



Conception de base de données

Introduction

L3 Informatique

Antoine Spicher

`antoine.spicher@u-pec.fr`

Bases de données

■ Définition

- Arrêté du 22/12/1981, enrichissement du vocabulaire informatique
Ensemble de données organisé en vue de son utilisation par des programmes correspondant à des applications distinctes et de manière à faciliter l'évolution indépendante des données et des programmes
- ISO 2682-4
Ensemble de données, qui est organisé selon une structure conceptuelle décrivant les caractéristiques de ces données ainsi que les relations entre leurs entités correspondantes et destiné à un ou plusieurs domaines d'application
- Code de la propriété intellectuelle, art. L.112-3
Parlement européen, art. 2 de la Directive n°96/9/CE
[...] recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique, et individuellement accessibles par des moyens électroniques ou par tout autre moyen

Bases de données



■ Définition

Collection structurée d'informations représentant les objets constituant une partie du monde réel et servant de support à une application

■ Types de BD

□ Moteur de recherche

Retrouver des ressources à partir de mots quelconques (e.g., Google)

Bases de données



■ Définition

Collection structurée d'informations représentant les objets constituant une partie du monde réel et servant de support à une application

■ Types de BD

- Moteur de recherche

- **Banque de données**

Collecter et publier un ensemble d'informations librement consultables et relatives à un domaine de connaissances

- Art : IMDb (*Internet Movie Database*), CDDB (*Compact Disc Data Base*)
- Droit : FNAEG (*Fichier national automatisé des empreintes génétiques*)
- ...

Bases de données



■ Définition

Collection structurée d'informations représentant les objets constituant une partie du monde réel et servant de support à une application

■ Types de BD

- ☐ Moteur de recherche
- ☐ Banque de données
- ☐ **Système d'information**

Collecter, classer et modifier les informations relatives à une activité

- ☐ Outils de gestion de stock
- ☐ Systèmes de réservation de billets (SNCF)
- ☐ CMS (*Content Management System*) comme Drupal, Joomla
- ☐ ...

Bases de données



■ Définition

Collection structurée d'informations représentant les objets constituant une partie du monde réel et servant de support à une application

■ Types de BD

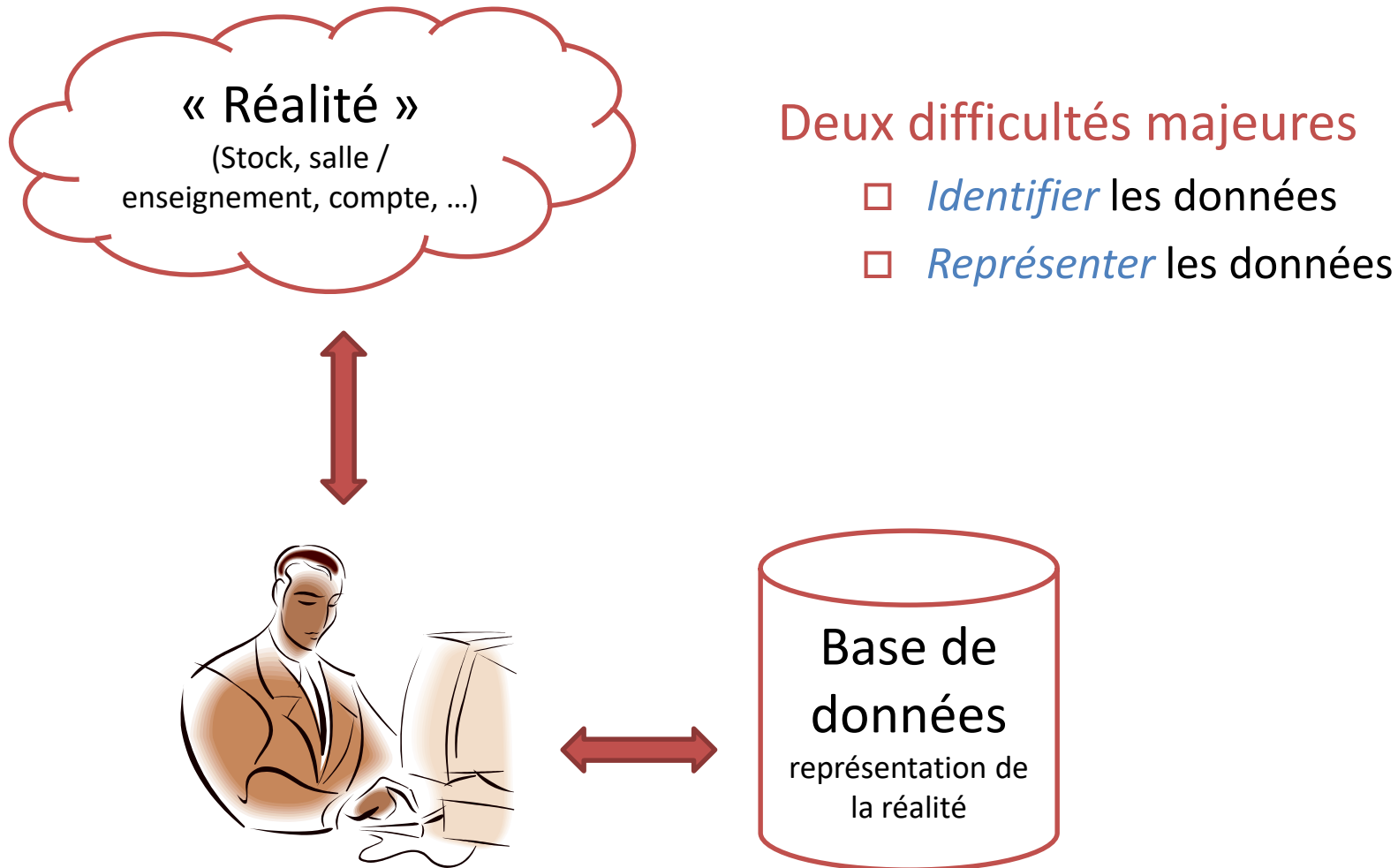
- ☐ Moteur de recherche
- ☐ Banque de données
- ☐ Système d'information
- ☐ **Entrepôt de données**

Collecter et stocker définitivement des informations historiques qui seront utilisées pour des statistiques et des analyses

- ☐ ...

Bases de données

■ *Big Picture...*



Identifier les données



- La philosophie du verbe *être*...
- Trois différentes compréhensions de l'*identité* (S. Ferret)
 - Identité *numérique*
 - *Relation qu'un objet entretient avec lui-même tout au long de son existence*
 - Être capable de distinguer chaque donnée présente dans la BD même si celle-ci évolue au cours du temps
 - Identité *spécifique*
 - *Appartenance de plusieurs objets à une même espèce ou sorte*
 - Être capable de regrouper les informations partageant les mêmes caractéristiques
 - Identité *qualitative* (ou *indiscernabilité*)
 - *Relation entre des objets qui ne sont distincts que par le nombre*
 - Être capable de représenter les informations suffisantes pour distinguer des objets réels différents

Représenter les données



■ Vocabulaire de la modélisation

□ Système

Objet ou phénomène du monde réel que l'on souhaite étudier

□ Modèle

Représentation abstraite (souvent dégradée) du système

□ Formalisme

Langage (formel, semi-formel ou informel) permettant la description du modèle

Représenter les données

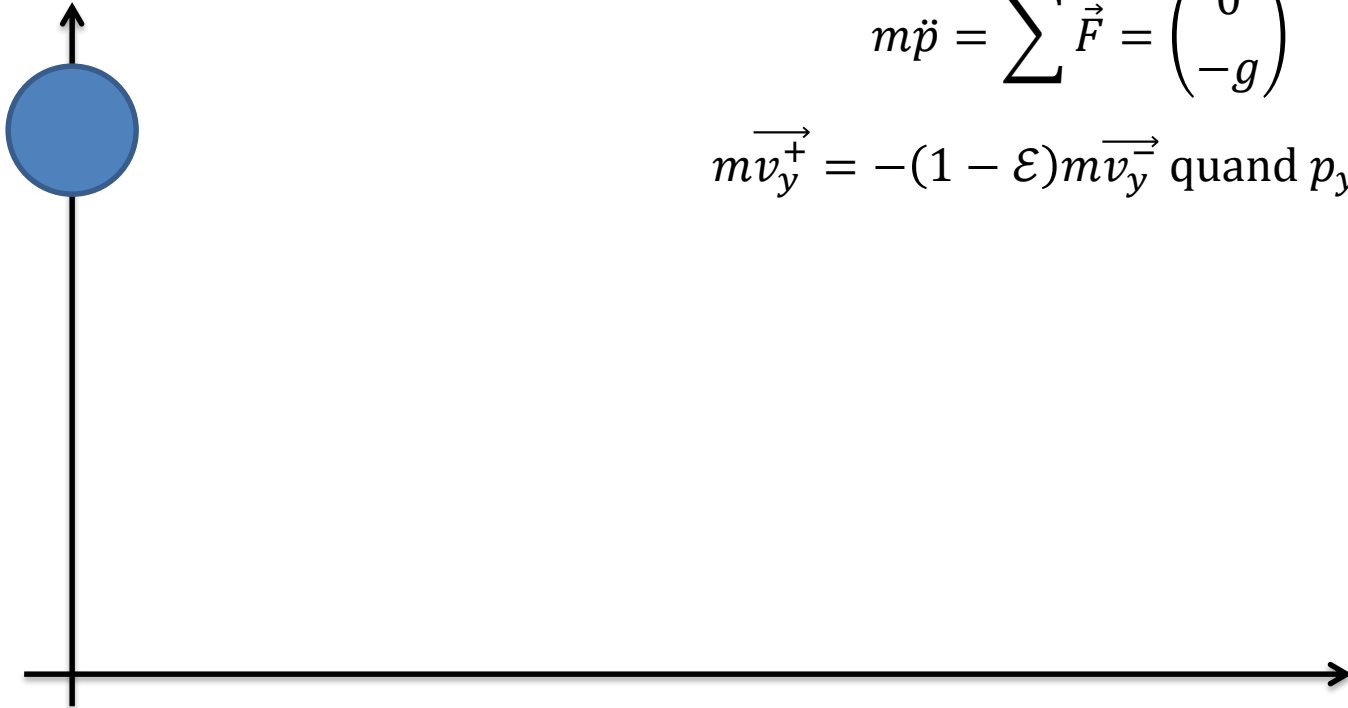
■ Exemple simple de modélisation

- **Système** : une balle rebondissante
- **Modèle** : mécanique du point, théorie de l'impact
- **Formalisme** : calcul différentiel

