

Bases de Données

S5 – INFO, DL BI

Tr. d'après Georges Gardarin

<http://www.gardarin.org/>

<http://e-campus2.uvsq.fr>

Administratif

- Cours magistral (à partir du 13 sept.)
 - Mercredi 9h45-11h15, Amphi B
- Responsable du cours : Iulian SANDU POPA (iulian.sandu-popa@uvsq.fr)
- Groupes de TD (à partir de la semaine du 25 sept.)
 - G1 (bât. Buffon, (pour le moment) salle RC9) : mardi 13h30-16h45
 - G2 (bât. Descartes, salle Alsace) : mercredi 13h30-16h45
- Chargés de TD :
 - G1 : Laurent Yeh
 - G2 : Iulian Sandu Popa

Plan du cours

- **Bases de données**
 - Introduction (1 cours)
 - Modèle Entité-Association (1/2 cours)
 - Modèle et algèbre relationnelle (1/2 cours)
 - Le langage SQL (3-4 cours)
 - Vues (1/2 cours)
 - Contrôle d'accès (1/2 cours)
- **Applications Web**
 - Introduction au Web et à la conception d'un site web (1/2 cours)
 - Le langage HTML (1/2 cours)
 - Le langage PHP (2-3 cours)
- Nous ne parlerons (presque) pas de l'optimisation !

Support de cours conseillé

- Bases de données, Eyrolles, Georges Gardarin
- Database System Concepts, Abraham Silberschatz, Henry F. Korth, S. Sudarshan
- Database Systems: The Complete Book, Hector Garcia-Molina, Jeff Ullman, Jennifer Widom

Aspects pratiques

- **TD sur papier et sur machine avec MySQL/HTML/PHP**
 - Autres BDs : PostgreSQL, Oracle, SQL Server, IBM DB2 ...
 - Cartable numérique ou ordinateur personnel
- **Evaluation**
 - CC (40%) : partie SQL (1-2 contrôles) (50%) + mini-site web (MySQL, HTML, CSS, PHP) (50%)
 - Examen final (60%)
- **Support de cours et TD disponible sur le Web**
 - <http://e-campus2.uvsq.fr/>
 - IN507 Bases de Données

Introduction

Les BD au quotidien

Dans la « vie de tous les jours »

- Site <http://www.transilien.fr/>

Horaires et Itinéraires

Rechercher mon horaire

Fiches horaires

Etat du trafic

Guide du Voyageur

Services +

Citylien

Plus sur Transilien

Mon compte

Identifiant

Mot de passe

Mémoriser

> Mot de passe oublié

> S'inscrire

> En savoir +

Vos déplacements en Ile-de-France

Obtenez pour vos itinéraires vos feuilles de route et fiches horaires.

Accueil > Horaires et Itinéraires

RECHERCHER MON HORAIRE

Point de départ et d'arrivée : indiquez le nom d'une gare, d'une station de métro ou d'une commune ou adresse postale complète (rue + ville) dans toute l'Ile-de-France (Horaires disponibles jusqu'au 27 septembre 2008)

Point de départ : paris saint lazare

adresse, arrêt, site, rue ...

Ville de départ : paris

Point d'arrivée : versailles montreuil

adresse, arrêt, site, rue ...

Ville d'arrivée : versailles

Gagnez du temps : enregistrez vos trajets les plus courants : créez votre espace personnalisé !

+ D'INFOS

Date: 8 Septembre 2008

Aujourd'hui


Heure: 13 H 00

de départ d'arrivée

Mode de transport: Tous modes

CHERCHER + D'OPTIONS AIDE ?

Dans la « vie de tous les jours »



Sommaire ↓

Horaires et Itinéraires

➤ **Rechercher mon horaire**

Fiches horaires

Etat du trafic

Guide du Voyageur

Services +

Citylien

Plus sur Transilien

Mon compte ↓

Identifiant

Mot de passe


Mémoriser ☐ OK

> Mot de passe oublié

> S'inscrire > En savoir +

HORAIRES ET ITINÉRAIRES

Vos déplacements en Ile-de-France
Obtenez pour vos itinéraires vos feuilles de route et fiches horaires.



Accueil > Horaires et Itinéraires > Rechercher mon horaire

➤ RECHERCHER MON HORAIRE

Point de départ et d'arrivée : indiquez le nom d'une gare, d'une station de métro ou d'une commune ou adresse postale complète (rue + ville) dans toute l'Ile-de-France (*Horaires disponibles jusqu'au 27 septembre 2008*)

Plusieurs réponses ont été trouvées, précisez dans la liste ci-dessous votre demande :

Point de départ :

----- Gares, stations, arrêts -----

GARE ST LAZARE - Paris

HAUSSMANN ST LAZARE - Paris

----- Lieux remarquables -----

Hôpital/Polyclinique Hôpital Saint-Lazare - Paris

Hôpital/Polyclinique Hôpital Saint-Lazare (Ap-Hp) - Paris

▲

▢

▢

▢

▼

Plusieurs réponses ont été trouvées, précisez dans la liste ci-dessous votre demande :

Point d'arrivée :

