

# Bases de données avancées

cours 1 : introduction

# Plan

- Objectifs du cours
- Bases de données : utilisation
- Système de Gestion de Bases de données (SGBD)
- Architecture d'un SGBD
- Ingénierie des bases de données

# Objectifs du cours

- savoir représenter et exploiter des données dans plusieurs systèmes
- comprendre les abstractions et les mathématiques qui leur sont liées
- comprendre la conception d'un SGBD, i.e. un système tel qu'un ingénieur peut être amené à en concevoir pendant sa carrière

# Bases de données : utilisation et SGBD

# L'information

- De grandes quantités d'information :
  - entreprises avec des millions de clients
  - services publics pour l'ensemble de la population d'un pays
  - l'ensemble de la production quotidienne de la presse
  - etc...

# L'information

- De grandes quantités d'information:
  - à conserver sur de longues périodes
  - il faut pouvoir retrouver l'information pertinente rapidement et simplement
  - il faut pouvoir la mettre à jour
  - il faut pouvoir la présenter

# Quelques exemples

- Google, Amazon.com, Yahoo!
- les cotations de la bourse, les pièces détachées dans le réseau Peugeot, les assurés de la sécurité sociale, les réservations d'avion
- les banques de génomes, le catalogue du patrimoine de l'état français

Systeme de Gestion de  
Base de Données :  
ce que l'on en attend

# Exemple : système de réservation de la SNCF.

Billet

nom client  
numéro train  
date  
classe  
numéro wagon  
numéro place  
gare départ  
date départ  
gare arrivée

Train

numéro train  
gares desservies  
dates départ

Arrêt

numéro train  
nom gare  
dates arrivée  
dates départ

# Définition des données

- Décrire les données à stocker sans référence à un SGBD particulier :
  - modélisation conceptuelle
- Elaborer/compiler une description équivalente pour le stockage des données dans le SGBD choisi :
  - modélisation logique
  - langage de description des données

# Création et modification des données

- Créer une représentation initiale du réseau SNCF :
  - langage permettant l'insertion de données
- Créer au fur et à mesure les données sur les réservations. Pouvoir modifier, voire annuler une réservation :
  - langage de manipulation de données (insertion, modification, insertion)

# Interrogation des données.

- Répondre à toute demande d'information portant sur les données contenues dans la base :
  - Durand Julien a-t-il une réservation pour aujourd'hui ?
  - Quels sont les trains de Bordeaux à Paris entre 9h00 et 10h00 le dimanche ?
  - Donner les destinations au départ de Bordeaux sans arrêt intermédiaire.
- langage de requête (langage d'interrogation)

# Intégrité et cohérence des données

- Il faut pouvoir exprimer et maintenir après insertion/modification les contraintes qui lient les valeurs enregistrées dans la base :
  - au plus un passager par siège
  - les dates de réservation doivent coïncider avec les dates de circulation des trains
  - etc...
- langage d'expression des contraintes d'intégrité

# Fiabilité/Concurrence

- Les informations ne doivent pas être perdues que ce soit en cas de panne système, erreur de programmation, panne d'ordinateur...
  - **garantie de fiabilité**
- Il ne faut pas qu'une réservation soit annulée parce que quelqu'un d'autre faisait une réservation au même moment.
  - **contrôle de la concurrence**

# Confidentialité

- Toute information doit pouvoir être protégée contre l'accès d'un utilisateur non autorisé :
  - en lecture ou en écriture
- Exemple : interdire au clients de modifier les numéros de trains ou les horaires de circulation.