$$(p, \vec{v}): \mathbb{R}^+ \rightarrow \mathbb{R}^2 \times \mathbb{R}^2$$

$$m\ddot{p} = \sum \vec{F} = \begin{pmatrix} 0 \\ -g \end{pmatrix}$$

$$m\vec{v}_y^+ = -(1 - \varepsilon)m\vec{v}_y^- \text{ quand } p_y = 0$$



Représenter les données



■ Modélisation et base de données

□ Modélisation des *données*

■ Système

Informations réelles à stocker dans la BD

■ Modèle

Organisation/structuration des informations

■ Formalisme

Quel formalisme pour décrire le modèle ?

□ Modélisation de la *base de données*

■ Système

Stockage physique des données

■ Modèle

Abstraction réaliste du stockage des données

■ Formalisme(s)

Théorie des ensembles, représentations diagrammatiques, ...

Représenter les données

■ Modélisation et base de données

□ Modélisation des *données*

■ Système

Informations réelles à stocker dans la BD

■ Modèle

Organisation/structuration des informations

■ Formalisme

Tout modèle de données...

Schéma

□ Modélisation de la *base de données*

■ Système

Stockage physique des données

■ Modèle

Abstraction réaliste du stockage des données

■ Formalisme(s)

Théorie des ensembles, représentations diagrammatiques, ...

*Modèle
de données
database model*

Modèles de données



■ Histoire des modèles de données

- ❑ Modèle hiérarchique : fin des années 60 et années 70
- ❑ Modèle réseau : les années 70
- ❑ Modèle relationnel : de 1970 à 1980
- ❑ Modèle entité-association : les années 70
- ❑ Modèle entité-association étendu : les années 80
- ❑ Modèle orienté objet : fin des années 80 et début 90
- ❑ Modèle relationnel objet : fin des années 80 et début 90
- ❑ Modèle semi-structuré (*e.g.*, XML) : fin 90 à maintenant

Modèles de données

■ Modèle *hiérarchique*

- Organisation arborescente des données
 - Représentées par des **enregistrements**
 - Liées les unes aux autres par une relation arborescente
 - Association **dirigée** (du fils vers l'**unique** père) de multiplicité **1 <- ***



Modèles de données

■ Modèle *hiérarchique*

- Organisation arborescente des données
 - Représentées par des **enregistrements**
 - Liées les unes aux autres par une relation arborescente
 - Association **dirigée** (du fils vers l'**unique** père) de multiplicité **1 <- ***
 - Absence de **classe d'association**



- Difficultés
 - Forte **dépendance physique/logique** (associations = « pointeur »)
 - Représentation **restrictive** et **redondante**
 - Optimisation **ad hoc** des requêtes

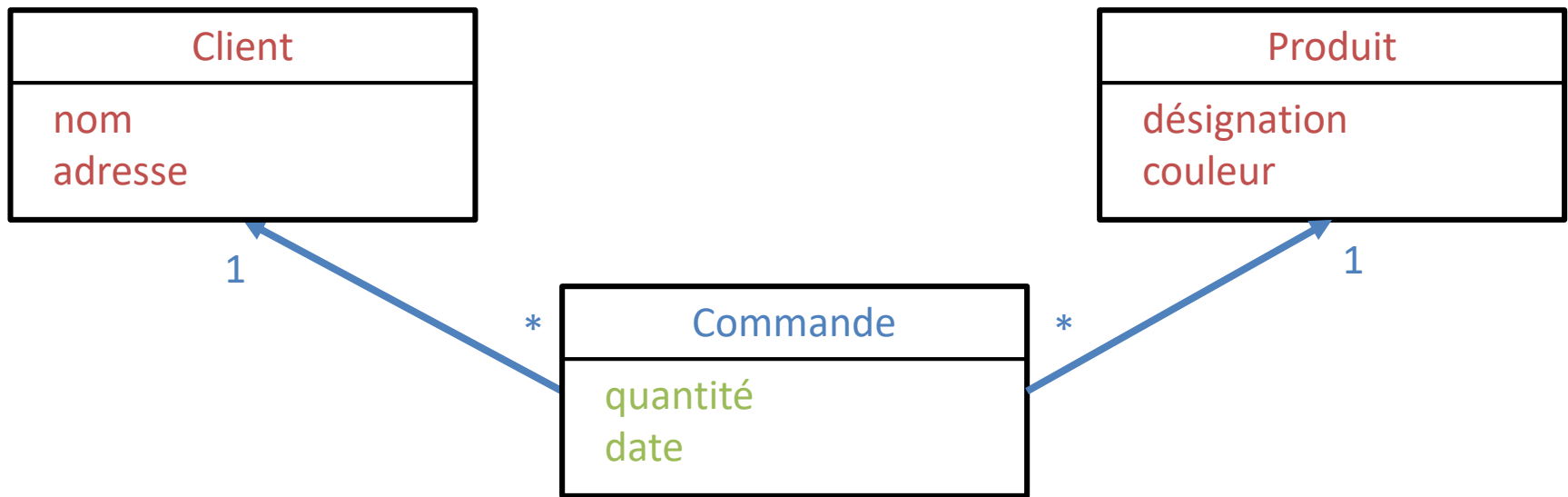
Modèles de données

■ Modèle *réseau*

□ Évolution du modèle hiérarchique

Le graphe des associations peut être quelconque

Toujours pas de classe d'association
mais on connaît une traduction



□ Gains et pertes

- Toujours une **forte dépendance physique/logique** (association = « pointeur »)
- Modèle plus **flexible** mais aussi plus **complexe** à implanter

Chargement et restauration des réseaux

Modèles de données

■ Modèle *relationnel*

- Nouveautés apportées par le modèle relationnel
 - Mathématisation des bases de données
 - Algèbre relationnelle fondée sur la théorie des ensembles
 - Caractérisation **formelle** de ce qu'est une « **bonne** » BD (normalisation)
 - **Indépendance physique/logique**

Associations représentées à travers la **notion de clé**

Client
<u>clientID</u>
nom
adresse

Commande
clientID
produitID
quantité
date

Produit
<u>produitID</u>
désignation
couleur

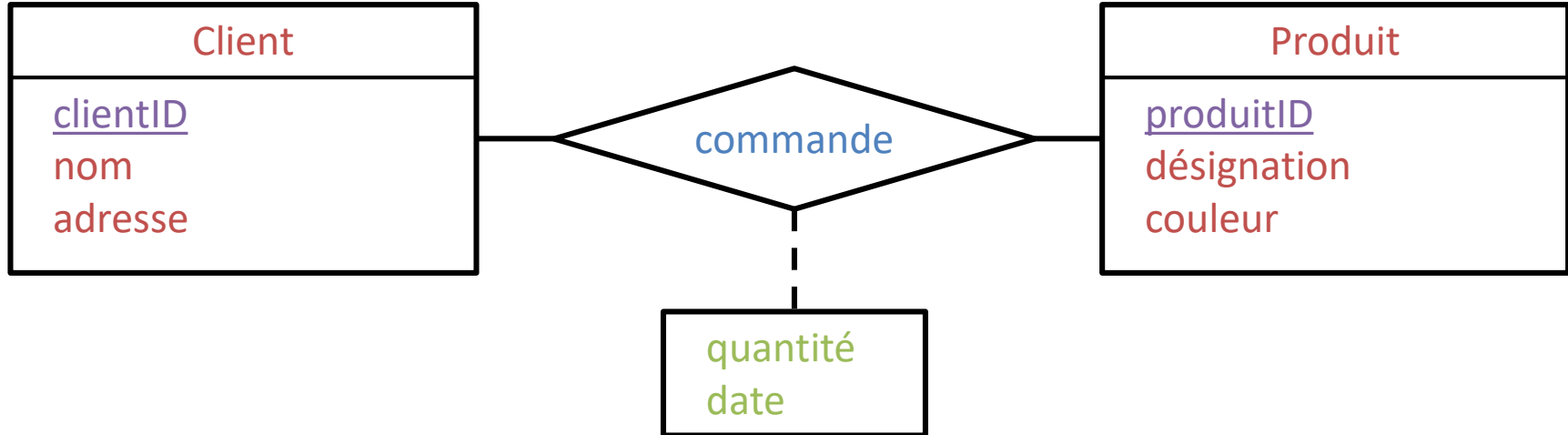
- Difficultés
 - Implantation efficace du modèle (vite résolu)
 - Représentation bas-niveau éloignée des soucis de conception (plus d'assoc.)

