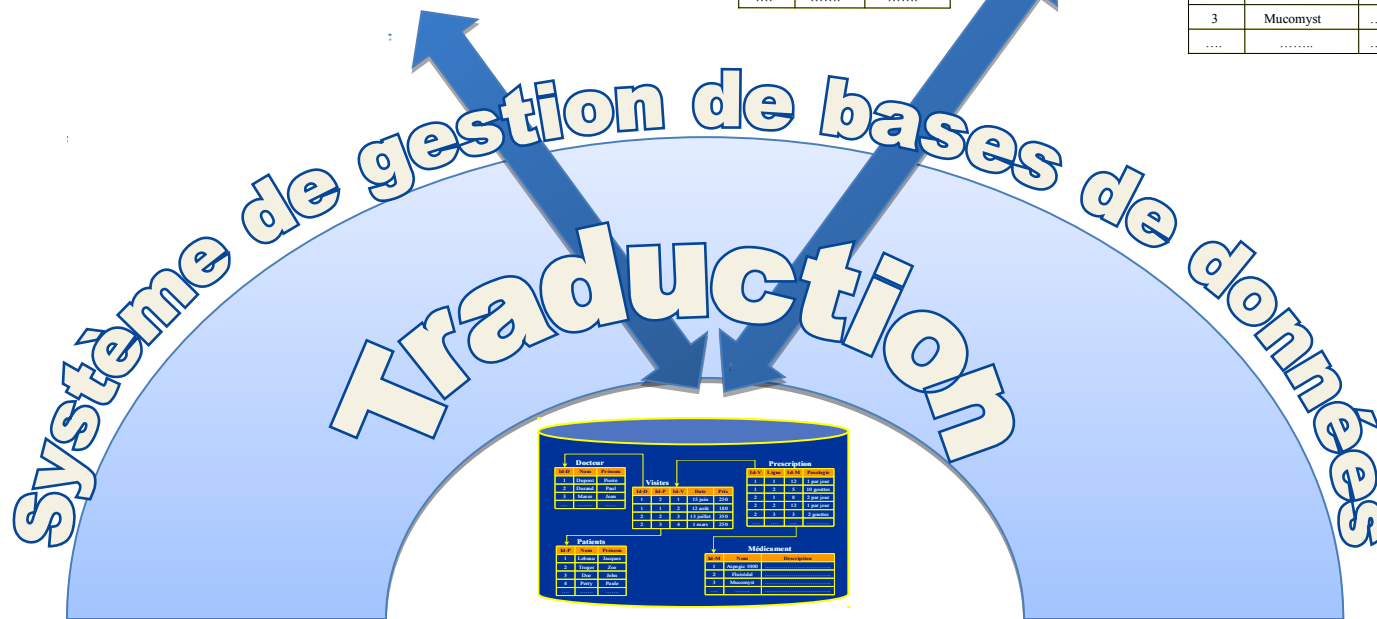
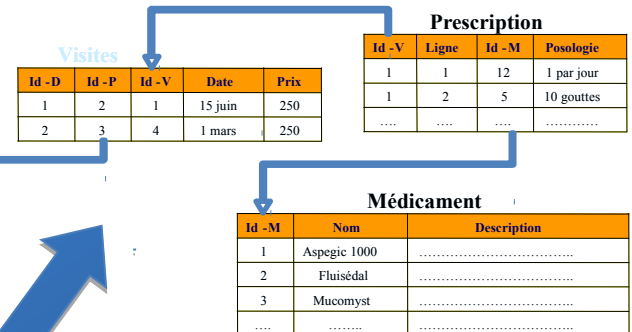
Les applications peuvent définir des vues logiques de la BD

Gestion des médicaments

Nombre_Médicaments

Id-M	Nom	Description	Nombre
1	Aspegic 1000	30
2	Fluisédal	20
3	Mucomyst	230
....