----- Gares, stations, arrêts -----

GARE DE MONTREUIL - Versailles

GARE DE VERSAILLES CHANTIERS - Versailles

Montreuil - Versailles

----- Lieux remarquables -----

Hôpital/Polyclinique Centre Hospitalier de Versailles - Versailles


▲

▢

▢

▢

▼

Date: 8 ▾ Septembre 2008 ▾ 

Dans la « vie de tous les jours »



Sommaire ↓

Horaires et Itinéraires

➤ **Rechercher mon horaire**

Fiches horaires

Etat du trafic

Guide du Voyageur

Services +

Citylien

Plus sur Transilien

Mon compte ↓

Identifiant

Mot de passe

Mémoriser ☐ OK

> Mot de passe oublié

> S'inscrire > En savoir +

HORAIRES ET ITINÉRAIRES

Vos déplacements en Ile-de-France
Obtenez pour vos itinéraires vos feuilles de route et fiches horaires.



Accueil > Horaires et Itinéraires > Rechercher mon horaire > Résultats de la recherche

➤ RÉSULTATS HORAIRES ET ITINÉRAIRES

Départ		GARE ST LAZARE - Paris			
Arrivée		GARE DE MONTREUIL - Versailles			
le		8 Septembre arrivée à 13:00			
Départ	Arrivée	Durée	Correspondance	Mode(s) de transport	Détail
12:17	12:42	25 min	0		
12:32	12:57	25 min	0		
12:47	13:12	25 min	0		

< Horaires précédents

Horaires suivants >

AUTRE RECHERCHE




TRAJET RETOUR

AIDE ?

Dans la « vie de tous les jours »

[Accueil](#) > [Horaires et Itinéraires](#) > [Rechercher mon horaire](#) > Feuille de route

FEUILLE DE ROUTE

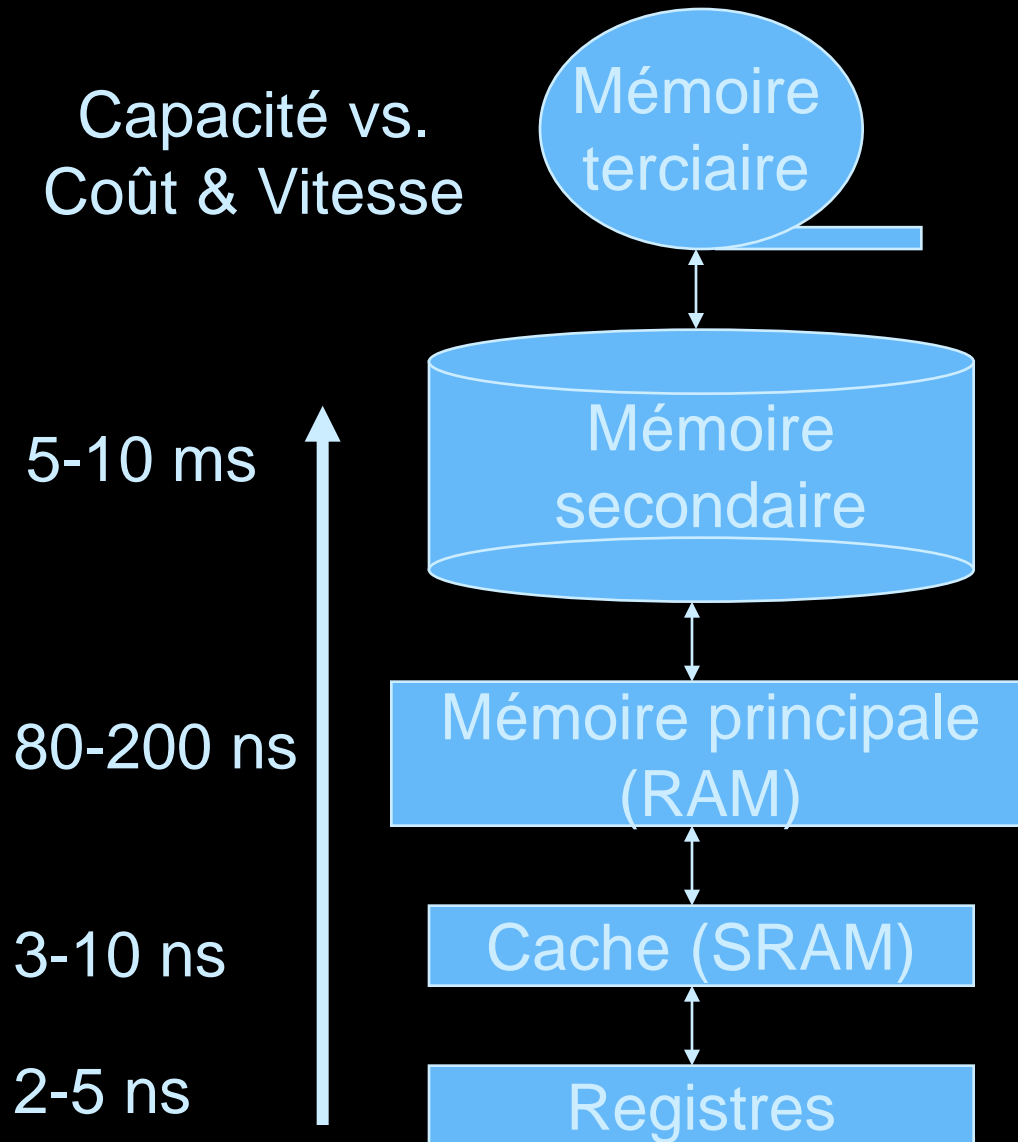
Départ	GARE ST LAZARE - Paris				
Arrivée	GARE DE MONTREUIL - Versailles				
le	8 Septembre arrivée à 13:00				
Heure	Mode	Lieu		Durée	Plan
12h32		De l'arrêt GARE ST LAZARE (Paris)	SNCF	25 min	
12h57		Prendre le Train VALE Direction GARE DE VERSAILLES RIVE DROITE jusqu'à GARE DE MONTREUIL (Versailles)			
Durée totale du trajet : 25 min					
< Hor. Précédents		Les horaires de votre recherche		Hor. Suivants >	