Modèles de données

■ Modèle *entité-association*

- Proposition d'amélioration du modèle relationnel

Réintroduction des associations



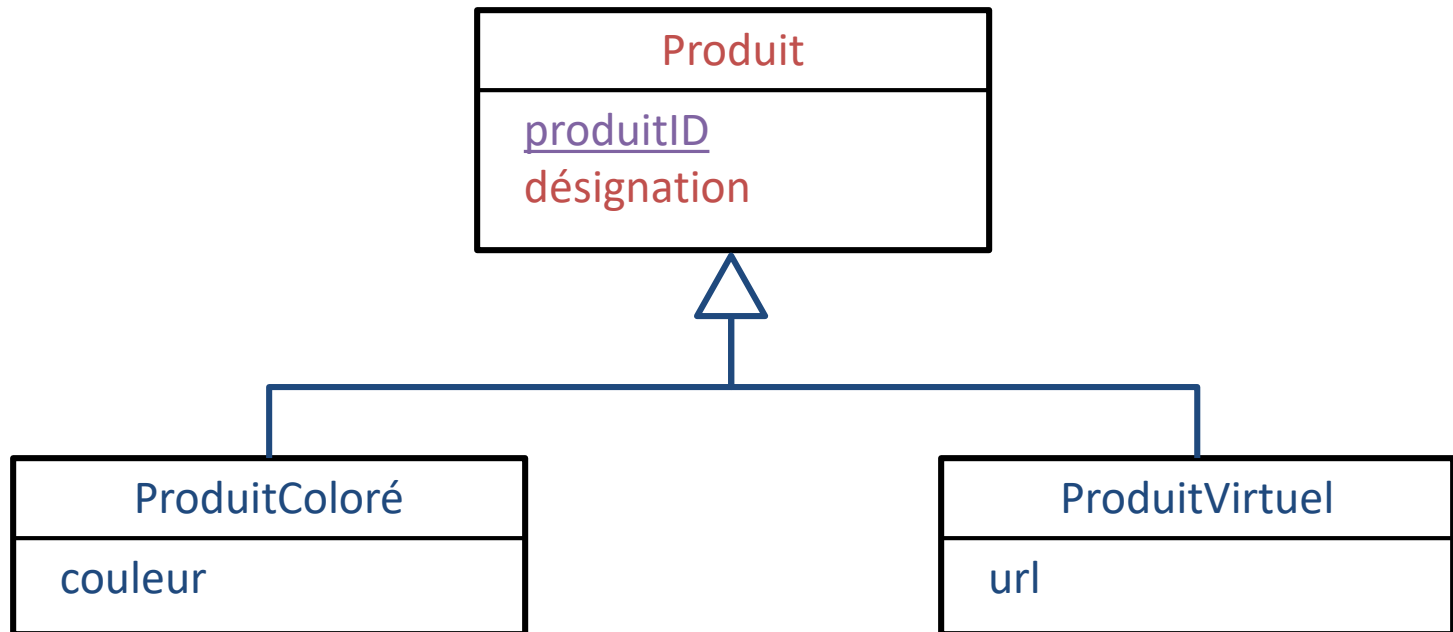
- Conséquences

- Outil de **conception** adapté et toujours utilisé...
- ... mais implantation directe trop difficile
 - Dépendances fonctionnelles (normalisation) difficiles à maîtriser
 - Traduction d'un schéma entité-association en schéma relationnel

Modèles de données

■ Modèle *entité-association étendu*

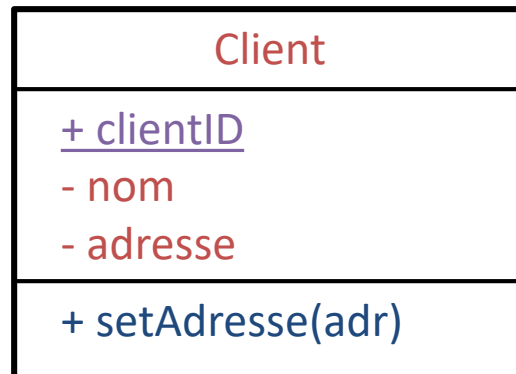
- Nouvelles propositions d'extension
 - Notion de domaine d'attributs
 - Agrégation (ordre supérieur)
 - Généralisation/spécialisation/héritage
 - ...



Modèles de données

■ Modèle *orienté objet* et modèle *relationnel-objet*

- Introduction des concepts « objets »
 - Possibilité de définition de nouveau type
 - Ajout de la notion de visibilité
 - Ajout de la notion de méthode/opération



Modèles de données

■ Modèle *orienté objet* et modèle *relationnel-objet*

- Introduction des concepts « objets »
 - Possibilité de définition de nouveau type
 - Ajout de la notion de visibilité
 - Ajout de la notion de méthode/opération
- Différences entre les deux approches
 - Orienté objet
 - Étendre un langage orienté objet pour la gestion d'objets persistants*
 - Relationnel-objet
 - Manipuler des objets persistants en utilisant les outils relationnels*
- Principales contributions
 - *User-defined data types*
 - *User-defined operators*
 - *User-defined functions*
 - *User-defined access methods*

Modèles de données



■ Modèle *semi-structuré*

- Aucune séparation entre données et schéma de données
 - Représentation d'informations ne pouvant être réduites à un schéma
 - Format d'échange souple et portable entre différentes (modèles de) BD
 - Modification simple et à la demande du schéma
- Exemple en XML (modèle orient document)

Modèles de données

■ Modèle *semi-structuré*

```
<clients>
  <client clientID=1>
    <nom>Chandler Bing</nom>
    <adresse>15 rue du Yemen, Yemen</adresse>
  </client>

  <client clientID=2> ... </client>
  ...
</clients>

<produits>
  <produit produitID=1>
    <designation>Billet d'avion pour le Yemen</designation>
    <commandes>
      <commande clientID=1 quantite=2 date="26/02/1998" />
    </commandes>
  </produit>

  <produit produitID=2> ... </produit>
  ...
</produits>
```

Modèles de données

■ Modèle *semi-structuré*

- Aucune séparation entre données et schéma de données
 - Représentation d'informations ne pouvant être réduites à un schéma
 - Format d'échange souple et portable entre différentes (modèles de) BD
 - Modification simple et à la demande du schéma
- Exemple en XML (modèle orient document)
- Rien n'est gratuit...
 - Requêtes « navigationelles » efficaces (de données en données)
 - Autres requêtes inefficaces (*e.g.*, croisement de données)
 - Vers les modèles NoSQL (*Not only SQL*)
 - Coût prohibitif des technologies relationnelles pour les BD du web
 - Prérequis ACID (cf. quelques slides) non réalisables en distribué
 - Déjà utilisés : Google (BigTable), Amazon (Dynamo), LinkedIn (Voldemort), Facebook (Cassandra, HBase), SourceForge (MongoDB), Ubuntu One (CouchDB), ...

Modèles de données

■ Histoire des modèles de données

- Modèle hiérarchique : fin des années 60 et années 70

- Modèle réseau : les années 70

- **Modèle relationnel : de 1970 à 1980**

modèles détaillés dans ce cours

- **Modèle entité-association : les années 70**

- **Modèle entité-association étendu : les années 80**

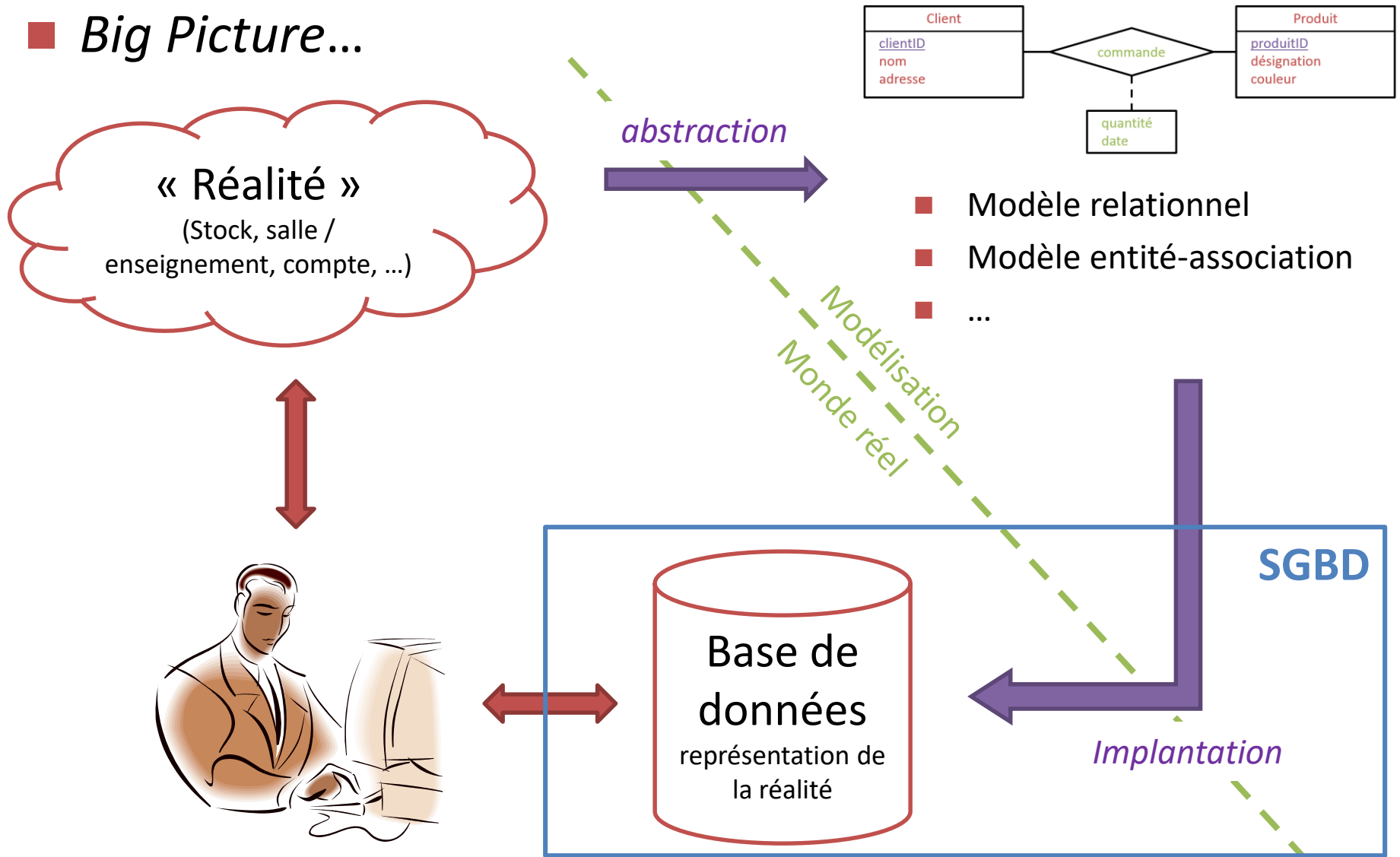
- Modèle orienté objet : fin des années 80 et début 90