Cabinet du Dr. Kamga



Avantages de l'indépendance logique

- **Possibilité pour chaque application d'ignorer les besoins des autres (bien que partageant la même BD).**
- **Possibilité d'évolution de la base de données sans réécriture des applications :**
 - **ajout de champs, ajout de relation, renommage de champs.**
- **Possibilité d'intégrer des applications existantes sans modifier les autres.**
- **Possibilité de limiter les conséquences du partage : Données confidentielles.**

III - Manipulation aisée

- **La manipulation se fait via un langage déclaratif**
 - **La question déclare l'objectif sans décrire la méthode**
 - **Le langage suit une norme commune à tous les SGBD**
 - **SQL : Structured Query Language**
- **Sémantique**
 - **Logique du 1er ordre ++**
- **Syntaxe (aperçu !)**
 - **SELECT <structure des résultats>**
 - **FROM <relations>**
 - **WHERE <conditions>**

IV – Des vues multiples des données

- **Les vues permettent d'implémenter l'indépendance logique en permettant de créer des relations virtuelles**
- **Vue = Question stockée**
- **Le SGBD stocke la définition et non le résultat**
- **Exemple :**
 - **la vue des patients parisiens**
 - **la vue des docteurs avec leurs patients**
 - **La vue des services statistiques**
 - **...**

V - Exécution et Optimisation

- **Traduction automatique des questions déclaratives en programmes procéduraux :**
 - ➔ **Utilisation de l'algèbre relationnelle**
- **Optimisation automatique des questions**
 - ➔ **Utilisation de l'aspect déclaratif de SQL**
 - ➔ **Gestion centralisée des chemins d'accès (index, hachages, ...)**
 - ➔ **Techniques d'optimisation poussées**
- **Economie de l'astuce des programmeurs**
 - **milliers d'heures d'écriture et de maintenance de logiciels.**

VI - Intégrité Logique

- **Objectif : Détecter les mises à jour erronées**
- **Contrôle sur les données élémentaires**
 - **Contrôle de types: ex: Nom alphabétique**
 - **Contrôle de valeurs: ex: Salaire mensuel entre 5 et 50kf**
- **Contrôle sur les relations entre les données**
 - **Relations entre données élémentaires:**
 - Prix de vente > Prix d'achat
 - **Relations entre objets:**
 - Un électeur doit être inscrit sur une seule liste électorale

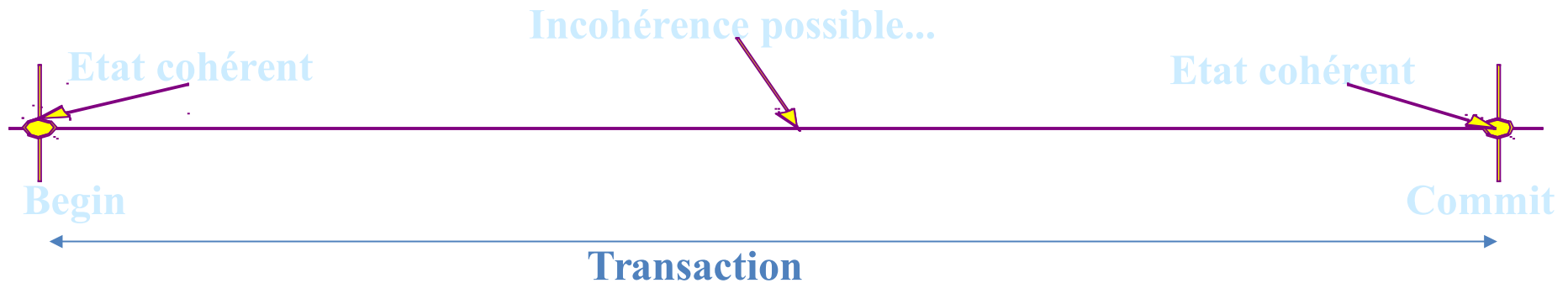
Contraintes d'intégrité

- **Avantages :**
 - simplification du code des applications
 - sécurité renforcée par l'automatisation
 - mise en commun des contraintes
- **Nécessite :**
 - un langage de définition de contraintes d'intégrité
 - la vérification automatique de ces contraintes

