

# Introduction aux bases de données et aux SGBD

Bernard ESPINASSE

Professeur à l'Université d'Aix-Marseille

- Généralités sur les BdD
- Des chaînes de traitements aux applications sur BdD
- Niveaux de description d'une BdD : le rapport ANSI/SPARC
- Qu'est ce qu'un SGBD :
  - historique,
  - objectifs,
  - architectures

## Généralités sur les bases de données

banque de données ≠ bases de données

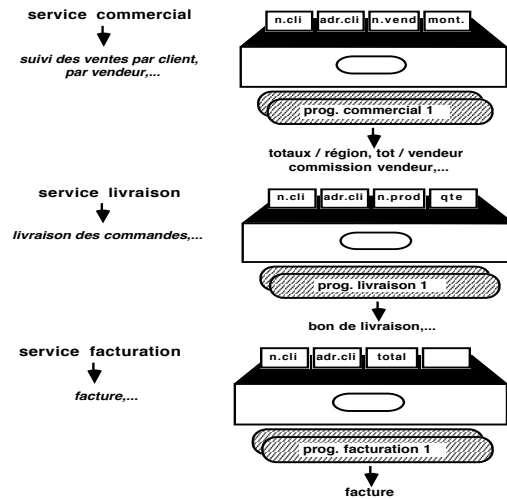
### Banque de données :

Ensemble de données, relatif à un domaine défini de connaissances, organisé pour être offert en consultation aux utilisateurs (producteur, serveur, transporteur, utilisateur).

### Base de données :

Ensemble structure de données enregistrées sur des supports accessibles par ordinateur, pour satisfaire un ou plusieurs utilisateurs, de façon sélective et en un temps opportun.

**Cas:** Soit une entreprise dont l'activité est assurée par plusieurs services ayant ses propres fichiers:

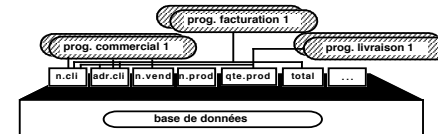


## Problèmes

liés à cette organisation des informations

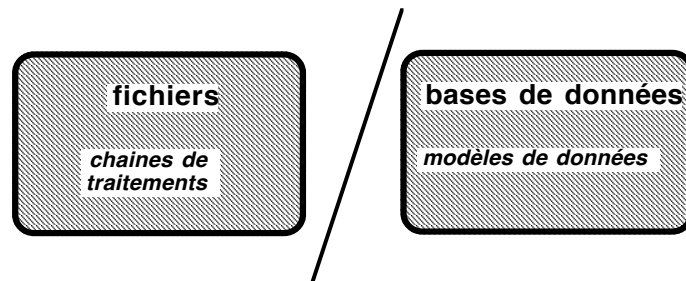
- un client change d'adresse ... → **redondance des données**
- litige entre un client et le service facturation sur le prix d'un produit : nom du vendeur non disponible (N.VEND) → **non partage des données**
- supposons que le programme d'application Facturation 1 doit être modifié (remise) et que le fichier facturation doit être restructuré tous les pgm utilisant ce fichier doivent être changés... (réaction en chaîne) → **non évolutivité**

**Solution :** placer toutes les informations dans un même système !

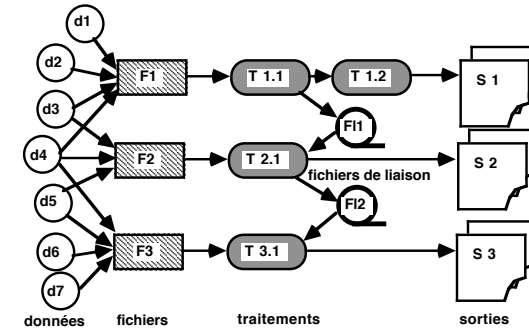


- **non redondance,**
- **partage des données,**
- **évolutivité,**
- les **pgm n'utilisent que les données nécessaires** (séparation programme/données)

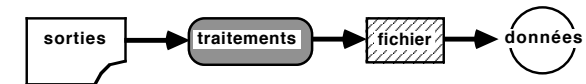
## 2 démarches fondamentales différentes :