- Modèle relationnel-objet : fin des années 80 et début 90

- Modèle semi-structuré (*e.g.*, XML) : fin 90 à maintenant

Bases de données

■ Big Picture...



Système de gestion de BD (SGBD)



■ Définition

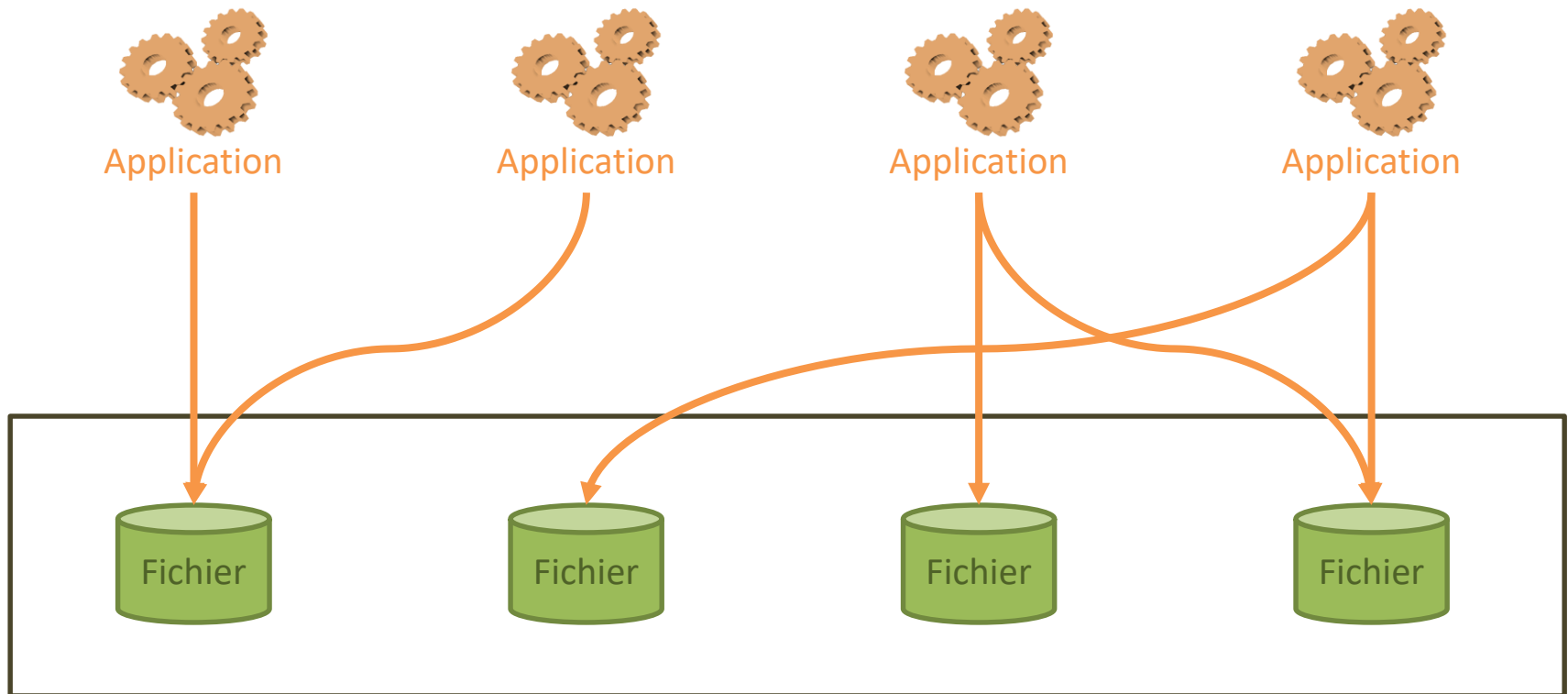
- *Ensemble coordonné de composants logiciel qui fournit un environnement pour décrire, modifier, interroger et administrer des ensembles de données d'une base de données tout en assurant la sécurité et la confidentialité*
- Attention, SGBD \neq BD
 - BD : ensemble des données stockées
 - SGBD : ensemble de fonctionnalités permettant de gérer la BD

■ Utilité des SGBD

- Limites de l'approche fichiers

Système de gestion de BD (SGBD)

■ Approche fichiers



Quelle cohérence entre les données ?

Système de gestion de BD (SGBD)



■ Définition

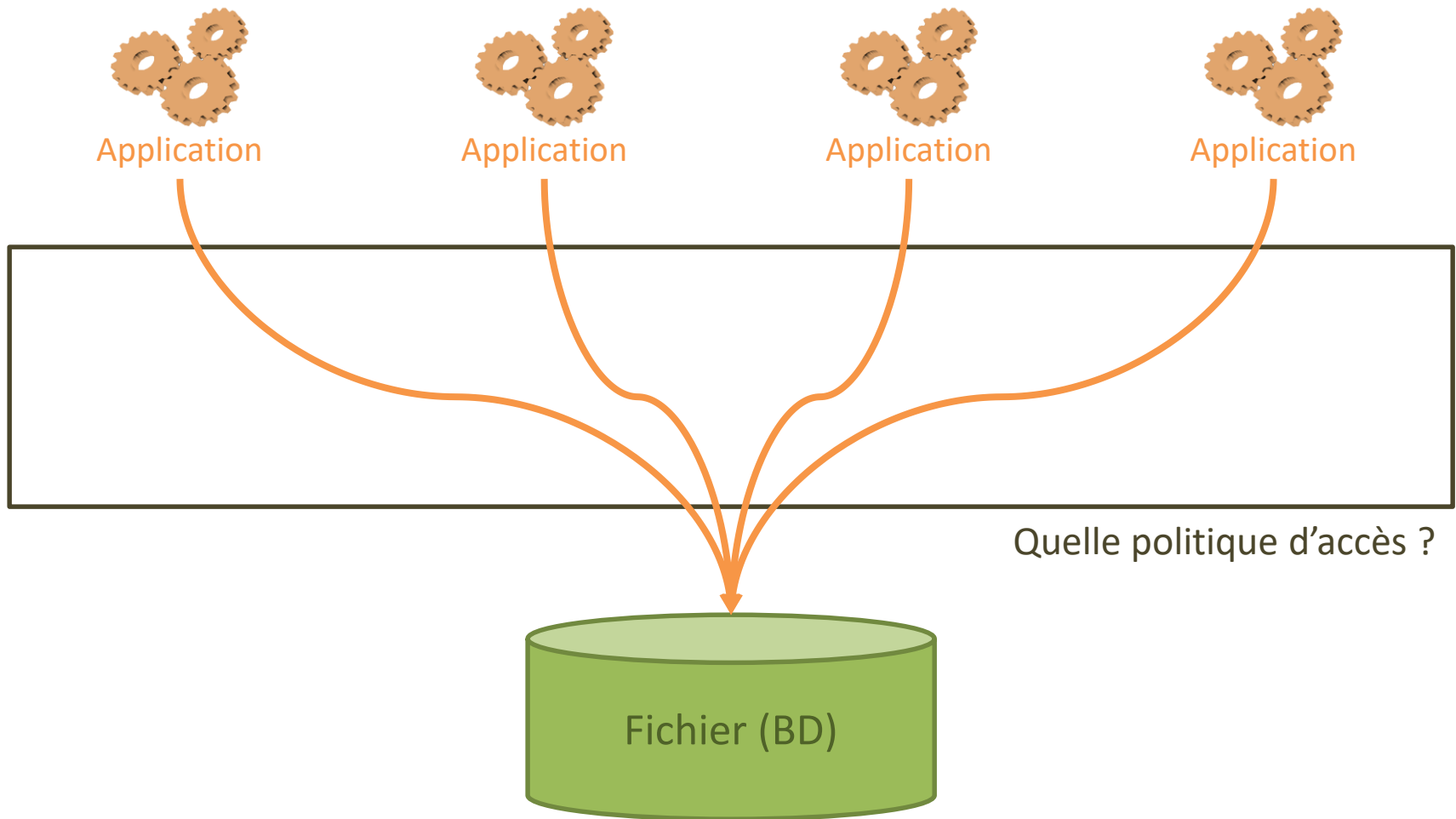
- *Ensemble coordonné de composants logiciel qui fournit un environnement pour décrire, modifier, interroger et administrer des ensembles de données d'une base de données tout en assurant la sécurité et la confidentialité*
- Attention, SGBD \neq BD
 - BD : ensemble des données stockées
 - SGBD : ensemble de fonctionnalités permettant de gérer la BD

■ Utilité des SGBD

- Limites de l'approche fichiers
 - Données associées sont mal définies, mal désignées (polysémie, synonymie), redondantes, peu accessibles ponctuellement, peu fiables, ...