# Performance

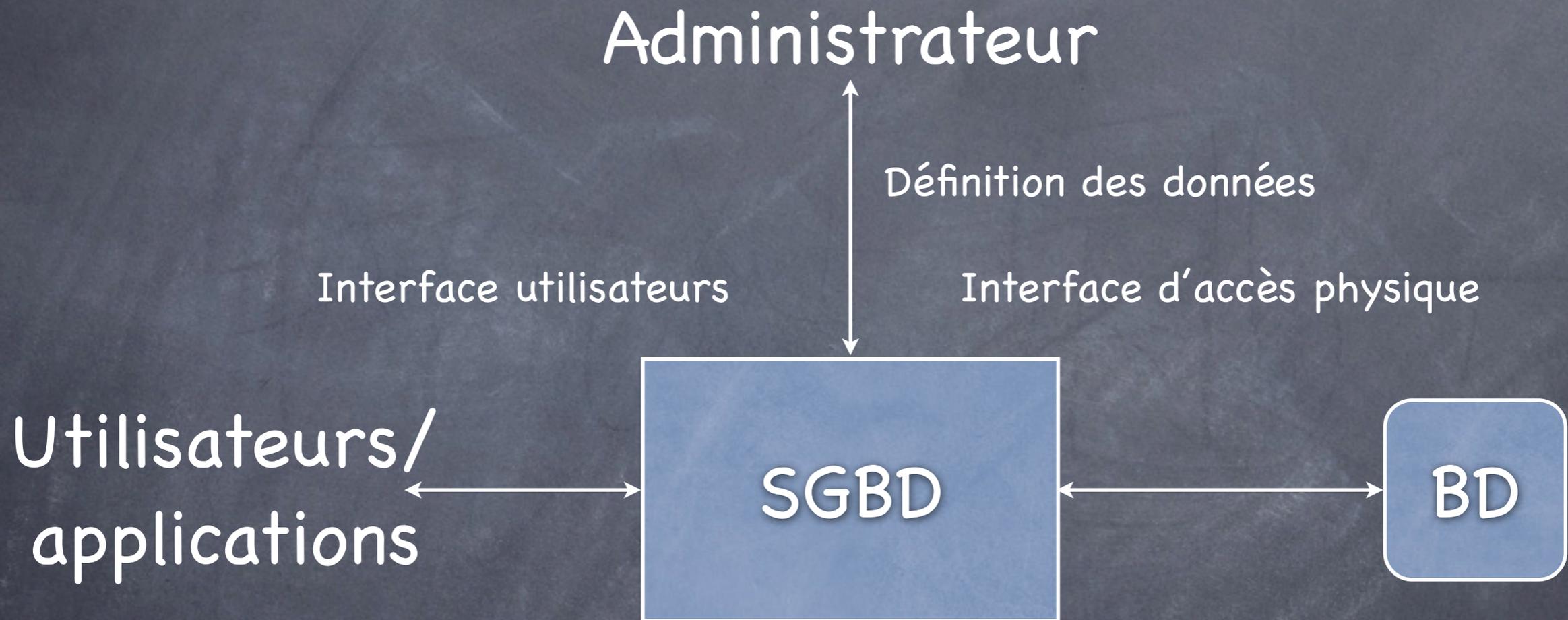
- Stockage de grandes quantités de données
- Durée du stockage de l'ordre de plusieurs années
- Rapidité du système :
  - mécanismes d'optimisation des requêtes
  - redondance des données
  - répartition des données sur plusieurs sites

# Architecture d'un SGBD

# Un peu de vocabulaire

- Base de données : ensemble cohérent, intégré partagé de données structurées, défini pour les besoins d'une application
- Système de Gestion de Bases de Données :
  - logiciel répondant aux contraintes que l'on vient de voir.

# Architecture



Analyse/vérification des requêtes  
Convivialité de l'interface  
Langages de haut niveau

Stockage/accès aux données  
Optimisation des performances

# Trois couches

- **Couche externe** : dialogue avec les utilisateurs et vues associées à chaque groupe d'utilisateurs
- **Couche logique** : stockage des données sur des supports physiques, gestion des structures de mémorisation (fichiers) et d'accès (gestion des index, des clés, ...)
- **Couche interne** : contrôle global et structure globale des données



# Traitement d'une requête

- Analyse syntaxique et sémantique de la requête
- Traduction au niveau logique
- Contrôle confidentialité/concurrence
- Optimisation et découpage en sous-requêtes internes
- Traduction des sous-requêtes action physique sur les disques et la mémoire