[NOUVELLE RECHERCHE](#)
[AJUSTER VOTRE RECHERCHE](#)
[PRIX DES TITRES](#)
[AIDE ?](#)
[ENVOYER À UN AMI](#)
[IMPRIMER](#)

1. Introduction

- Les entreprises gèrent des volumes de données très grands
 - Giga, Terra, Péta –octets
 - en 2014 Google stockait plus de 10 exabytes (1 exabyte = 1 million de terrabytes)
 - Numériques, textuelles, multimédia (images, films,...), capteurs (GPS, température, pollution ...)
- Il faut pouvoir facilement
 - Archiver les données sur mémoires secondaires permanentes
 - 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:
 - Récipients logique 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
- Années 90:
 - Orientation décisionnelle (Data mining, OLAP)
- Années 2000:
 - Avènement du Web (Alta-vista, yahoo!, google...)

BD sur le Web

- Archivage et recherche de données multimédias
 - Texte (livres, articles, journaux, ...)
 - Images (instagram, flickr)
 - Films (youtube, etc.)
 - Données géographiques (cartes 2D, 2.5 D – OpenStreetMap, google maps)
 - Données spatiales (3D), temporelles et spatio-temporelles
- Recherche par proximité
 - Textes : liste de mots-clés (à la Google)
 - Images : par proximité (couleur, forme, texture ...)
 - Cartes : par rectangle englobant, distance, zoom

Systemes de fichiers

Caractéristiques¹⁵



Comptabilité



Chirurgie

Consultations



Psychiatrie



Problèmes

Format des fichiers



Dupont

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

Dupond

Turlututusqjsk
Symptom: yyyy
Analyses xxxx

Turlututudhjsd
Analyses :xx



Duhpon

Symptomes : yy
Analyses : xxxx

Symptomes : yy

Duipont

Turlututu : sq

Symptomyyyy
Analysesxxxx

Turlututudhjsd



Caractéristiques

16

Plusieurs applications

- plusieurs formats
- plusieurs langages

Problèmes

- Difficultés de gestion

Redondance (données)

Caractéristiques

17

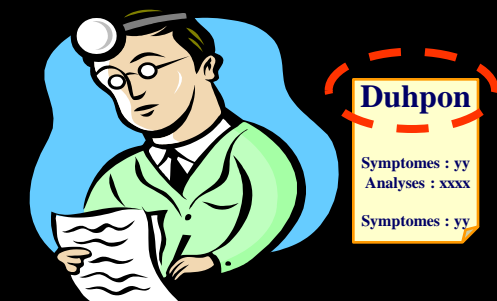
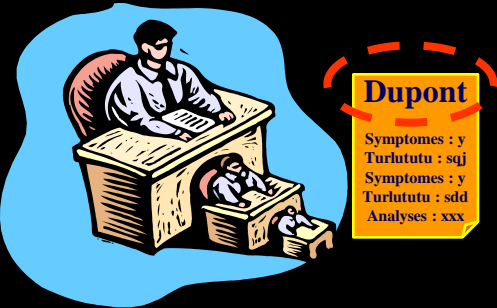
Plusieurs applications

- plusieurs formats
- plusieurs langages

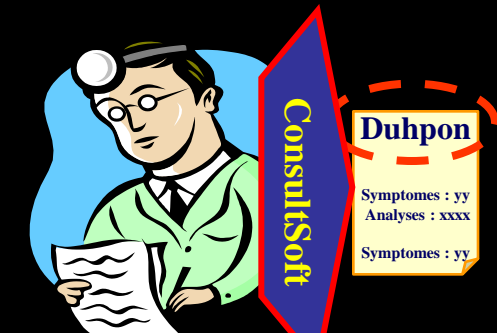
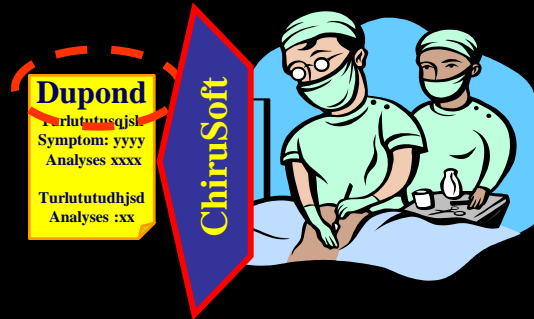
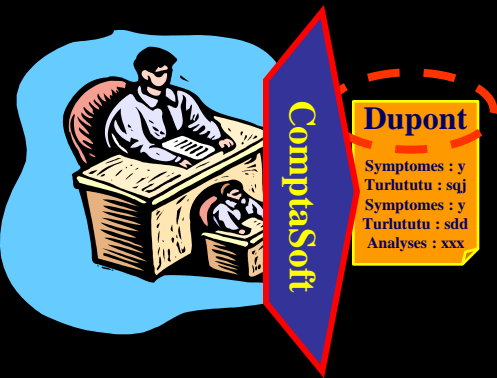
Redondance de données