Système de gestion de BD (SGBD)

■ Approche intermédiaire



Système de gestion de BD (SGBD)



■ Définition

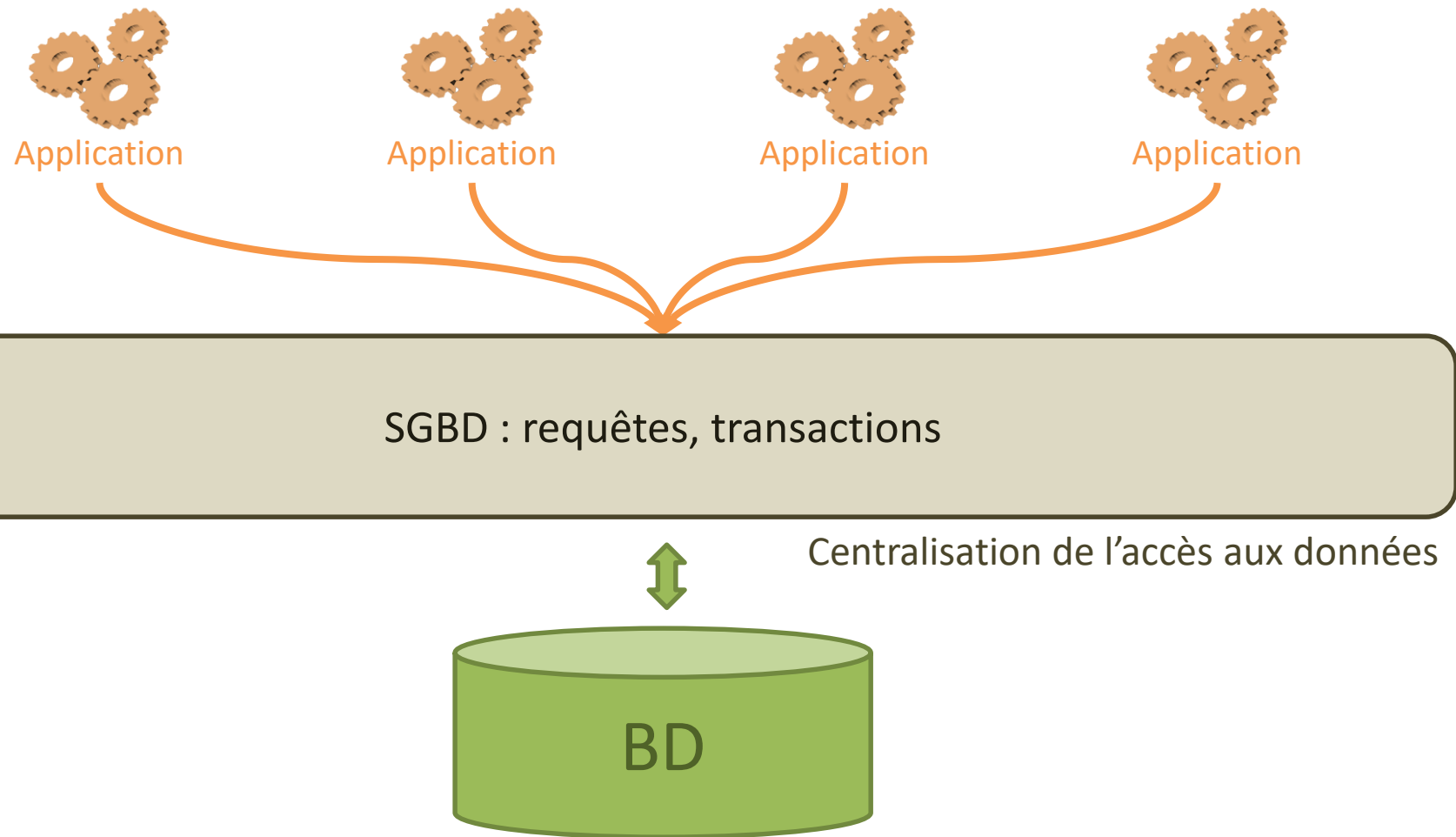
- *Ensemble coordonné de composants logiciel qui fournit un environnement pour décrire, modifier, interroger et administrer des ensembles de données d'une base de données tout en assurant la sécurité et la confidentialité*
- Attention, SGBD \neq BD
 - BD : ensemble des données stockées
 - SGBD : ensemble de fonctionnalités permettant de gérer la BD

■ Utilité des SGBD

- Limites de l'approche fichiers
 - Données associées sont mal définies, mal désignées (polysémie, synonymie), redondantes, peu accessibles ponctuellement, peu fiables, ...
 - Dépendance forte entre programmes et données
 - Modification des fichiers \Rightarrow modification des programmes (et vice-versa)
 - Difficultés pour le développement de nouvelles applications

Système de gestion de BD (SGBD)

■ Approche SGBD



Qualités logiciel des SGBD



- Indépendance des programmes aux données
 - Dépendance physique
 - Dépendance logique

Qualités logiciel des SGBD



- Indépendance des programmes aux données
- Simplicité des manipulations des données
 - Langage non-procédural (quoi et non comment, cf. prog. fonctionnelle)
 - Recherche, insertion et mise-à-jour

Qualités logiciel des SGBD



- Indépendance des programmes aux données
- Simplicité des manipulations des données
- Efficacité des accès aux données
 - Temps de réponse
 - Débit global

Qualités logiciel des SGBD



- Indépendance des programmes aux données
- Simplicité des manipulations des données
- Efficacité des accès aux données
- Partage et sécurité des données
 - Confidentialité (authentification, droits d'accès, ...)
 - Gestion de la concurrence des accès
 - Simultanéité lectures/écritures
 - Restauration après une panne (journalisation, sauvegarde)

Qualités logiciel des SGBD



- Indépendance des programmes aux données
- Simplicité des manipulations des données
- Efficacité des accès aux données
- Partage et sécurité des données
- Redondance contrôlée des données
 - Ne pas conserver deux fois les mêmes données
 - Sauf pour conserver une copie (sauvegarde)

Qualités logiciel des SGBD



- Indépendance des programmes aux données
- Simplicité des manipulations des données
- Efficacité des accès aux données
- Partage et sécurité des données
- Redondance contrôlée des données
- Conception facilitée des applications
 - Conception visuelle des BD (diagrammes E/R, E/A, objets)
 - Conception des traitements (diagrammes de flux entre modules)
 - Dictionnaire de données (objets BD, graphiques, applicatifs)

