

Base de données relationnelle

Objectifs :

- Qu'est-ce qu'une BD ?
- Comment on la fabrique
- Comment on s'en sert
- Vocabulaire
 - BD
 - SGBD
 - SGBD-R
 - SQL
 - Intégrité des données
 - Architecture Client-Serveur
- 5 parties
 - Généralités sur les BD
 - Généralités sur les SGBD
 - Architecture des SGBD
 - Petit historique des SGBD
 - Généralités sur le SQL

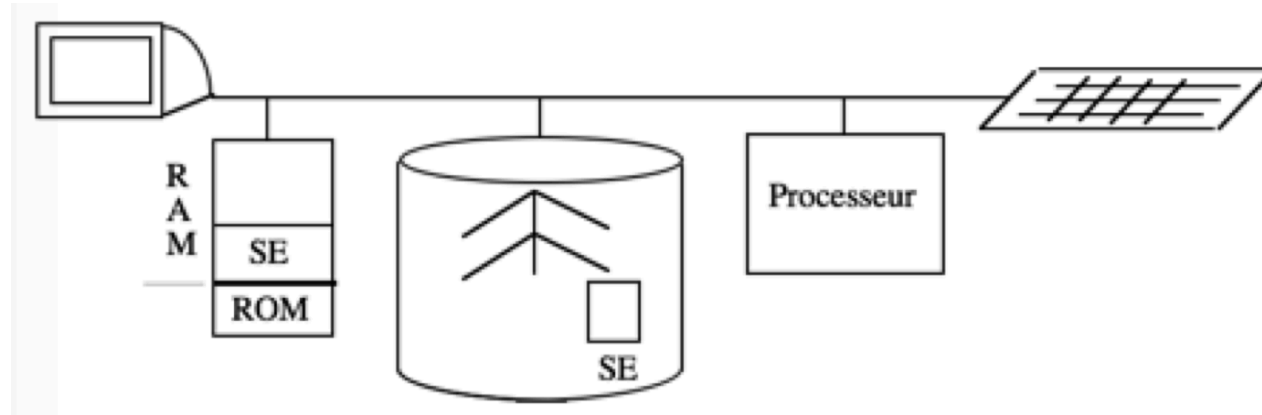
Partie 1

Généralités sur les BD

BD

Requisits

- Aucun !
- Architecture et fonctionnement d'un ordinateur



- Système d'exploitation
- Disque dur, répertoires, fichiers
- Processeur
- RAM

BD

Base de données en général, pas forcément informatisée

- Système qui permet de manipuler de façon plus aisée des objets, des choses, en grand nombre.
- Une BD est constituée de deux éléments :
 - **La BD à proprement parler** : un regroupement de données (d'informations) en grand nombre. Ces données décrivent (représentent) des objets qu'on veut manipuler.
 - **Le SGBD (système de gestion de la base de données)** : c'est un ensemble d'outils de gestion permettant de ranger et d'utiliser les objets qu'on veut manipuler. Ces outils permettent donc d'ajouter, de retirer et de modifier ces objets et les données qui les représentent (partie rangement) mais aussi de les consulter (et donc de les trouver !)

BD

Base de données en général, pas forcément informatisée

- Exemples
 - **Une bibliothèque** : des fichiers d'adhérents et de livres, des index (fichiers pour les recherches).
 - **Un magasin** avec ses catalogues d'inventaires, sa comptabilité.
 - **Un disque dur** : on a des données, on peut les trouver assez facilement
 - **Un cerveau** : on a des données, on les trouve plus ou moins facilement !
 - **Les sociétés d'assurances et les banques**. Cf. les films : la Foule (King Vidor, 1928), la Garçonnière (Billy Wilder, 1960, bureau de société d'assurance) et aussi Brasil (Terry Gilliam, 1985, utilisation de tubes pneumatiques)

BD

Base de données informatisée

- Les données et les outils de gestion sont informatisées : c'est la situation actuelle !
- Exemples
 - **Une bibliothèque** : les adhérents et les catalogues sont gérés par informatique. Le retrait et le retour des livres est géré via une machine qui scanne les codes-barres des livres. Les informations (livres disponibles, livres actuellement empruntés par l'adhérents, etc.) sont accessibles sur des terminaux dans la bibliothèque et sur internet.
 - **Le système de réservation de la SNCF** : il permet d'accéder à tous les trains et à toutes les places dans les trains, de savoir si elles sont disponibles, etc.
 - **Les sites internet de journaux** (Le Monde, le Figaro, etc.) : les articles et les commentaires utilisent des bases de données.
 - **Les sites web en général** utilisent des bases de données.