## informatique "classique" : les chaînes de traitements



les fichiers sont définis pour chaque traitements à partir des sorties demandées



## Problèmes liés aux chaînes de traitements :

### • redondance et inconsistance des données :

- infos répliquées dans plusieurs fichiers
- volume des fichiers exploités importants
- inconsistance des données mémorisées
- données stockées sous différents formats (époques différentes): accès difficile

### • multiplicité des mises à jour :

- accès concurrents en mises à jour → besoin d'un **superviseur**, difficile à écrire (prog. et fichiers époques différentes)

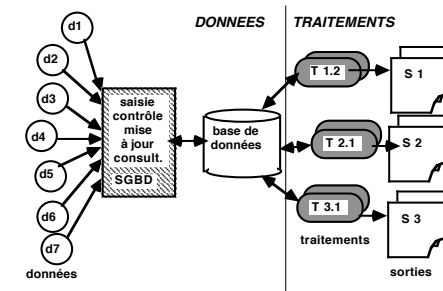
### • sécurité et confidentialité :

- confidentialité assurée par les programmes → **verrouillage** personnalisé dans tous les programmes

### • intégrité des données :

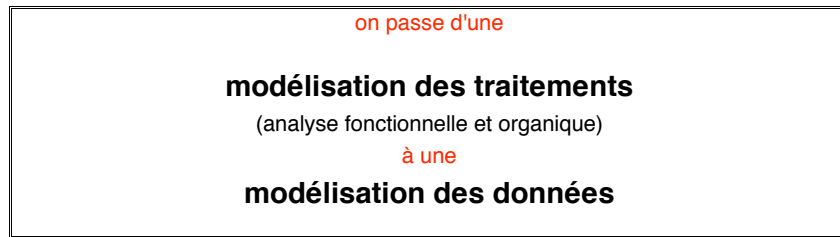
- données doivent respecter des contraintes lors de la saisie → développement dans les programmes de **modules spécifiques...**

## Applications autour d'une base de données



- **saisie unique** : non redondance, moindre coût, moins de support
- **mise à jour unique** : moins de maintenance, moins de risque d'incohérences
- **accès facile à l'information stockée** : par les pgm d'application, indépendant des pgm d'application
- **partage des données entre application**
- **évolutivité**

## Applications autour d'une base de données

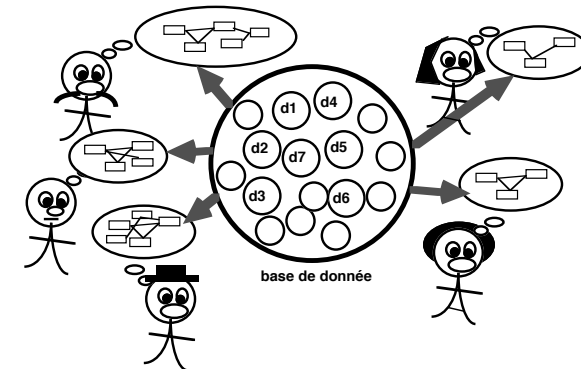


### modèles de données:

= combinaison :

- de **types de structures de données**
- d'**opérateurs ou règles d'inférences** sur ces structures
- de **règles d'intégrité** définissant le jeu des états ou changements d'états admis de données

## Conception d'une BdD : difficultés...



- une BdD = **collection de données** dont sont tirées **différentes vues des utilisateurs finaux**.
- conception d'une BdD = **saisir les points de vue des utilisateurs et les synthétiser dans une structure de données**.

## Différents niveaux de description d'une BdD (1)

### Le rapport ANSI/SPARC (1975)

(standards planning and requirements committee (SPARC) of the American National Standard Committee  
X3/SPARC study group on  
Database systems (x3 Project 226)

- rapport interim 1975
- publié dans "ANSI/SPARC DBMS MODEL", D.A.Jardine Ed. 1977, North Holland. (ANSI/SPARC)

### 1• Recherche de niveaux, étapes, modèles :

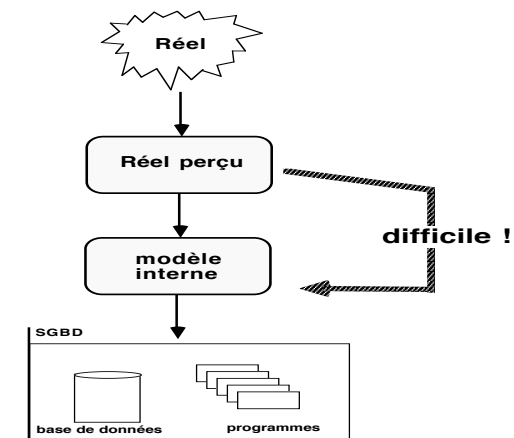
que devraient comporter le processus de structuration d'une BdD