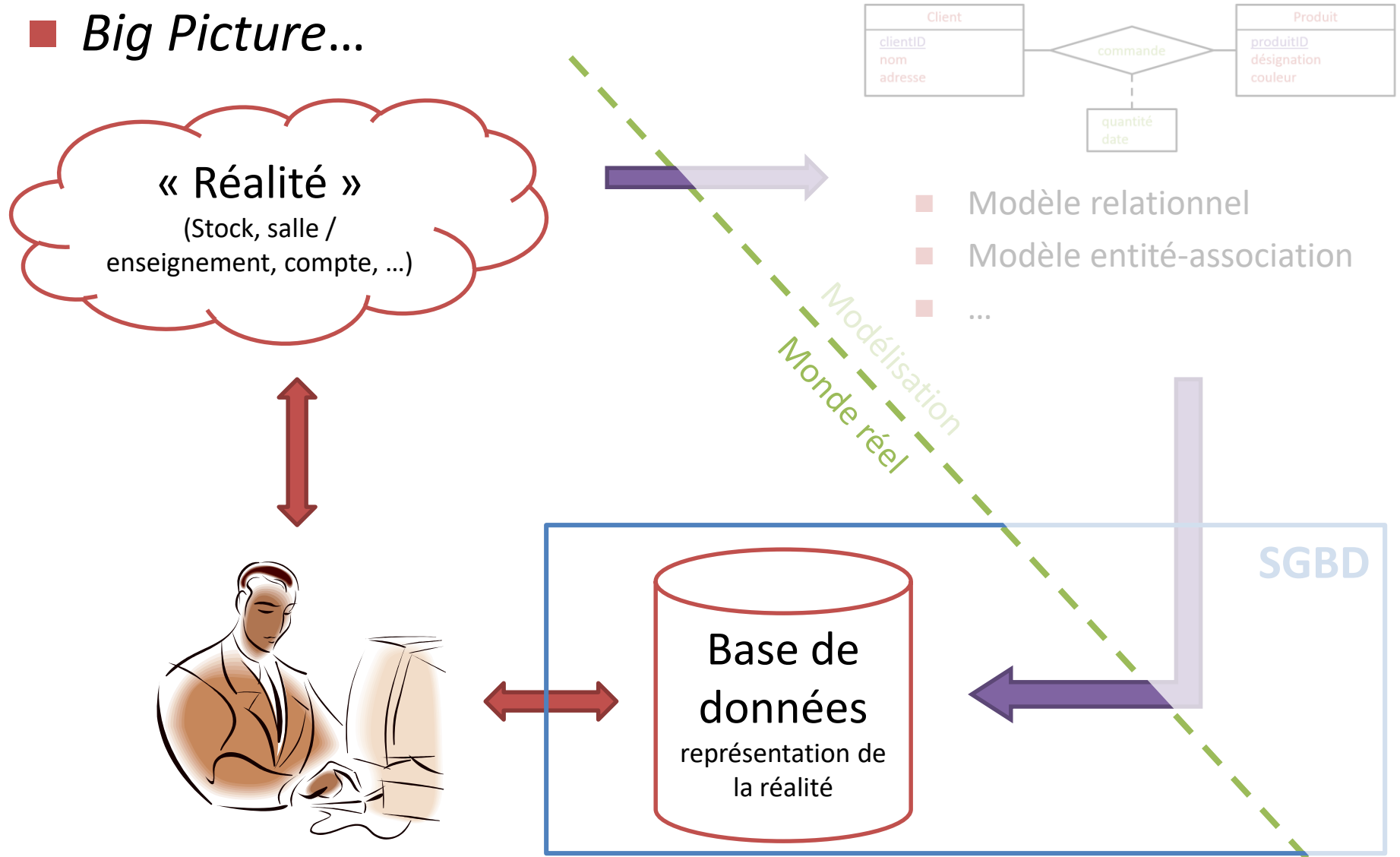
Qualités logiciel des SGBD



- Indépendance des programmes aux données
- Simplicité des manipulations des données
- Efficacité des accès aux données
- Partage et sécurité des données
- Redondance contrôlée des données
- Conception facilitée des applications
- Facilité de l'administration système des BD
 - Outils d'audit & de *tunning*
 - Visualisation des plans d'accès
 - Élaboration de statistiques

SGBD et monde réel

■ Big Picture...



SGBD et monde réel

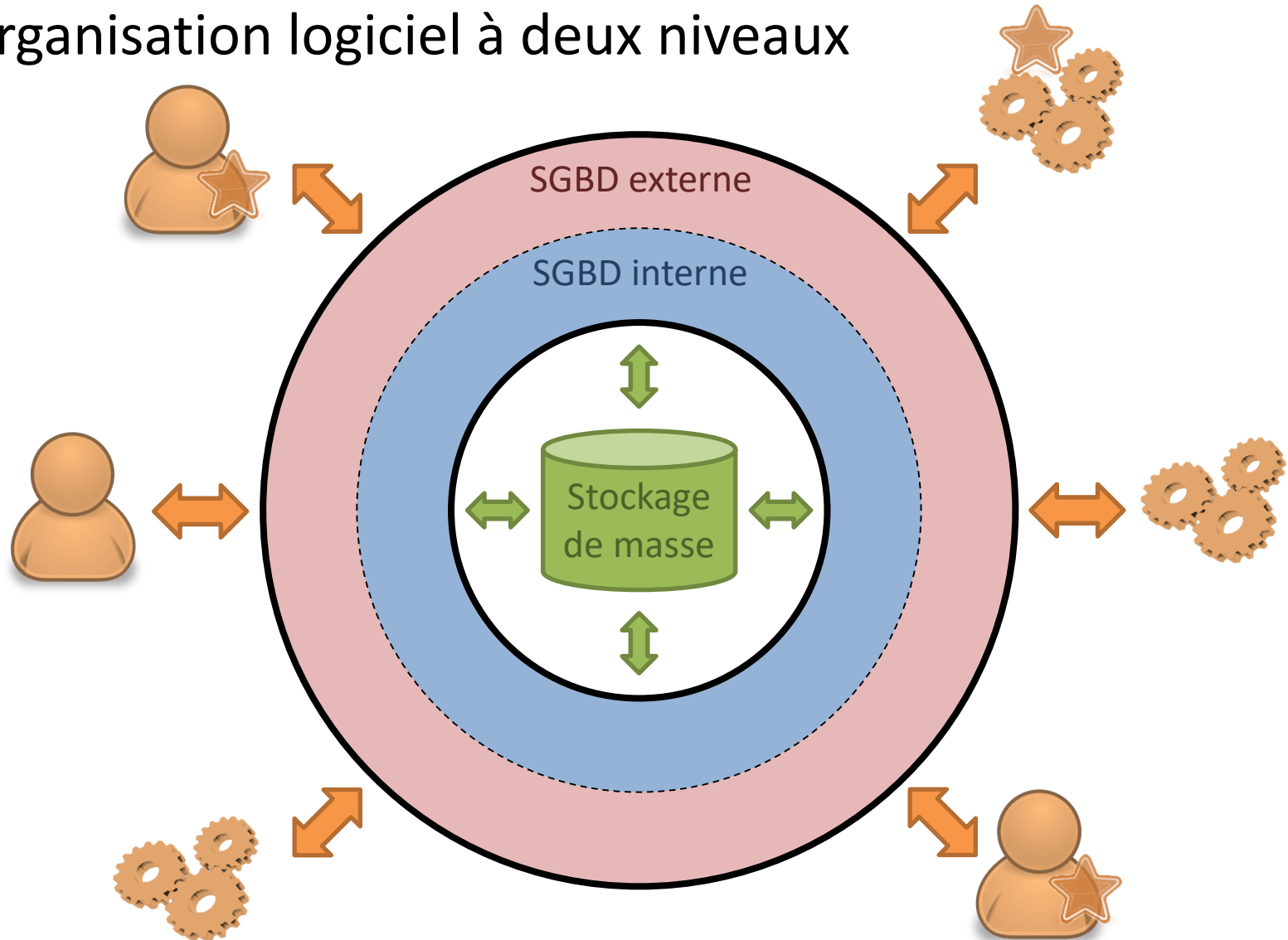


■ Organisation logiciel à deux niveaux

- Niveau interne : gestion de la mémoire physique
 - Interface avec le(s) unité(s) de stockage (hardware)
 - Types d'enregistrements internes
 - Représentation des champs mémoires
 - Séquencement physique
- Niveau externe : gestion de la logique utilisateur
 - Interface vers les utilisateurs
 - Différents outils de communication
 - Langages/bibliothèques de requêtes
 - Différents types d'utilisation
 - Interactive : requêtes saisies par l'utilisateur
 - Intégrée : requêtes saisies par une application étrangère
- Indépendance entre les niveaux physiques et logiques

SGBD et monde réel

■ Organisation logiciel à deux niveaux



SGBD et monde réel

■ Deux types d'utilisations

□ Modificateur de données

■ *Langage de modification de données* (LMD)

■ Utilisation conventionnelle de la BD

□ *Interrogation* de la BD

□ Mise-à-jour de la BD



□ Administrateur★

■ *Langage de définition de données* (LDD)

■ Responsable de la structure de la BD

□ Définition et gestion conceptuelles de la BD

□ *Contrôle* des droits d'accès (vues), sécurité et fiabilité



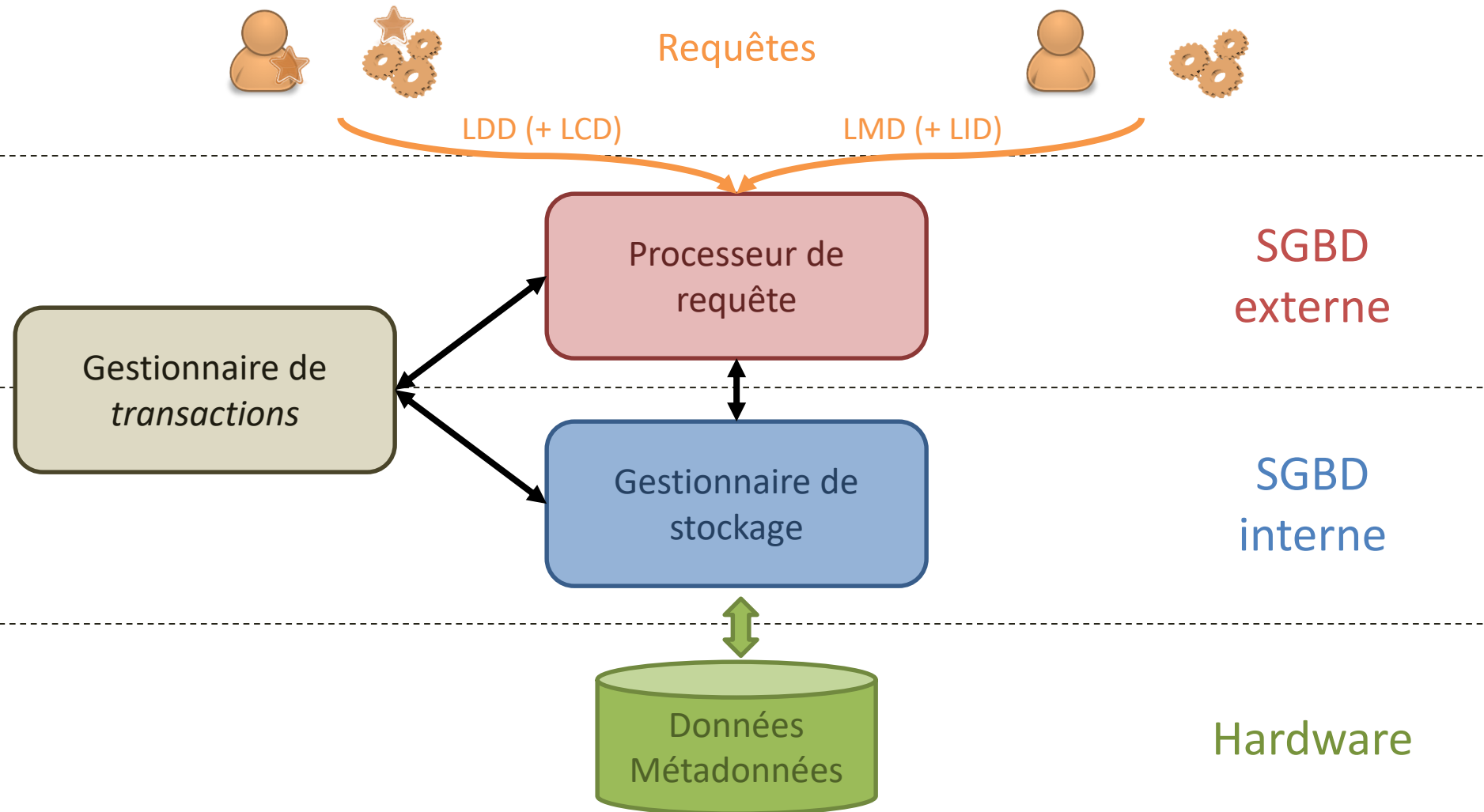
□ Structuration alternative des langages de requêtes