VII - Intégrité Physique

- **Motivations : Tolérance aux fautes**
 - **Transaction Failure : Contraintes d'intégrité, Annulation**
 - **System Failure : Panne de courant, Crash serveur ...**
 - **Media Failure : Perte du disque**
 - **Communication Failure : Défaillance du réseau**
- **Objectifs :**
 - **Assurer l'atomicité des transactions**
 - **Garantir la durabilité des effets des transactions commises**
- **Moyens :**
 - **Journalisation : Mémorisation des états successifs des données**
 - **Mécanismes de reprise**

Transaction



Begin

$CE_{pargne} = CE_{pargne} - 3000$

$CC_{courant} = CC_{courant} + 3000$

Commit T1

Atomicité et Durabilité

ATOMICITE

Begin

CEpargne = CEpargne -
3000

CCourant = CCourant +
3000

Commit T1

Panne



→ Annuler le débit !!

DURABILITE

Begin

CEpargne = CEpargne -
3000

CCourant = CCourant +
3000

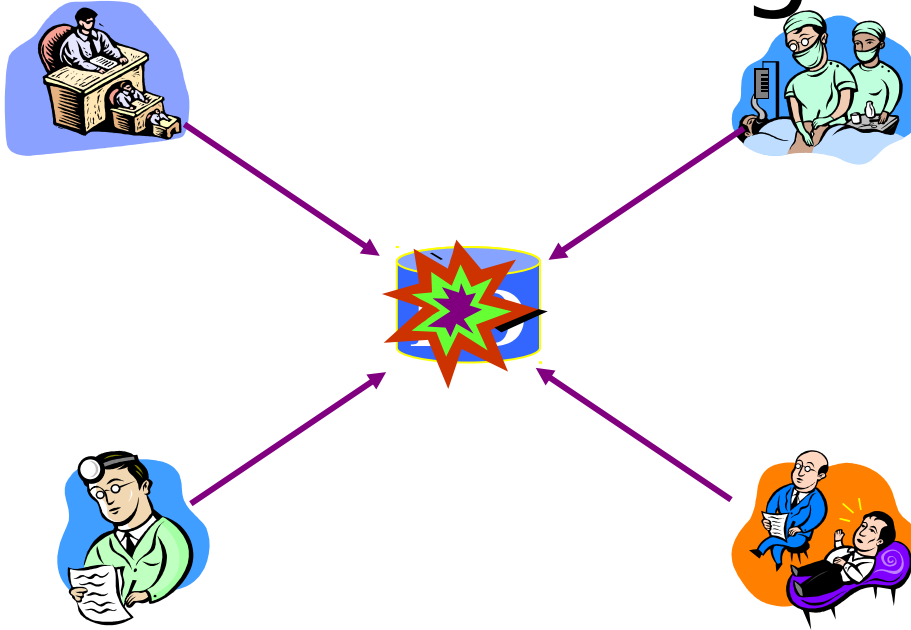
Commit T1

Crash disque



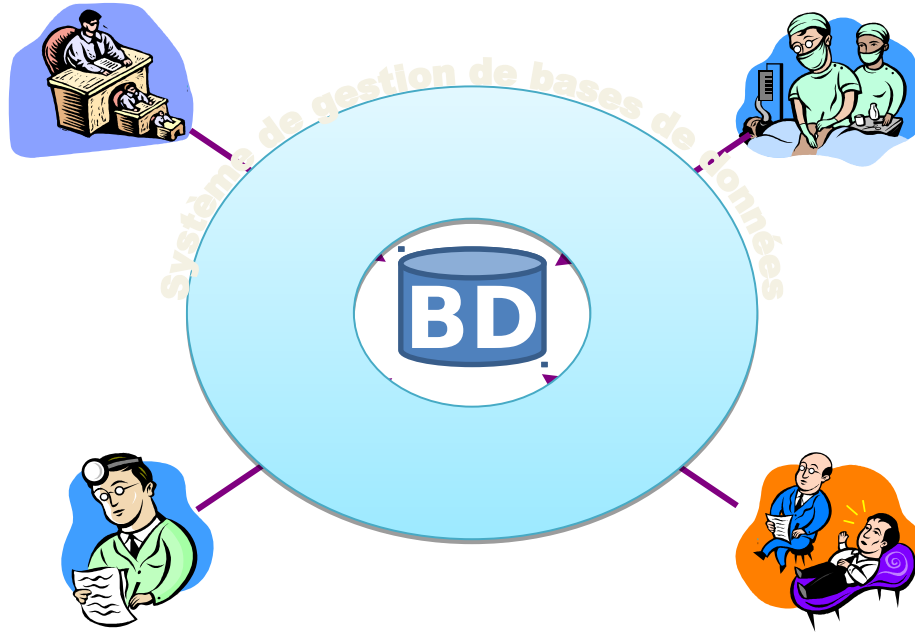
→ S'assurer que le
virement a été fait !