Problèmes

- Difficultés de gestion
- Incohérence des données



Interrogations



Caractéristiques

18

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

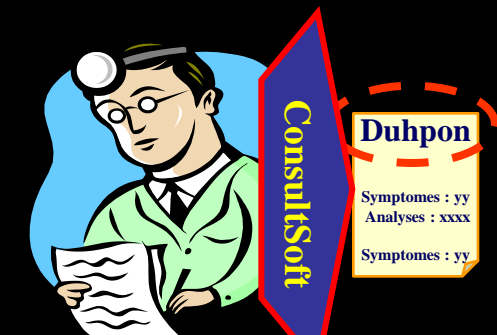
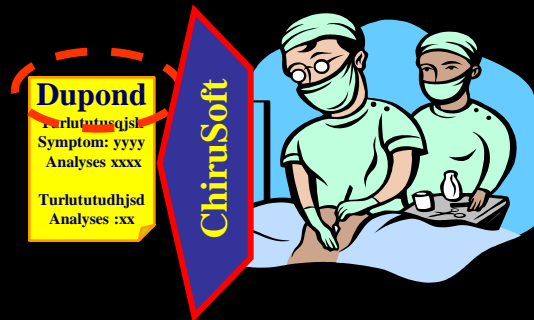
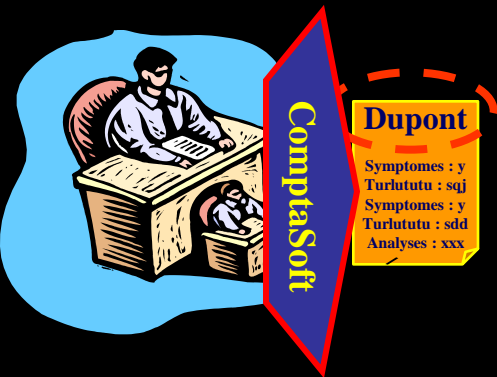
Pas de facilité d'interrogation

- Question ⇒ développement

Problèmes

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

Pannes ???