■ *Langage d'interrogation de données* (LID) : usuellement inclus dans LMD

■ *Langage de contrôle de données* (LCD) : usuellement inclus dans LDD

SGBD et monde réel

■ Traitement d'une requête par le SGBD



SGBD et monde réel

■ Transactions

□ Définition

- *Séquence considérée atomique d'opérations (read/write) sur une BD*
- Toutes transactions doit respecter la cohérence de la base
 - Contraintes d'intégrité (propriétés logiques à ne pas violer)
 - Aucune information sémantique dans le SGBD

□ Propriétés ACID

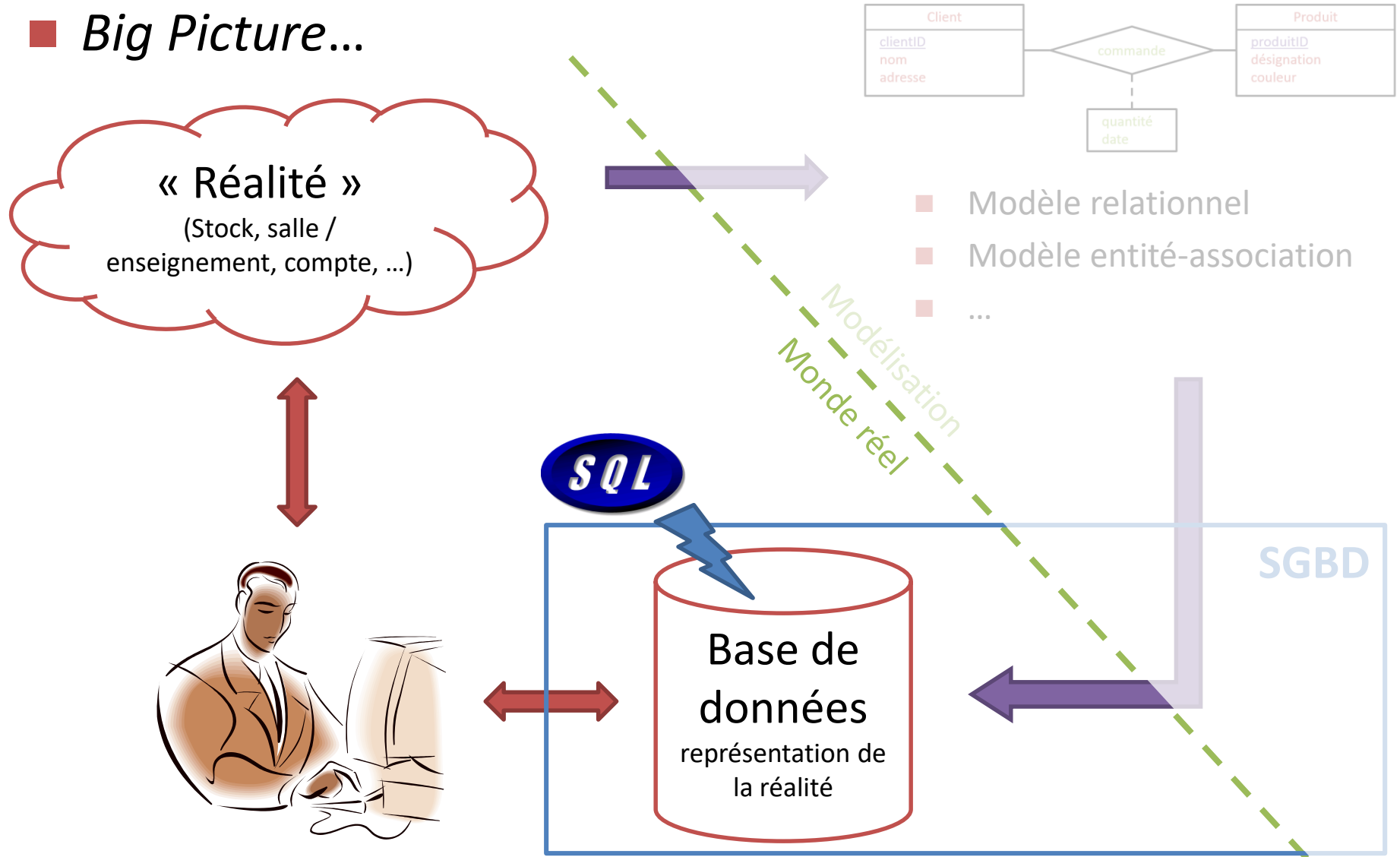
- **Atomicité** : pas d'entrelacement possible avec d'autres transactions
- **Cohérence** : respect des contraintes référentielles et d'intégrité
- **Isolation** : exécution indépendante

Les modifications ne sont visibles qu'une fois la transaction terminée
- **Durabilité** : préservation des modifications effectuées

Succès \Rightarrow mäj de la BD, échec \Rightarrow état de la BD inchangé

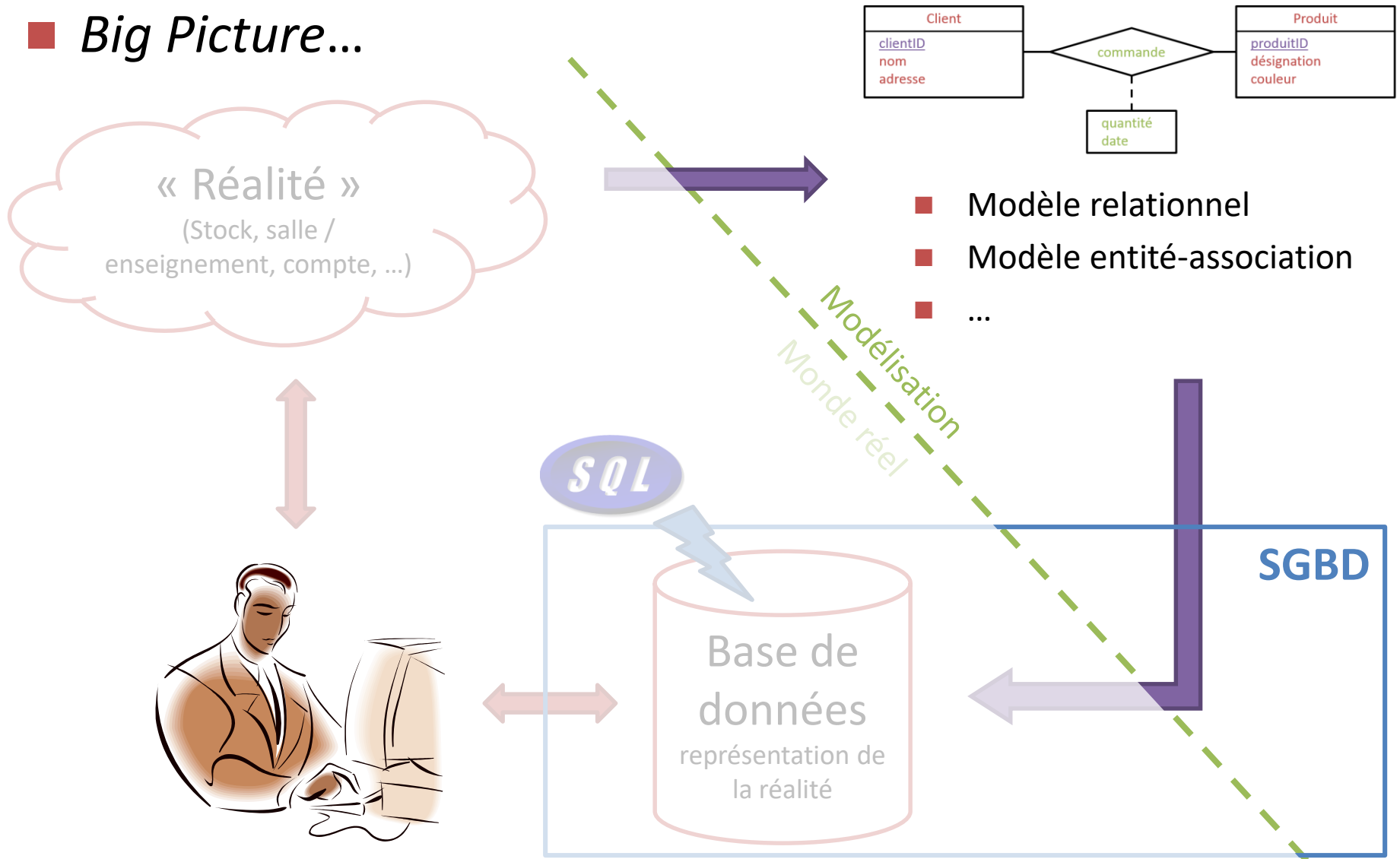
SGBD et monde réel

■ Big Picture...