VIII - Partage des données



- Accès concurrent aux mêmes données
➔ Conflits d'accès !!

Isolation et Cohérence



- Le SGBD gère les accès concurrents
 - ➔ Chacun à l'*impression* d'être seul (Isolation)
 - ➔ Cohérence conservée (Pas de maj conflictuelles)

IX – Confidentialité

- **Objectif : Protéger les données de la BD contre des accès non autorisés**
- **Deux niveaux :**
 - **Connexion restreinte aux usagers répertoriés (mot de passe)**
 - **Privilèges d'accès aux objets de la base**
- **Usagers : Usager ou groupe d'usagers**
- **Objets : Relation, Vue, autres objets (procédures, etc.)**

X - Standardisation

- **L'approche bases de données est basée sur plusieurs standards**
 - **Langage SQL (SQL1, SQL2, SQL3)**
 - **Communication SQL CLI (ODBC / JDBC)**
 - **Transactions (X/Open DTP, OSI-TP)**
- **Force des standards**
 - **Portabilité**
 - **Interopérabilité**
 - **Applications multisources...**

3. Architecture des SGBD

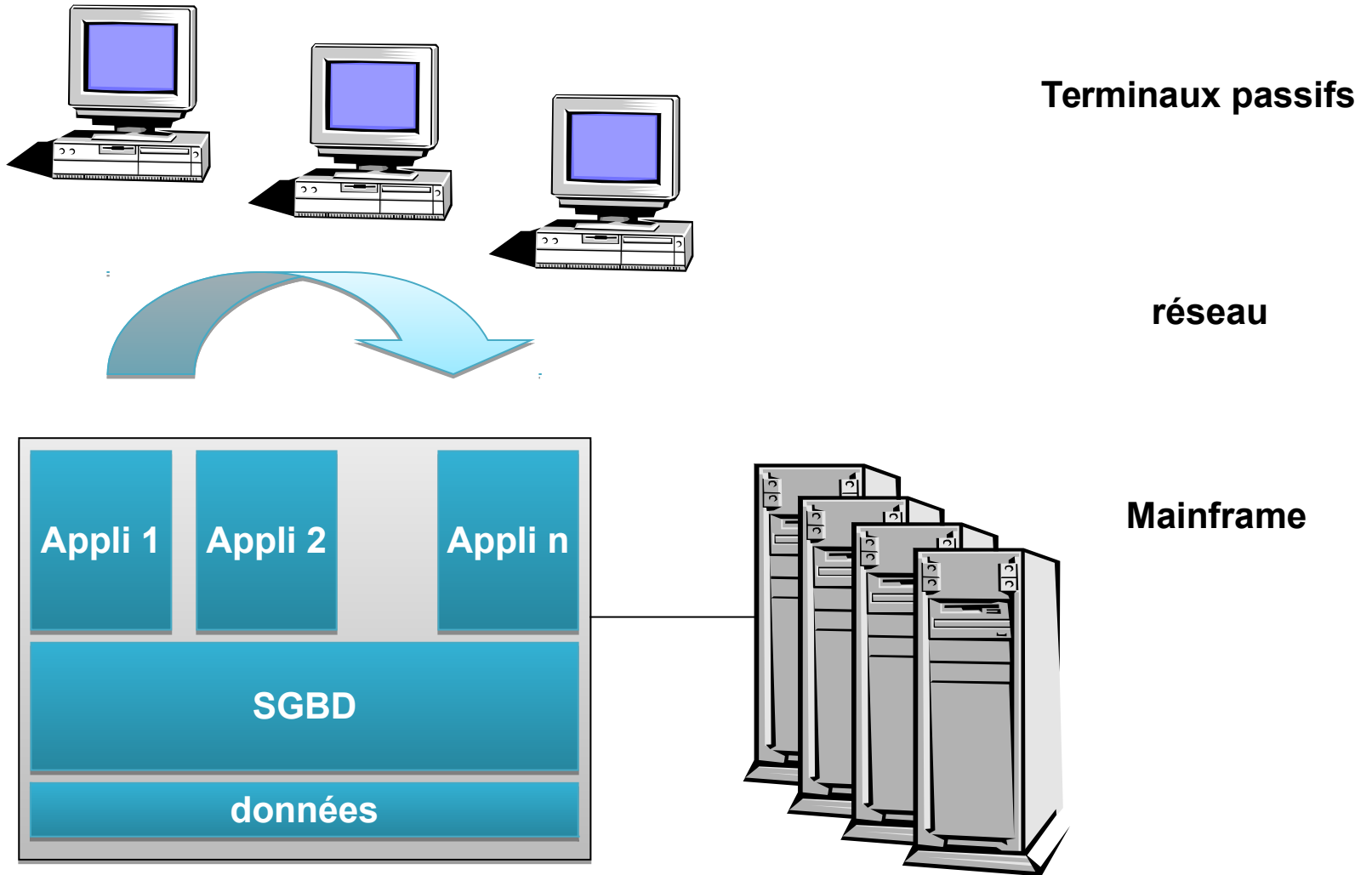
👉 Les architectures physiques de SGBD sont très liées au mode de répartition.