→ **émergence d'un niveau intermédiaire le niveau CONCEPTUEL**

### 2• Définition des interfaces entre les différents modèles et acteurs du processus :

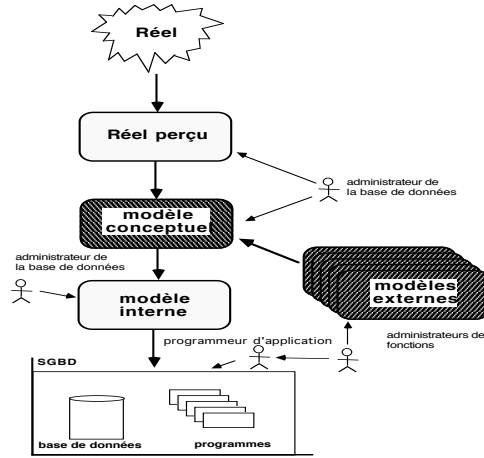
→ **architecture générale des SGBD**

## Différents niveaux de description d'une BdD (2)



- conception de BdD : très grande **difficulté** à **décrire le réel perçu** dans un formalisme informatique.

## Différents niveaux de description d'une BdD (3)



- émergence des niveaux Conceptuel et Externe
- rôles et niveaux d'information dans l'organisation

## Niveaux de description liés à une base de données

- **réel perçu** : représentation du réel que l'on se construit selon
  - les finalités
  - le phénomène observé

*exprimé dans le langage, le vocabulaire du domaine étudié.*
- **niveau interne** : définit dans le système informatique la réalisation de la structure de données et dépend :
  - du SGBD choisi (réseaux, relationnels, autres...)
  - d'objectifs d'optimisation

*exprimé dans le formalisme informatique lié à l'outil. formalismes CODASYL, RELATIONNEL.*
- **niveau conceptuel** : description en termes d'objets, propriétés et relations du réel perçu, permettant d'établir/domaine :
  - l'unicité de la mémorisation des informations (exprimé au niveau interne)
  - la stabilité de la structure
  - la rencontre gens du domaine/informaticiens

*formalisme: nature/gens du domaine et rigoureux/informatique*
- **niveau externe** : description en terme d'objets, propriétés et relations d'une utilisation particulière d'information du domaine.
  - la validation d'un modèle externe.

*même formalisme qu'au niveau conceptuel*

## Qu'est ce qu'un S.G.B.D. ?

**SGBD = Système de Gestion de Bases de Données :**

### 1 • un écran entre les usagers et les mémoires secondaires permettant :

- de **créer l'illusion** que les données désirées par tout usagers sont stockées sur mémoire secondaires :
  - assemblées et codées comme souhaitées
  - comme si l'utilisateur était seul à utiliser ces données

### 2 • un outil informatique permettant :

- d'**insérer, modifier efficacement les données** spécifiques dans une grande masse
- **rechercher des informations partagées** par tous les usagers (milliards d'octets) giga octets

## Historique des S.G.B.D.

**1960** premiers développements des systèmes de fichiers

### 1965 première génération de SGBD : (années 70)

- séparation de la description des données des programmes d'applications
- avènement de langages d'accès navigationnels (structures de graphes)
  - recommandations CODASYL
  - système IMS (IBM) / IDMS
  - **SGBD RESEAUX,...**

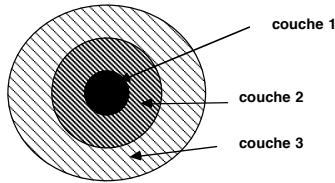
### 1970 deuxième génération de SGBD : (années 80, 90)

- enrichir le SGBD pour faciliter d'accès des usagers aux données
- langage assertionnels basés sur la logique spécifier les données souhaitées sans dire comment les accéder
  - modèle relationnel (TED CODD)
  - **SGBD RELATIONNELS**

### 1990 troisième génération de SGBD (années 2000...)

- langages d'accès plus puissants, plus naturels
- types de données variées
- possibilités de déduction (Bd déductives...)