SGBD et modélisation

■ *Big Picture...*



SGBD et modélisation

■ Organisation des données à trois niveaux

□ Niveau *interne*

- Point de vue du **SGBD** (donc de son concepteur)
- *Modèle et schéma physiques* de données, **agnostique** sur la BD
- Structure de données, moteur de la BD, ...

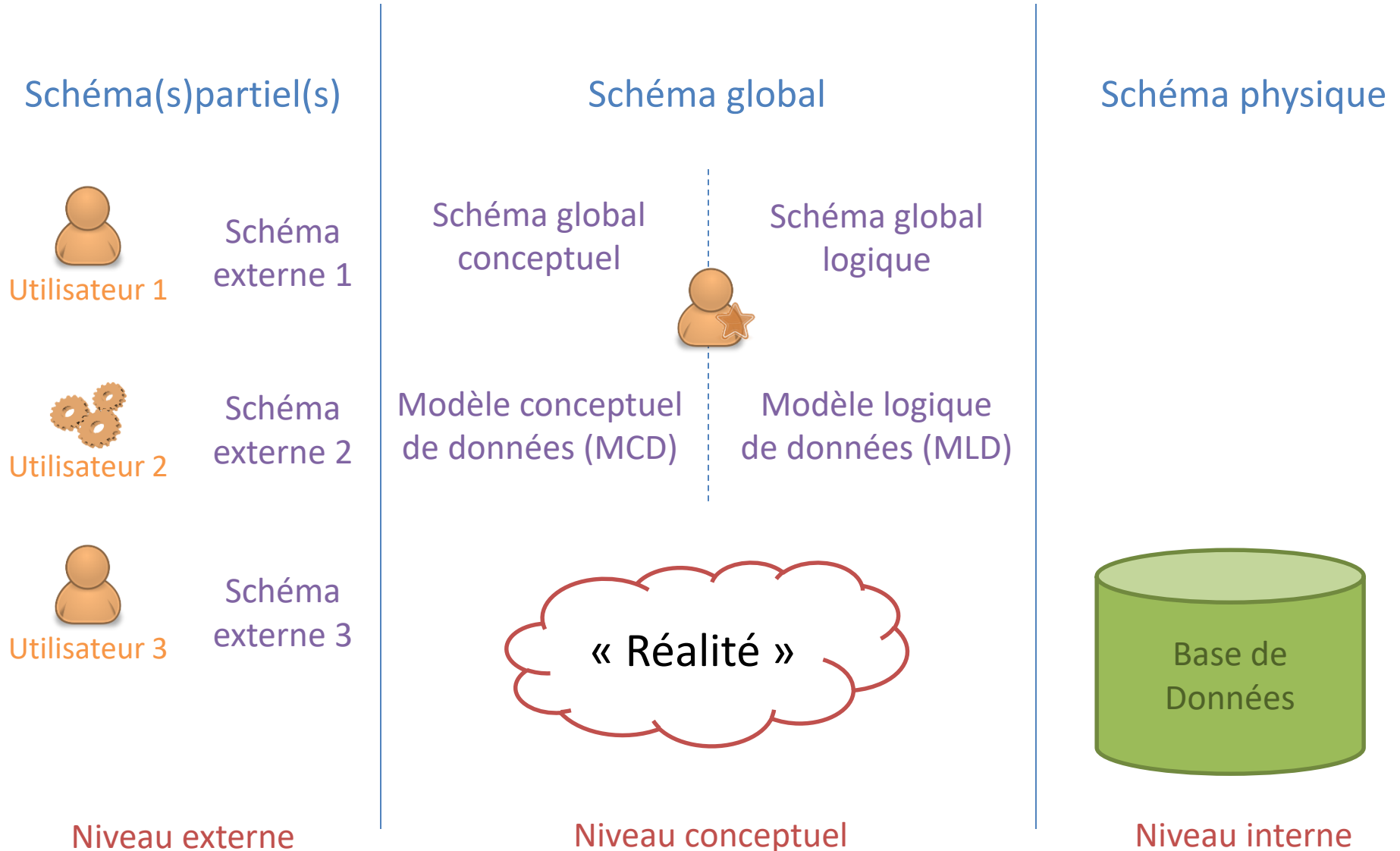
□ Niveau *conceptuel/logique* ★

- Point de vue de l'**administrateur**
- Structure et sémantique des données, **connaissance globale**
- *Modèle et schéma globaux* de la BD (différents niveaux de description)
 - **Modèle conceptuel de données** (MCD)
 - **Modèle logique de données** (MLD)

□ Niveau *externe*

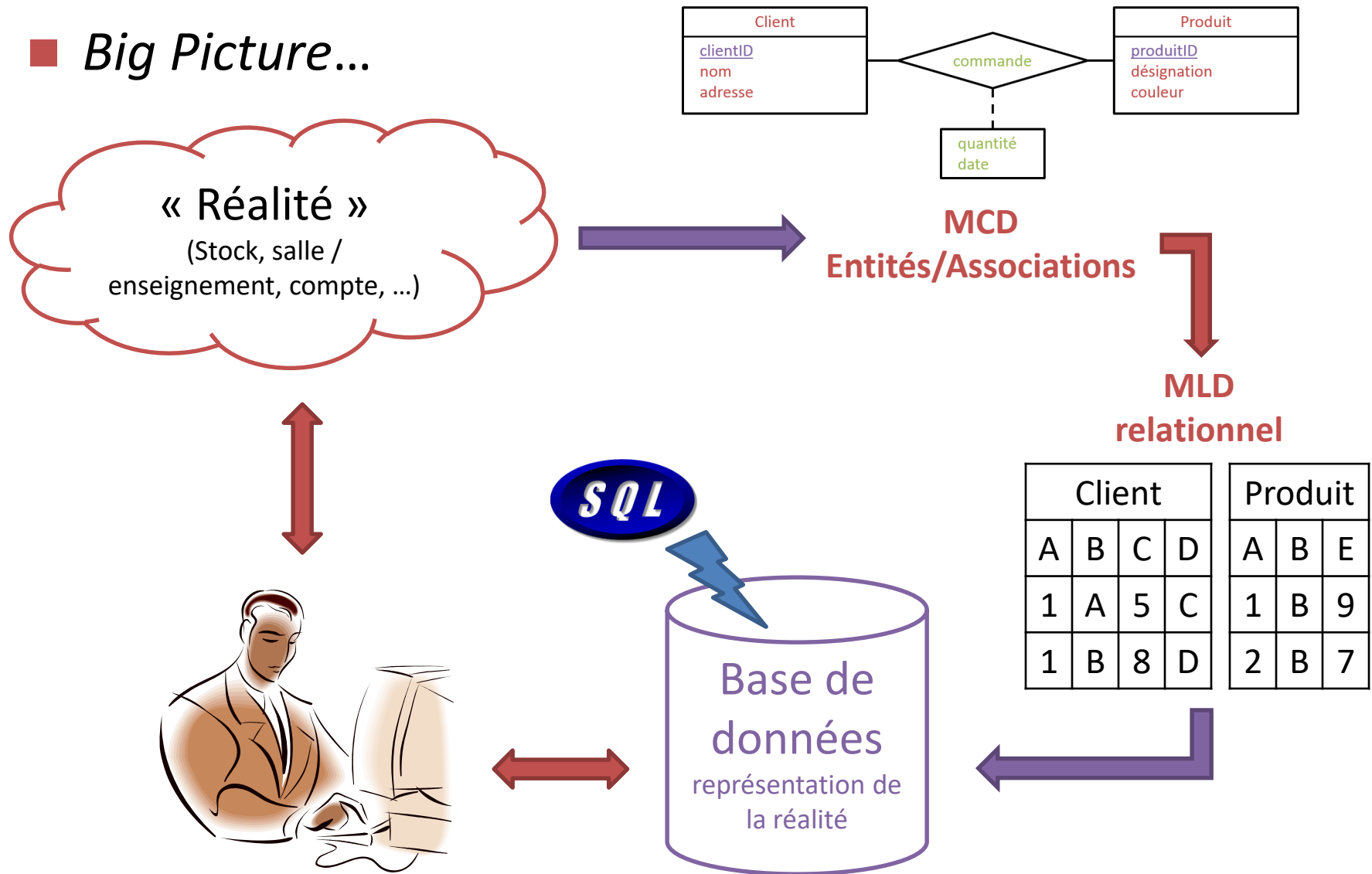
- Point de vue des **utilisateurs**
- Prise en compte des droits d'accès, **connaissance partielle**
- *Modèles et schémas externes partiels*

SGBD et modélisation



SGBD et modélisation

■ Big Picture...



Glossaire



- BD : base de données
- SGBD : système de gestion de bases de données
- Modélisation
 - Modèle de données : modélisation d'une base de données
 - Schéma de données : modélisation des données
 - MLD : modèle logique de données
 - MCD : modèle conceptuel de données
 - Modèle/schéma relationnel, modèle/schéma entité-association
- Langage(s) de requêtes
 - LDD : langage de définition de données
 - LMD : langage de manipulation de données
 - LCD : langage de contrôle de données
 - LID : langage d'identification de données

Plan du cours



■ Introduction aux BD

■ BD au niveau logique

- Modèle et algèbre relationnels
- Langage d'interrogation de données SQL

■ BD au niveau conceptuel

- Modèle Entité-Association
- Traduction vers le niveau logique
- Langage de définition et de modification de données SQL

■ Développement d'un programme d'application

Mise en place d'un site web avec la technologie PHP/MySQL (ou PostgreSQL)



-- FIN --