- BD centralisée
- BD client/serveur
- BD client/multi-serveurs
- BD répartie
- BD hétérogène
- BD mobile

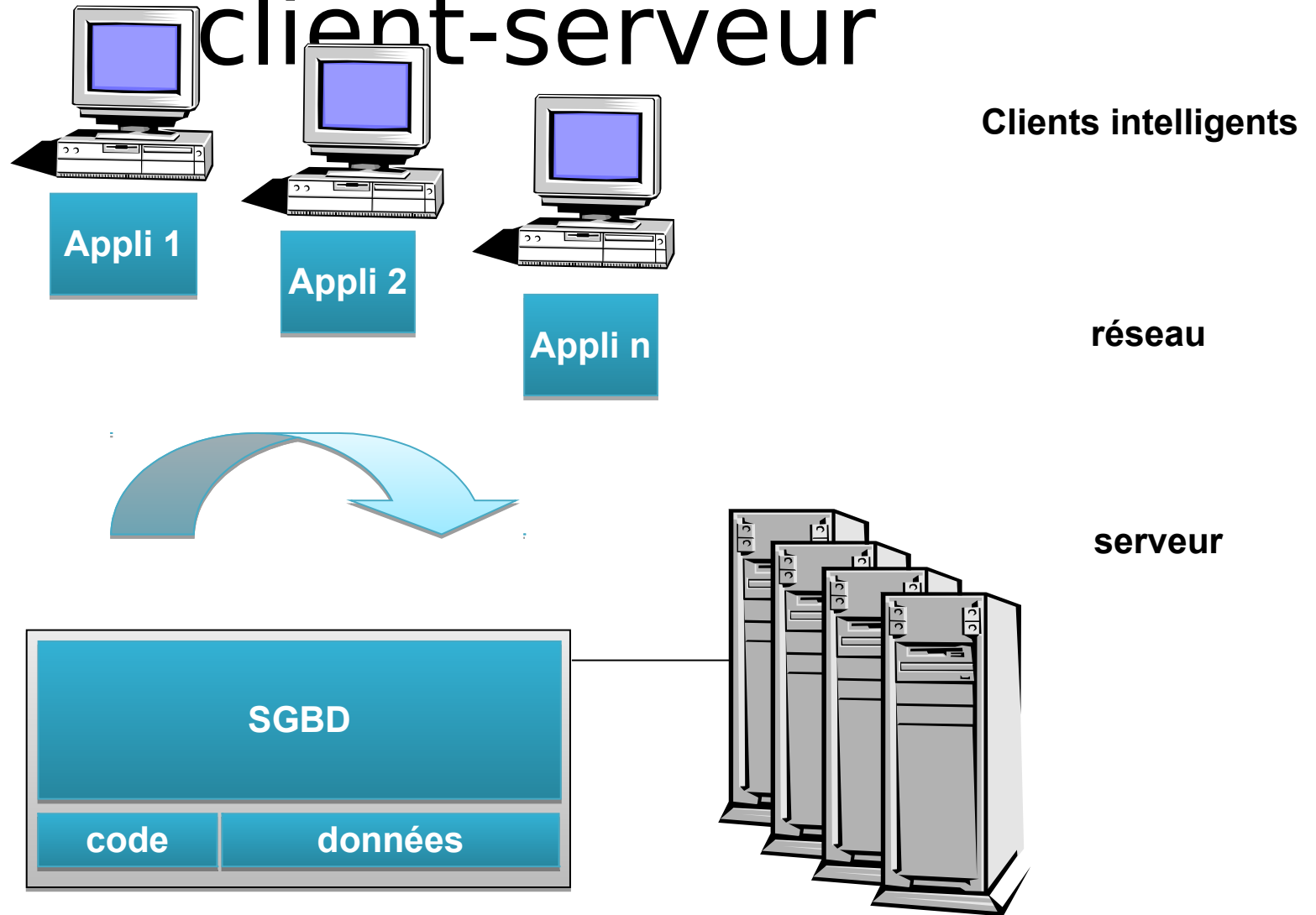
👉 Le challenge se déplace des Péta-bases aux Pico-bases.

- Péta-bases => parallélisme et grandes mémoires
- Pico-bases => faible empreinte et forte sécurité