## Structure fonctionnelle d'un S.G.B.D



### couche 1

- gestion des récipients de données sur mémoire secondaire : système de gestion de fichiers (fonctions de base)

### couche 2 : SGBD interne

- gestion des données stockées dans les fichiers
- placement, assemblage de ces données
- gestion des liens entre données et structures de recherche rapide (index)

### couche 3 : SGBD externe

- présentation des données aux programmes d'applications et aux usagers ayant formulé leurs besoins en langage ± élaborés (requêtes, rapports, L4G...)

## Objectifs des S.G.B.D.

### • orientés données :

- non redondance des données
- partageabilité des données
- sécurité des données
- cohérence des données

### • orientés traitements :

- indépendance physique des données
- indépendance logique des données
- manipulation facile des données
  - par informaticien
  - par non informaticien
- cohérence physique (pannes, ...)

### • organisationnels :

- administration centralisée des données

## Objectifs orientés données

### • non redondance des données :

avec une approche BdD, les fichiers ± redondants seront intégrés en un seul fichier ou plusieurs fichiers contenant des données distinctes.

- mémoire secondaire réduite
- effort de saisie et mise à jour

### • partageabilité des données :

permettre le partage des données de la base à plusieurs applications, usagers, simultanément : **problème des accès concurrents**

### • sécurité des données :

- les données doivent être protégées contre les accès
  - non autorisés, mal intentionnés
  - les contrôler (consultation, mise à jour...)

### • cohérence des données :

les données sont soumises à certaines règles :

- niveau élémentaire exemple :  $salaire\ 4.000 \leq S \leq 10.000$
- niveau d'ensemble : *une commande doit être rattachée à un client et un seul.*

le SGBD doit vérifier que les applications respectent ces règles ... contraintes d'intégrités... : connaissances sur les données, leur signification.

## Objectifs orientés traitements (1)

- **indépendance physique des données** : organisation physique des données, pb informatique : performance, flexibilité d'accès

- le changement d'organisation physique ne doit pas être répercuté sur l'exploitation par les applications de la base

- **indépendance logique des données** : organisation "conceptuelle" synthétique et notion de vue (user-view) VIEW

**VEHICULE (N°VEH, MARQUE, TYPE, COULEUR) ; PERSONNE (IDENT, NOM, PRENOM)**

**PROPRIETAIRE (IDENT, N°VEH, DATE)**

**vue 1 Personne (Ident, prénom, nom, n° de voiture) : personne possédant une voiture**

**vue 2 voiture (n° véh., type, marque, date vente) : véhicules vendus à une certaine date**

## Objectifs orientés traitements (2)

- **manipulation facile des données : pour les non-informaticiens**
  - pouvoir les consulter, les interroger, les mettre à jour
  - langages "non procéduraux" : décrire les données que l'on souhaite retrouver (ou mettre à jour) sans décrire la manière de les retrouver dans la base: SQL,...
  - langages procéduraux de haut niveau : permettent d'élaborer de petites applications eux mêmes...
- **manipulation facile des données : pour les informaticiens**
  - langage de développement de haut niveau
  - manipulation en langage hôte : COBOL, PASCAL, C, PLI

... langages dits de « 4ème génération » ...

## Objectifs orientés traitements (3)

- **cohérence physique/fiabilité :**

Le SGBD doit être capable de retrouver un état cohérent de la base de données après une panne : **mécanisme de reprise après panne**

  - **panne de mémoire secondaire (cas le plus rare)**

informations sur disque perdues  
→ *reprise à froid : archive sur bande + journal des modifs des données (journal des images avant et après modifs)*
  - **panne système :**

informations en mémoire centrale perdues  
→ *reprise à chaud : utilise le journal + copie des pages sur disque (pages ombres dans SQL/DS)*
  - **panne de transaction :**

toutes les modifs faites par la transaction doivent être défaites

    - demandé par usager : RESTORE (SQL/DS)
    - automatique système : cas de dead lock (concurrence)  
→ *utilise le journal*

## Objectifs organisationnels

- **administration centralisée des données :**
  - permettre un contrôle efficace des données,
  - résoudre conflits entre divers points de vue d'utilisateurs,
  - optimisation des accès aux données,
  - optimisation des moyens informatiques
- **fonctions**
  - d'Administrateur des Données
  - d'Administrateur de Base de Données
- **centralisation / décentralisation :**
  - BdD + Réseaux : distribué/partagé  
INGRES STAR, ORACLE, ...