# Ingénierie des bases de données

# Cycle de vie en quatre phases

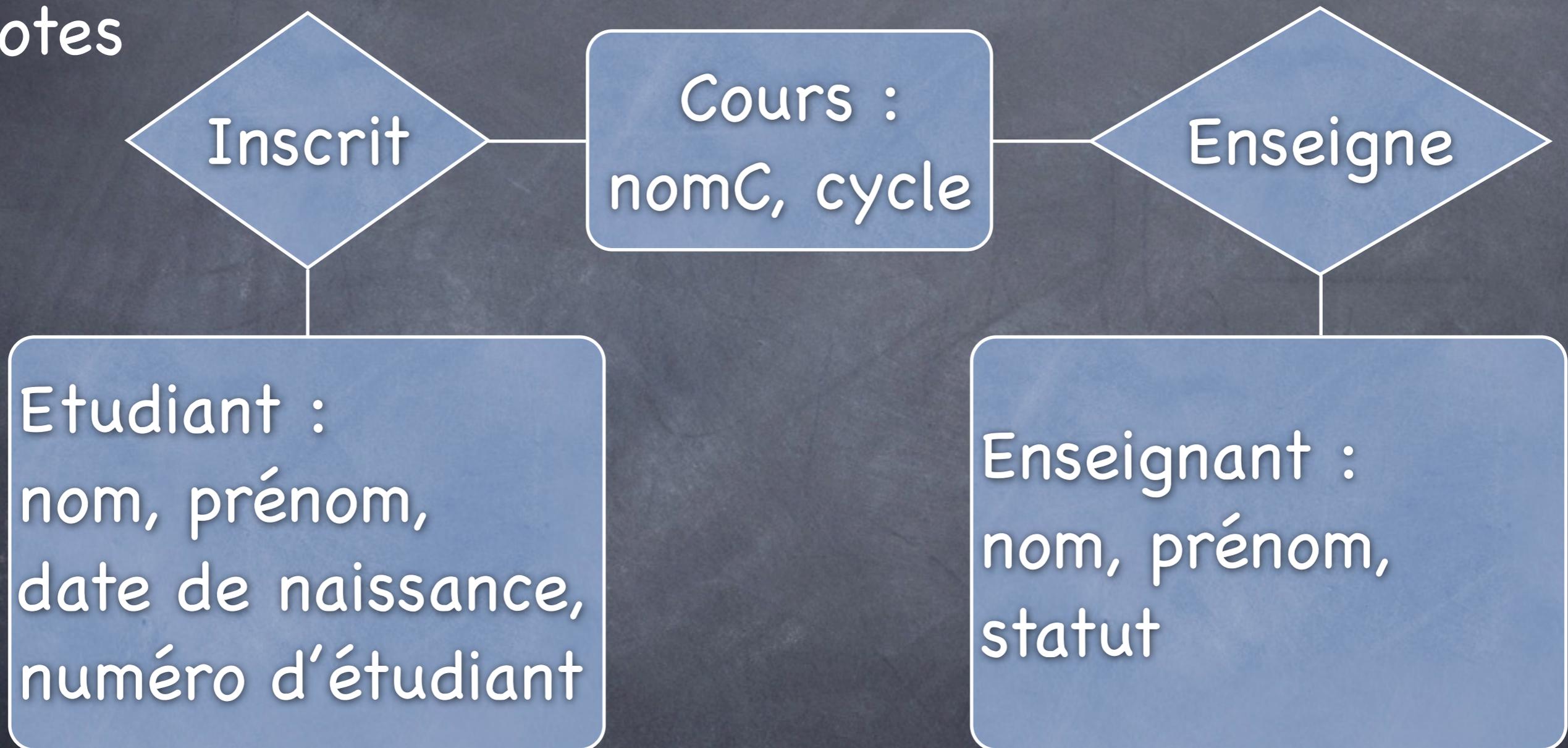
- Conception de la base de données (schéma conceptuel)
- Implantation des données (schéma logique)
- Utilisation (interrogation et mise à jour)
- Maintenance (correction, évolution)

# Le niveau conceptuel

- Description des besoins : modèle conceptuel
  - support du dialogue concepteur/utilisateur
  - indépendant de la solution informatique
  - deux parties couvertes par le modèle
    - statique (la structure des données)
    - dynamique (règles et opérations)
- Contraintes d'intégrité : inhérentes aux données ou traduisant les règles d'utilisation de la base

# Exemple

notes



Un étudiant suit au plus 6 cours

Un cours est assuré par un unique enseignant

# Implantation du schéma

- Traduction du schéma conceptuel en un schéma logique dans les concepts du modèle utilisé par le SGBD choisi
- Exemple précédent dans le modèle relationnel :
  - Etudiant : nom, prénom, date de naissance, numéro étudiant
  - Enseignant : nom, prénom, statut, id
  - Cours : nomC, cycle, id\_enseignant
  - Inscription : numéro étudiant, nomC, note1, note2

# Implantation des données

- Choix des structures de stockage des données par les administrateurs
- Schéma interne : description des choix d'enregistrement des données dans les fichiers
  - fait appel à un nouveau modèle où les concepts sont ceux de fichiers, organisation de fichiers, index, chemin d'accès, clé, etc...

# Plan indicatif du cours

- Modèle relationnel (1 séance)
- Diagramme entités/associations (1 séance)
- Datalog (2 séances)
- SQL (2 séances)
- Contrôle continu (1 séance, le 18 novembre)
- Dépendances fonctionnelles et formes normales (3 séances)
- Transactions : concurrence, reprise sur panne (2 séances)

# NB

- Il n'y aura pas cours les
  - 7 octobre et 28 octobre