Caractéristiques

19

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

Pas de facilité d'interrogation

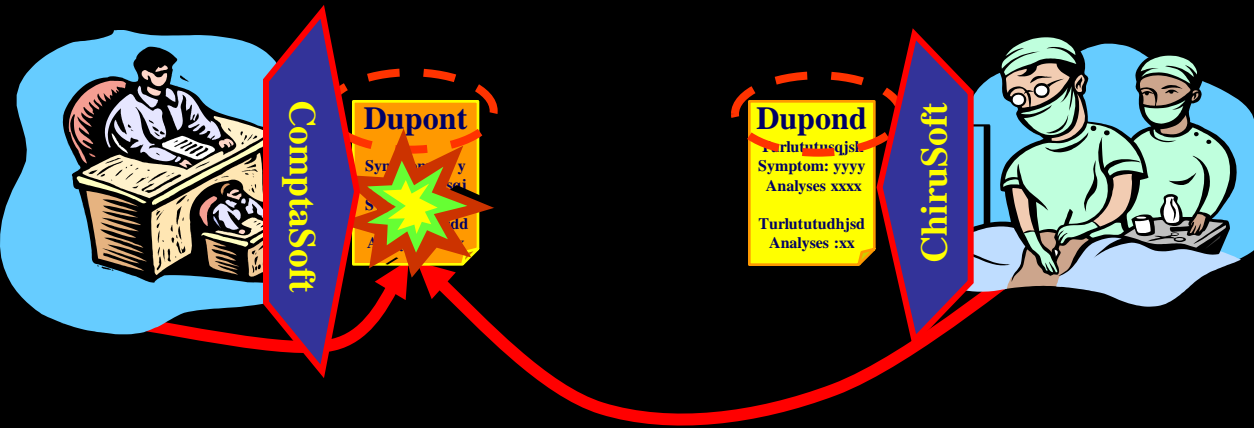
- Question ⇒ développement

Redondance de code

Problèmes

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

Partage de données



Caractéristiques

20

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

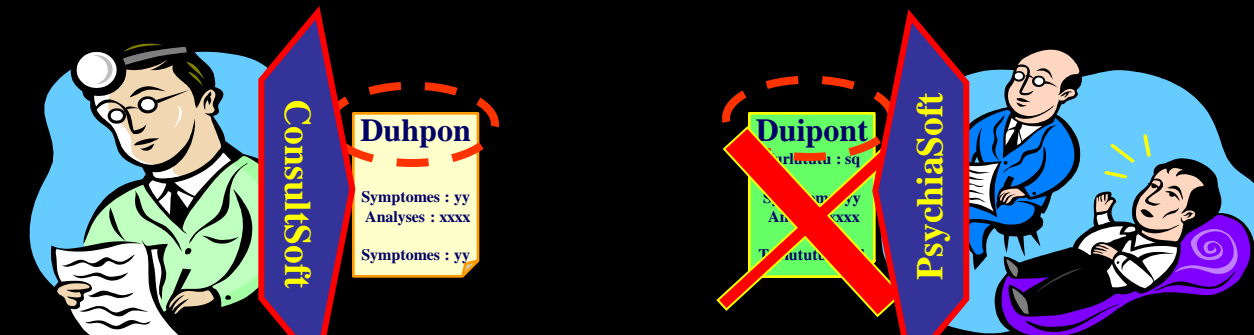
Pas de facilité d'interrogation

- Question \Rightarrow développement

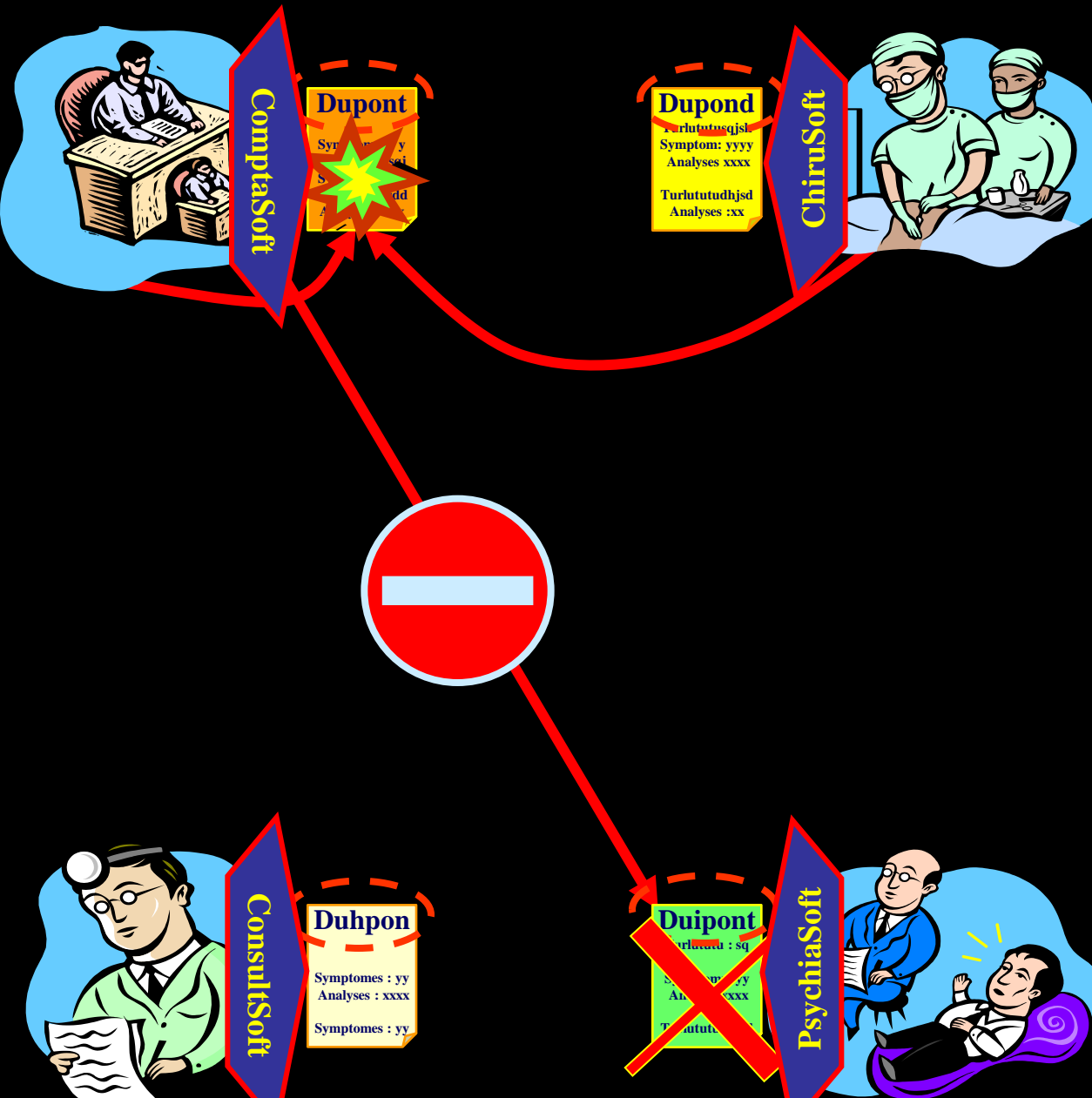
Redondance de code