## Administrateur de données

- **petite annonce :**

**Administrateur de Données**

**Autres intitulés :**  
- Architecte, Spécialiste bases de données.

**POSTE :**  
**Une fonction études qui concerne le contenu même des données.**  
L'administrateur de données élabore le schéma conceptuel des données de l'entreprise (définition, format... ) crée et maintient une nomenclature, sélectionne et implante le dictionnaire des données, centralise les connaissances relatives à la structure logique et à la finalité des données et des programmes.  
de formation supérieure (ingénieur informaticien ou généraliste, ou universitaire) il a une expérience de l'entreprise (de préférence comme analyste ou chef de projet) et connaît bien ses rouages ; c'est un homme d'organisation et à l'aise dans les contacts.

**Une fonction exploitation qui concerne l'architecture physique de la base.**  
L'administrateur de données assure la conception de la base, il organise sa gestion physique et celle du dictionnaire des données. Il garantit la sécurité et l'intégrité des données, et optimise le fonctionnement de la base (accès, mises à jour).  
Ingénieur ou universitaire, c'est un technicien de l'informatique qui a une expérience système ou exploitation. Il connaît les SGBD (Systèmes de Gestion des Bases de Données) et les réseaux.  
Son rôle de conseil et de garant des normes exige des qualités de rigueur, d'organisation ainsi que d'excellentes capacités relationnelles.

## Administrateur de Base de Données et/ou de Données

- très grandes entreprises
- très grandes BdD

### Administrateur de Base de Données (DBA)

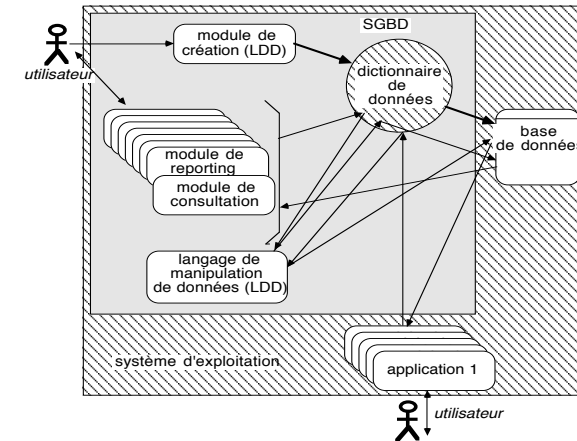
- performance
- sécurité
- partage
- formation/information
- maîtrise de l'outil (tuning, ...)
- évolution de la technologie
- ...

### Administrateur de Données (DA)

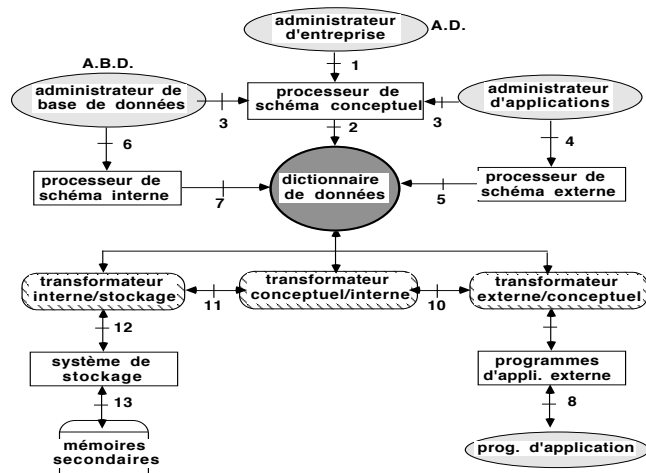
- pertinence
- sécurité
- partage
- formation/information
- diffusion
- vision globale S.I.
- évolution de l'organisation
- ...

profils différents...