Partie 2

Généralités sur les SGBD

SGBD

Généralités

- Le SGBD, Système de Gestion de Base de Données, est le logiciel qui réalise les outils de gestion de la base de données.
- Il permet essentiellement de :
 - Définir les données qu'on veut manipuler,
 - Ajouter, modifier, supprimer des données
 - Consulter des données
- Il réalise les objectifs de la BD : intégrité, distinction données / traitements et performances

BD et SGBD

Problème de vocabulaire

- Quand on parle de base de données (BD), on parle :
 - Soit des données uniquement. Par exemple : « la BD des utilisateurs a été piratée ».
 - Soit de l'ensemble constitué par les données et le SGBD. Par exemple : « un Cours de BD »
 - Soit du SGBD uniquement. Par exemple « j'utilise une BD MySQL »

BD et SGBD - Principe fondamental :

la distinction entre les données et les traitements

- Modèle des données indépendant des traitements
 - La BD offre un modèle synthétique des données indépendant des traitements qui leur sont appliqués.
 - Ce modèle permet une meilleure compréhension du système d'informations représenté, que ce soit celui d'une entreprise ou d'un système scientifique.
 - L'information correspondant à une donnée doit être directement intelligible : elle ne correspond pas à un calcul ou à un codage (pas de code spéciaux dans la BD).
- Traitements rationalisés qu'on applique aux données
 - Une fois les données définies, les traitements se ramènent essentiellement à ajouter, retirer, modifier et consulter les données.
 - A cela, s'ajouteront ensuite les algorithmes de transformations (calculs) plus ou moins compliqués.

BD et SGBD

2 objectifs majeurs

- **Garantir l'intégrité des données**
- **Assurer la performance.**

SGBD - objectifs majeurs

L'intégrité des données - 1

❖ Problématique fonctionnelle :

- BD = **réservoir d'informations unique et partagé** dont l'intégrité est garantie.
- Garantir l'intégrité des données = **éviter l'altération et l'incohérence** des données.

❖ L'altération des données

- Il y a plusieurs sources d'altération possibles : l'usure, les pannes, les erreurs, les malveillances. Une BD (et un SGBD) aura comme objectif d'en limiter la possibilité.

❖ L'incohérence des données

- Une donnée est incohérente si elle est contradictoire avec une autre donnée.
- Il y a deux grands types d'incohérence :
 - **La duplication des données** avec des valeurs différentes. Exemple : deux adresses différentes pour une même personne.
 - **Les valeurs aberrantes.** Exemples : un âge négatif ou supérieur à 150 ; une donnée faisant référence à une autre donnée qui n'existe pas.

SGBD - objectifs majeurs

L'intégrité des données - 2

Problématique technique :

- Conformité au modèle : vérifier que les données saisies sont cohérentes par rapport à la définition qu'on en a donnée.
- Accès concurrents des utilisateurs : éviter des incohérences dues à des modifications faites en même temps sur une même donnée.
- Gestion des pannes : garantir le maintien de la cohérence des données, même quand une panne intervient au cours d'une période de modification des données.
- Sécurité : gérer des droits des utilisateurs pour limiter les possibilités des utilisateurs sur les données.

SGBD - objectifs majeurs

Performance et optimisation

- Une BD doit fournir des **performances** acceptables par l'utilisateur. C'est la problématique de **l'optimisation**. C'est traité essentiellement par la notion d'**index**.