Problèmes

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



Confidentialité



Caractéristiques

21

Plusieurs applications

- plusieurs formats
- plusieurs langages

Redondance de données

Pas de facilité d'interrogation

- Question \Rightarrow développement

Redondance de code

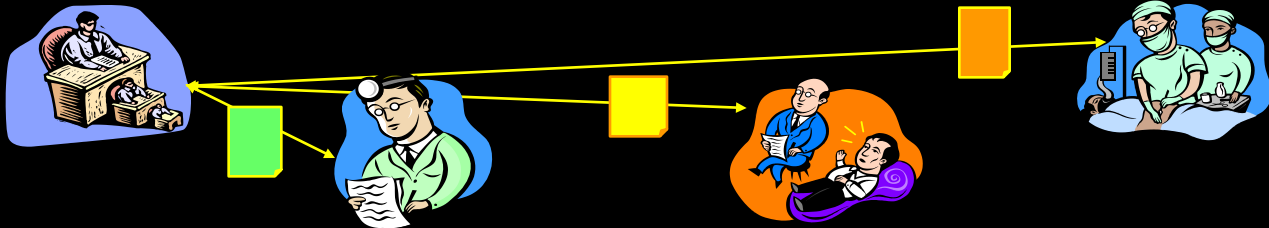
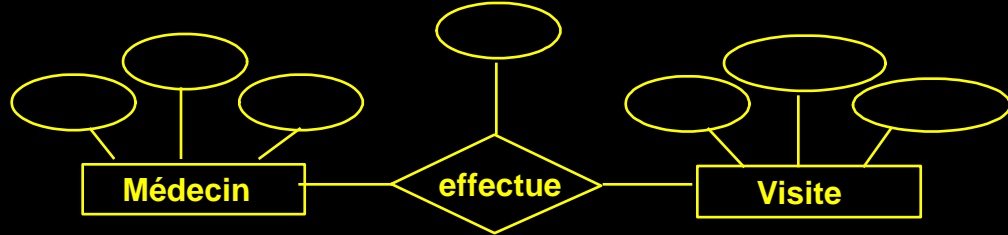
Problèmes

- 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**
 - ➔ Eliminer 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)			

Modélisation Relationnelle (1)