## Architecture générale simplifiée



## Architecture ANSI/SPARC



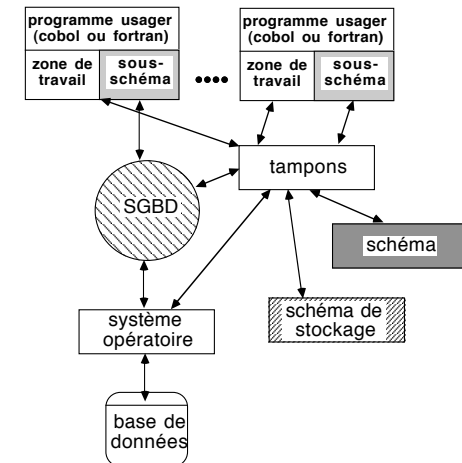
Architecture SGBD de référence Groupe ANSI/X3/SPARC (1975)

## Architecture du DBTG CODASYL

- Data Base Task Group du comité CODASYL (Cobol)
- recommandations pour la construction d'un SGBD (dès 1971)

- LDD
- LMD orientés COBOL

- **schéma** : conceptuel ANSI/SPARC (+ interne)
- **sous-schéma** : externe ANSI/SPARC
- **schéma de stockage** : interne ANSI/SPARC



## SGBD relationnels

- données organisées sous forme de **tables**
- de plus en plus **nombreux** (depuis 1980)

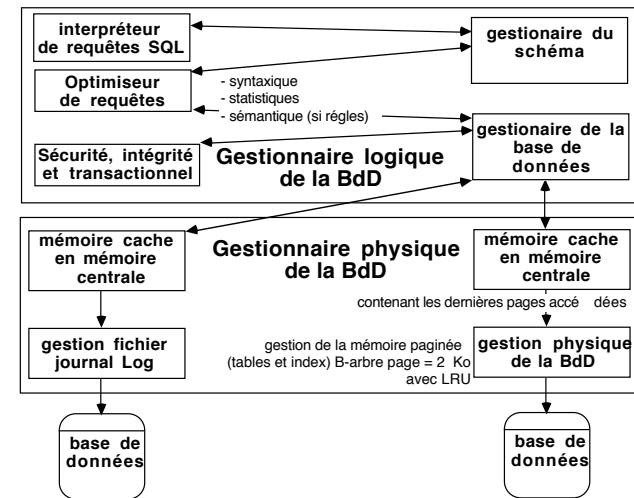
### la plupart comportent :

- un **schéma relationnel (conceptuel/interne)**  
description des tables implantés sur disque + chemins d'accès ... (index)
- **vues (externe)**  
décrit les tables utilisées par les usagers + règles d'élaboration des vues à partir du schéma relationnel

### SGBD relationnels (portant l'étiquette relationnelle ...) :

SQL/DS IBM 80, DB2 IBM  
ORACLE 79, INGRES 80, SYBASE xx  
ADABAS 72, DATACOM/DB 74  
IDMS/R 72, RAPPORT 76  
INFORMIX ...

## Architecture fonctionnelle des SGBD relationnels



## Architecture de SQL/DS - IBM

### DSC (Data System Control)

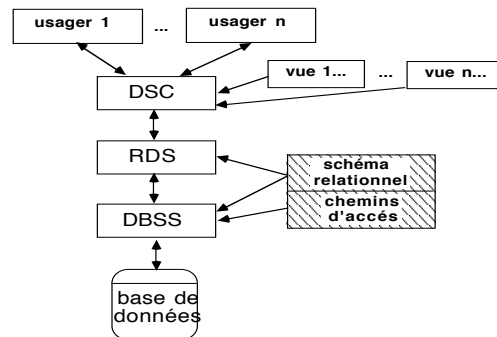
- communications du SGBD avec autres programmes (CICS, ...)
- superviseur initialisation, terminaison plusieurs usagers

### RDS (Relational Data System)

- traitement des requêtes usagers, analyse, optimisation, compilation

### DBSS (Data Base Storage System)

- effectue les accès demandés par RDS
- gère l'allocation d'espace
- accès concurrents
- reprises en cas de pannes



## Architecture de INGRES

### P1 moniteur de terminaux

- permet à l'utilisateur de formuler, imprimer, lancer l'exécution de commandes INGRES

### P2

- analyse syntaxique des requêtes
- modification des requêtes pour prise en compte des vues
- protection, contrôle de cohérence des données
- accès concurrents aux données

### P3

- décompose les requêtes portant sur plusieurs tables en suite de requêtes mono-tables
- traite les requêtes mono-tables

### P4

- utilitaires de : création de tables, d'index, destruction
- résistance aux pannes, retard des mises à jour