SGBD - objectifs majeurs

Performance et optimisation

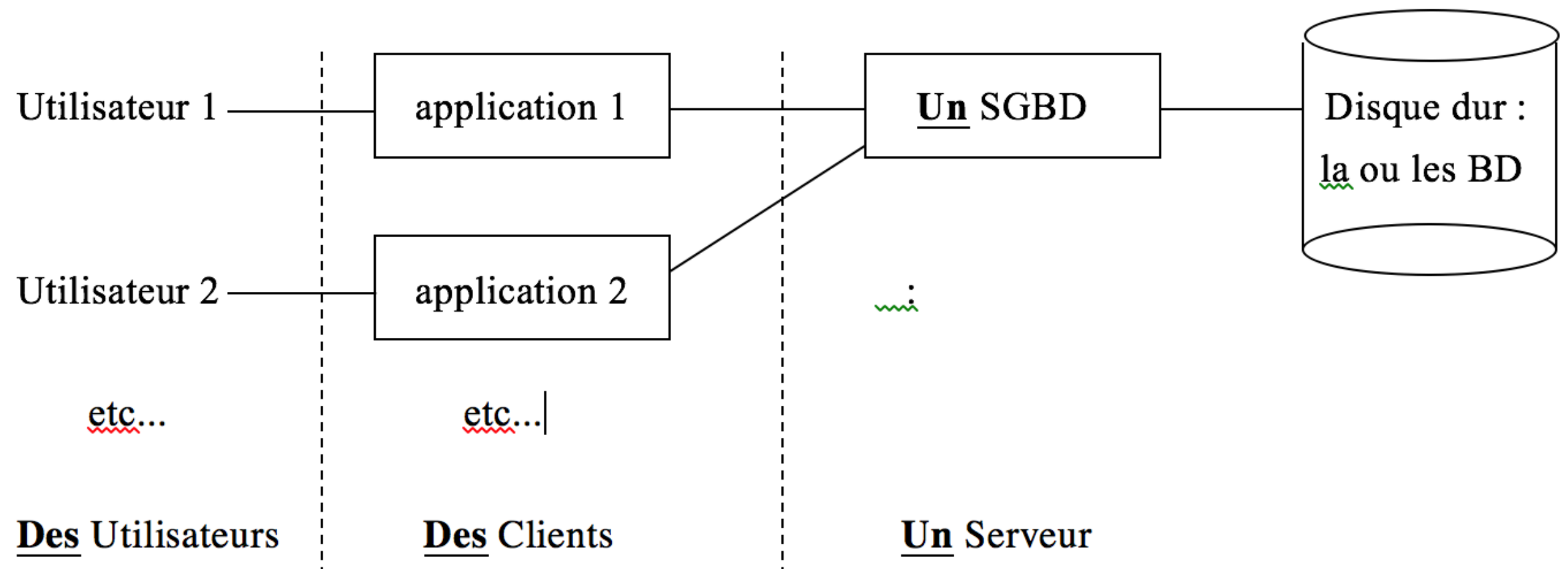
- L'objectif de la performance se déploie sur plusieurs registres :
 - La performance de calcul : les SGBD traitent les requêtes plus ou moins efficacement permettant ainsi une réponse plus ou moins rapide.
 - Les index : ils permettant d'améliorer les performances.
 - La gestion des mémoires cache : la question des accès disque est centrale pour la performance. La gestion de mémoires cache permettant de limiter les accès disques.
- Arbitrage entre sécurité et performance
 - Sécurité et performance sont en partie contradictoires.
 - Certains SGBD sont plus orientés sécurité (ORACLE), d'autres plus orientés performance (MySQL).

Partie 3

Architecture des SGBD

SGBD - Architecture

- Modèle fondamental : architecture client serveur



SGBD - Architecture

- Modèle fondamental : architecture client serveur
 - Une BD est stockée sur un ou plusieurs disques durs.
 - Pour une BD, il y a un SGBD et un seul. Le SGBD et la BD sont en général sur la même machine.
 - Le SGBD peut gérer plusieurs BD.
 - Le SGBD est le serveur. Chaque application est un client. Plusieurs clients peuvent communiquer avec le serveur.
 - Une application est un programme qui envoie des commandes au SGBD pour manipuler d'une façon ou d'une autre les données de la BD sur le disque dur. Par exemple : un programme PHP enregistre une donnée dans la BD ou récupère des informations dans la BD.

SGBD - Architecture

- Qu'est-ce qu'un serveur ?
 - Un serveur est un programme qui n'a pas de communication directe avec un utilisateur humain.
 - Un serveur est un programme qui communique uniquement avec d'autres programme. Il est en attente (il écoute avec un « listener ») de messages qui lui sont envoyés par d'autres programme. Il répond aux messages qu'il a reçu : il « sert » ses « clients ».
 - Par abus de langage, la machine qui héberge le serveur (et le fait fonctionner) est appelée serveur. On peut aussi parler de machine-serveur hébergeant un ou plusieurs serveurs.

SGBD - différents types d'objets gérés

- Un SGBD gère un certain nombre d'objets :
 - Des bases de données : pour distinguer des applications différentes
 - Des tables : pour mettre les données
 - Des index : pour optimiser les requêtes
 - Des contraintes d'intégrité : pour garantir la cohérence des données
 - Des utilisateurs : pour se connecter à la BD et leur donner des droits particuliers
 - Des droits : pour limiter l'accès des utilisateurs à la BD
 - Etc.
- Pour chacun de ces objets, le SGBD permet la création, modification, suppression, la consultation.

SGBD - 3 types d'utilisateur

- **L'administrateur**

- Il gère les accès et les droits des utilisateurs : problème de sécurité.
- Il gère aussi le bon fonctionnement des bases de données : problème d'efficacité.