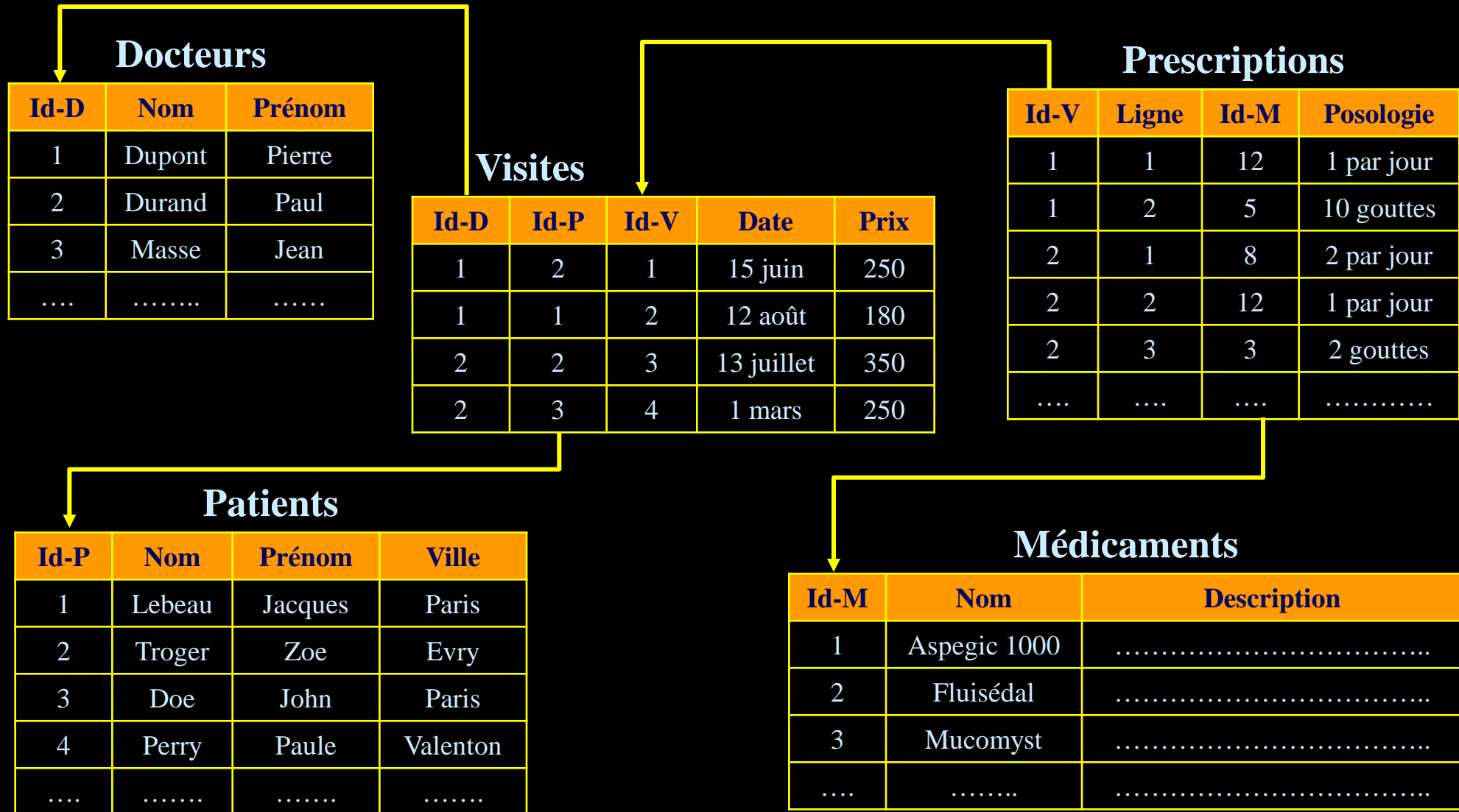
Relation ou table

Champs, attributs,
colonnes

Id-D	Nom	Prénom
1	Dupont	Pierre
2	Durand	Paul
3	Masse	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. Masse

Visites

Id - D	Id - P	Id - V	Date	Prix
1	2	1	15 juin	250
2	3	4	1 mars	250

Prescription

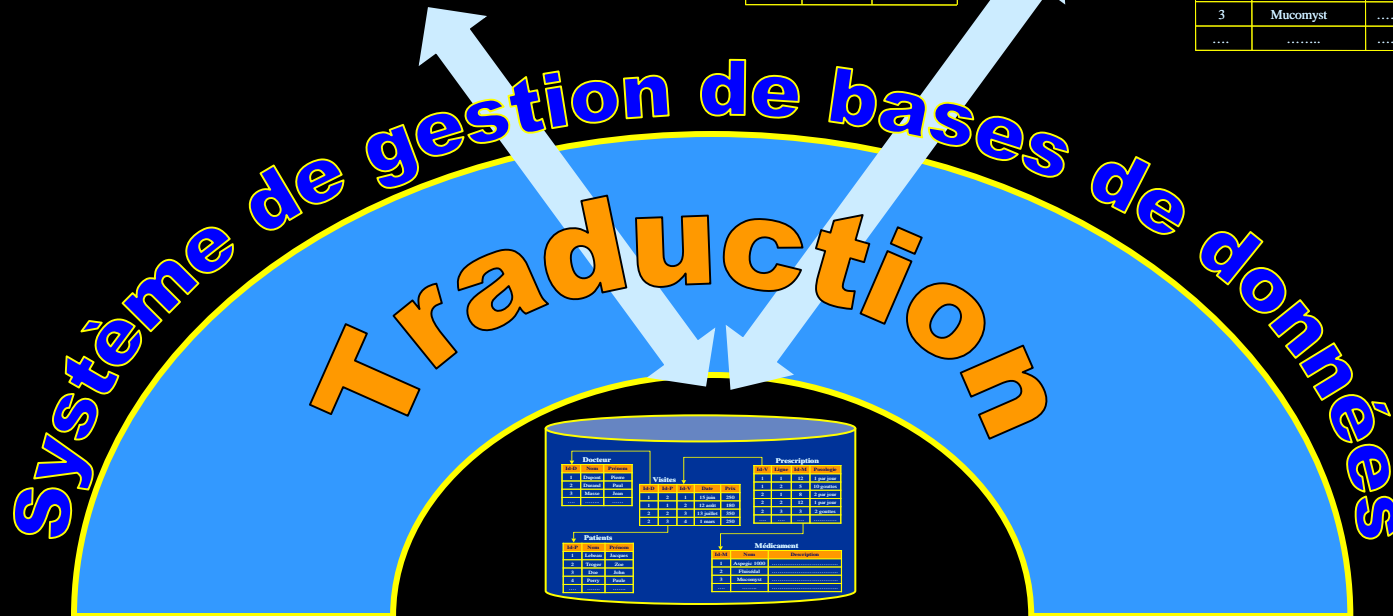
Id - V	Ligne	Id - M	Posologie
1	1	12	1 par jour
1	2	5	10 gouttes
....

Patients

Id - P	Nom	Prénom
1	Lebeau	Jacques
2	Troger	Zoe
....

Médicament

Id - M	Nom	Description
1	Aspegic 1000
2	Fluisédal
3	Mucomyst
....



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:
 - $\text{Prix de vente} > \text{Prix d'achat}$
 - Relations entre objets:
 - Un électeur doit être inscrit sur une seule liste électorale