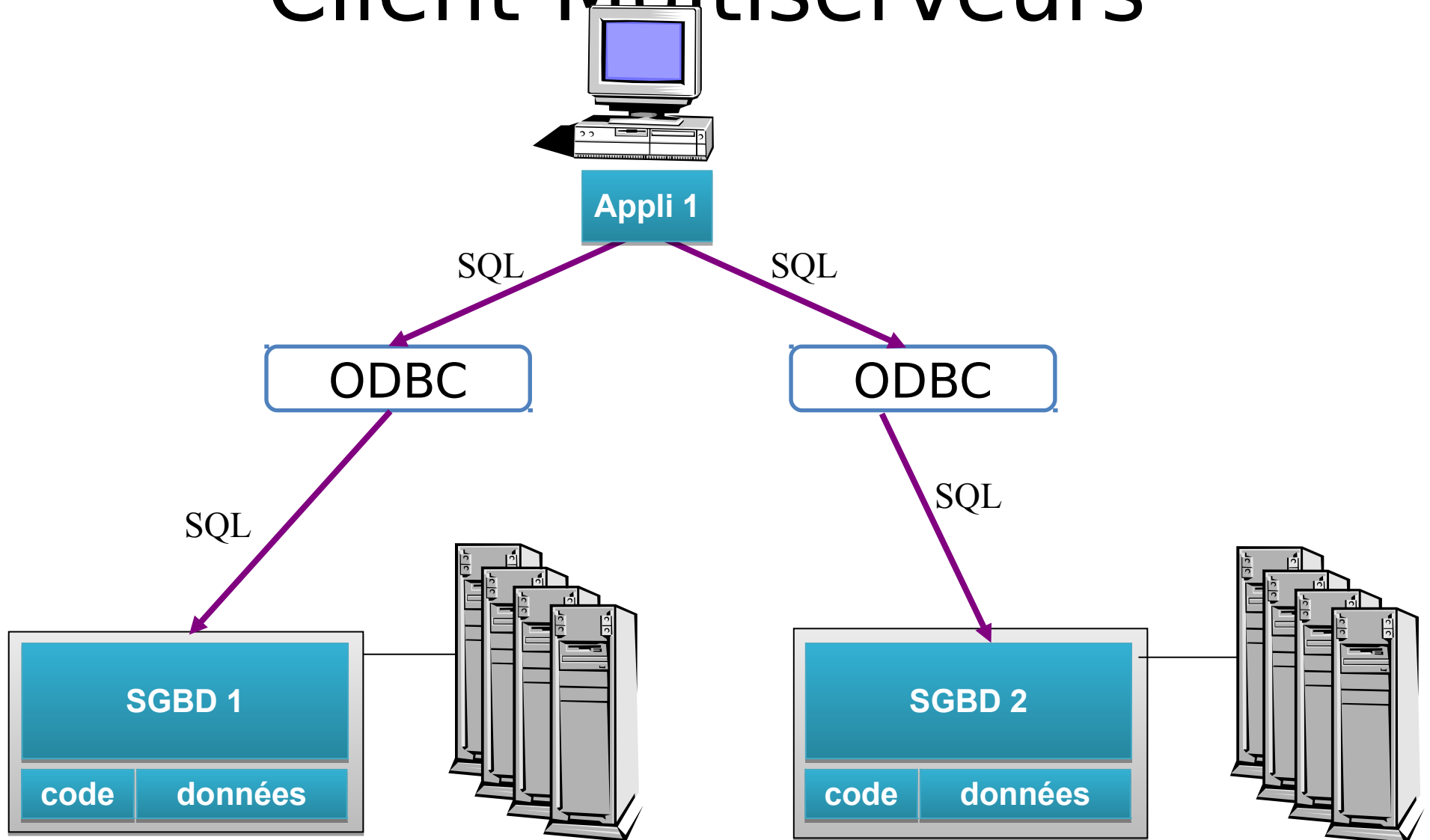
Architecture centralisée



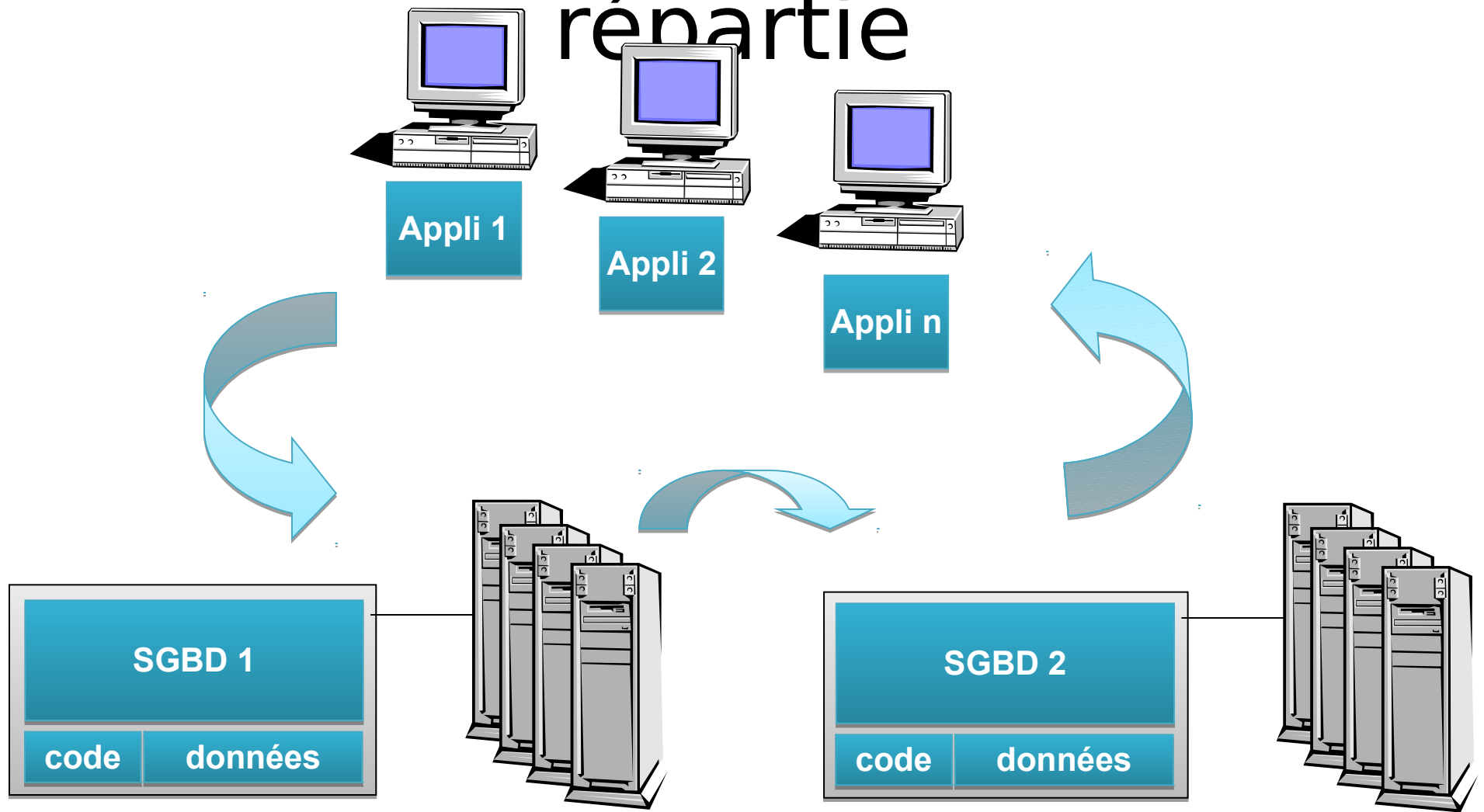
Architecture client-serveur



Architecture Client-Multiserveurs



Architecture répartie



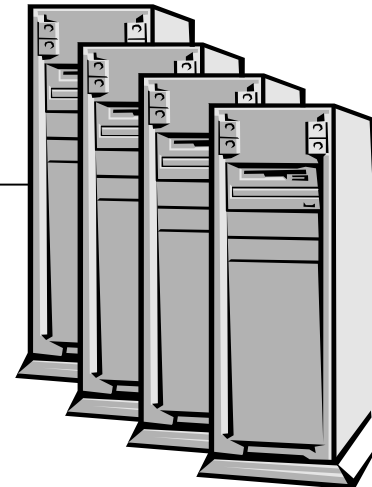
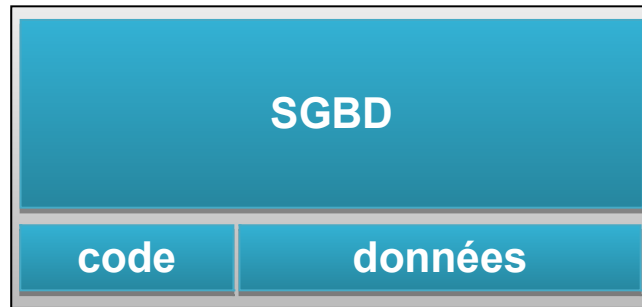
Architecture mobile



**Clients intelligents
mobiles**

**Données répliquées
et/ou personnelles**

Réseau sans fil











serveur

4. Applications traditionnelles des SGBD

- **OLTP (On Line Transaction Processing)**
 - **Cible des SGBD depuis leur existence**
 - **Banques, réservation en ligne ...**
 - **Très grand nombre de transactions en parallèle**
 - **Transactions simples**
- **OLAP (On Line Analytical Processing)**
 - **Entrepôts de données, DataCube, Data Mining ...**
 - **Faible nombre de transactions**
 - **Transactions très complexes**

Evolution des BD

	BD d'entreprise	BD personnelles	BD 'light' (PDA / Tél.)	PicoDBMS carte à puce
Capacité				
Prix				
Nombre	