- **Le développeur**

- Il programme une application qui sera cliente du SGBD. Il a accès à la BD à travers l'application cliente dite : « calculatrice SQL ».

- **L'utilisateur final**

- Il utilise une application cliente du SGBD. En général, il n'a accès à la BD qu'à travers une application cliente spécifique.

SGBD - 2 grands types d'applications

- **1 : L'informatique « transactionnelle » : OLTP**
 - Les applications d'informatique transactionnelle regroupent à la fois les applications de gestion (système d'information, site marchand, etc.) et des applications industrielles qui utilisent une BD. Ce sont toujours des applications de grosse taille.
 - On parle aussi d'OLTP : On-Line Transaction Processing

SGBD - 2 grands types d'applications

- **2 : L'informatique « décisionnelle » : OLAP**

- Les applications d'informatique décisionnelle servent à exploiter les données le plus souvent issues d'applications d'informatique transactionnelle.
- Ces applications mettent en œuvre des notions de statistiques, d'analyse de données et de datamining. Elles peuvent s'appuyer sur des fonctionnalités SQL proposées par le SGBD (Group By, Pivot, etc.).
- Ces applications sont en général de petite taille. Leur finalité est l'aide à la décision.
- Les notions de «Business Intelligence » : BI, data-mining et data-warehouse sont des notions connexes à ce domaine.
- On parle aussi d'OLAP : On-Line Analytical Processing.
- Le Big Data, c'est l'utilisation du BI sur des données personnelles. Il vient du développement des smartphones.

Partie 4

Petit historique des SGBD

SGBD-Relationnel

historique des SGBD

- Il existe différents types de BD en fonction du modèle d'organisation des données qu'elles utilisent.
- A ce modèle est associé un langage de manipulation des données.
 - Années 60 : les anciens modèles. Modèle hiérarchique, modèle réseau, langages d'accès navigationnels permettant de circuler dans des structures de types graphes.
 - 1970 : Création théorique du modèle relationnel. Article de CODD.
 - Années 80 : Premiers SGBD-Relationnels sur le marché (ORACLE, Ingres, Informix, sybase, Dbase IV, etc.). 1979 : première version d'ORACLE. Création du SQL.
 - Années 90 : Premiers SGBD-Objets. XML.
 - Fin des années 2000 : Premiers SGBD « NoSQL ». Le NoSQL se développe dans des environnements à données dont la structure est très dynamique et pour lesquelles la question de l'intégrité n'est pas primordiale.
 - Aujourd'hui : L'Objet n'a pas remplacé le Relationnel. Le marché est dominé par les SGBD-R (ORACLE, SQL-Server (microsoft), DB2 (IBM), MySQL, PostgreSQL, ACCESS...). Les technologies sont parfois mixtes (Relationnel et objet). Le XML cohabite avec le SQL. Le « NoSQL » se développe.

SGBD-Relationnel

GBDR-R

- Aujourd'hui, le modèle relationnel domine le marché et la théorie des bases de données.
- Les SGBD correspondant sont des SGBD-R.

SGBD-Relationnel

Le marché

Parts de marché en 2012

| Base de données | Part de marché |
|----------------------|---|
| Oracle | 44 % |
| IBM DB2 | 20 % |
| Microsoft SQL Server | 19 % |
| Total | 28.69 milliards de dollars +7.8% en 1 an |

Usages

| Base de données | % |
|-----------------|------|
| MySQL | 58 % |
| Oracle | 37 % |
| PostgreSQL | 32 % |
| MS SQL Server | 31 % |

- Oracle a racheté Sun en 2009 qui avait racheté MySQL en 2008.
- MariaDB est la continuation open source du projet MySQL par ses fondateurs.

Partie 5

Généralités sur le SQL

SQL - Présentation - 1

- SQL signifie Structured Query Language, c'est-à-dire langage structuré de requêtes.
- Le SQL est LE langage qui permet d'envoyer toutes les requêtes possibles à un SGDB-R : création, modification, suppression, consultation, etc.
- Le SQL un langage de programmation fondé sur une théorie mathématique : l'algèbre relationnelle.