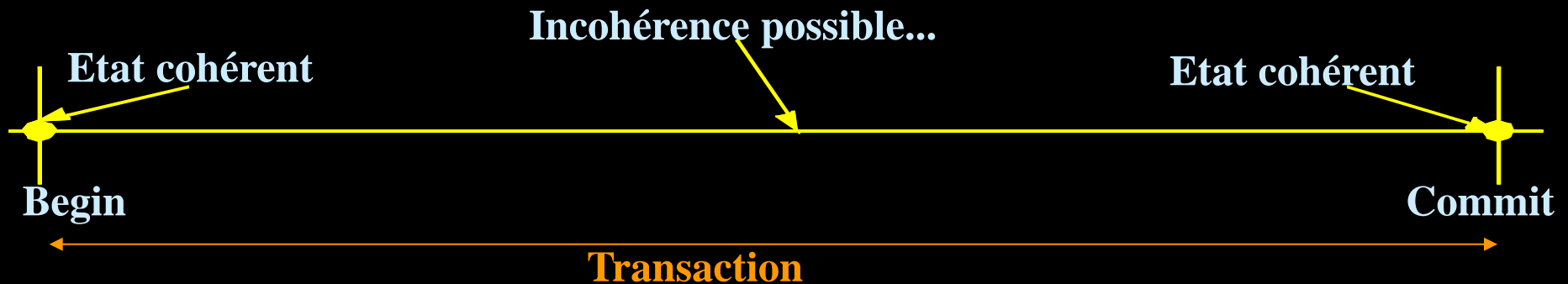
Contraintes d'intégrité

- **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

$CEpargne = CEpargne - 3000$

$CCourant = CCourant + 3000$

Commit T1

Atomicité et Durabilité

ATOMICITE

Begin

$CEpargne = CEpargne - 3000$

$CCourant = CCourant + 3000$

Commit T1

Panne

→ Annuler le débit !!

DURABILITE

Begin

$CEpargne = CEpargne - 3000$

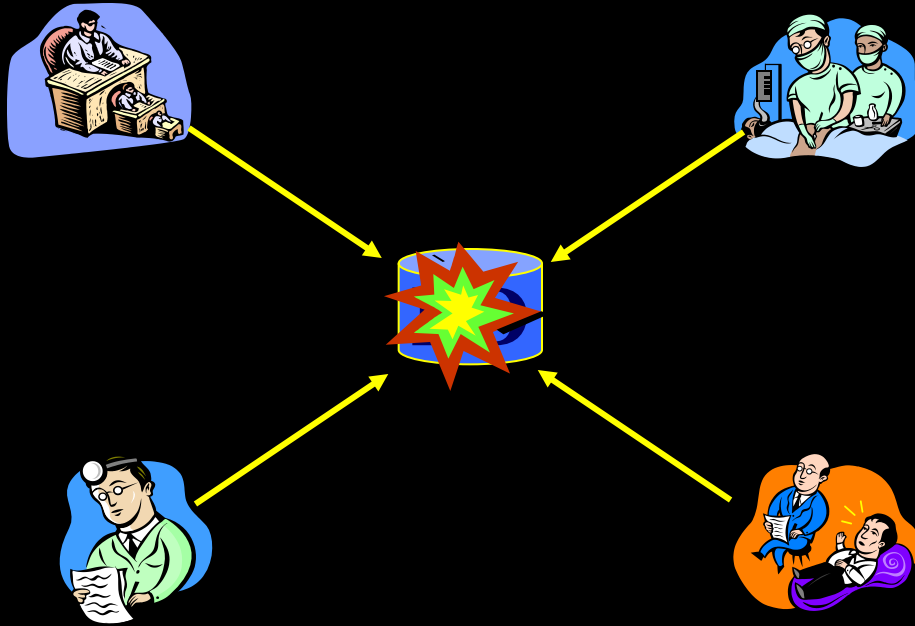
$CCourant = CCourant + 3000$

Commit T1

Crash disque

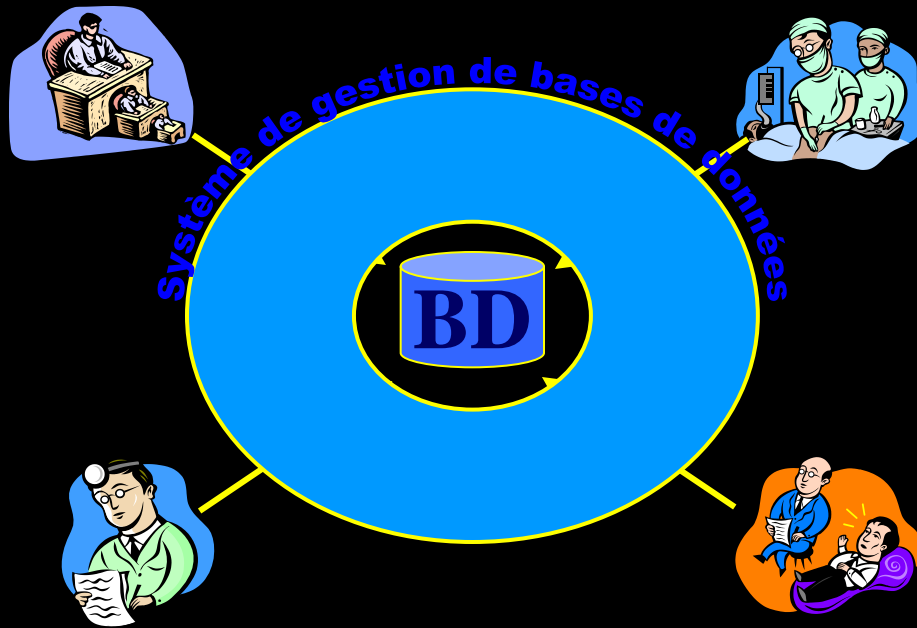
→ S'assurer que le virement a été fait !

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...