SQL - Présentation - 2

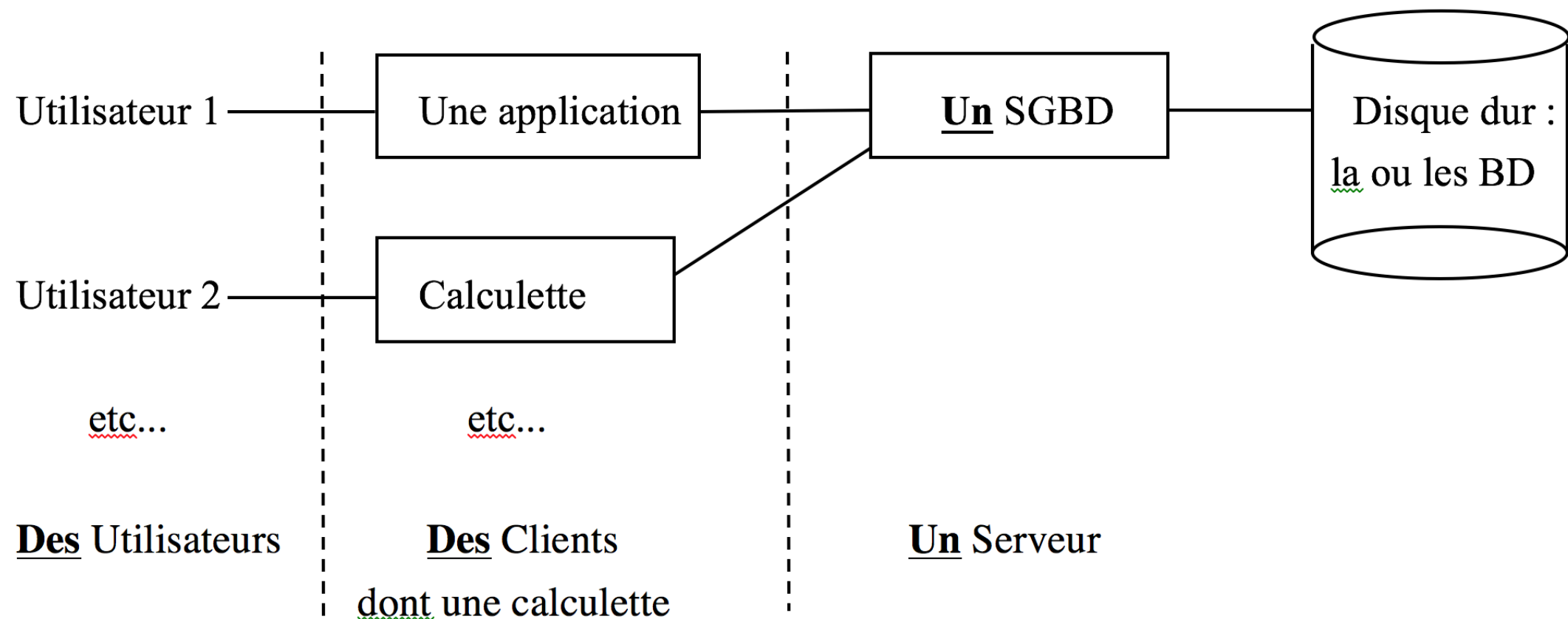
- L'algèbre relationnelle est la théorie des opérations qu'on peut appliquer à un objet mathématique particulier : les tableaux de données (appelées « relations » en algèbre relationnelle).
- Le SQL est un ensemble d'opérateurs (qui traduisent les opérations de l'algèbre relationnelle).
- Ces opérateurs permettent d'écrire les commandes pour créer la BD et pour la consulter.
- Ce n'est pas un langage de programmation qui permet d'écrire des programmes avec des interactions avec l'utilisateur (comme le langage PHP ou le Python par exemple)

SQL - La calculette SQL

- L'application cliente du SGBD qui permet d'utiliser directement le SQL peut être considérée comme une « calculette SQL » qui, au lieu de travailler sur des nombres, travaille sur des tableaux de données.
- Il existe plusieurs formes de cette calculette :
 - les versions « en ligne de commande » (non graphique) : le « Command Line Client » et ses dérivés : mysql, psql (Postgresql), SQL*PLUS (ORACLE), etc.
 - les version graphiques fournies par le SGBD : MySQL GUI tools, ORACLE developer, SQL Server Management Studio Express (SSMSE), etc.
 - les versions graphiques fournies par le serveur web (PHP-MyAdmin, phpPGAdmin, etc.

SQL - La calculette SQL

- La calculette est une application cliente :



- Installer un SGBD c'est installer au moins un serveur et une calculette. On accède au serveur avec la calculette.

SQL - Le PL-SQL

- Dans la plupart des SGBD-R, il existe un langage de programmation à proprement parler qui permet d'écrire des successions d'opérations relationnelles (SQL) sous la forme d'un programme en utilisant les mécanismes usuels de la programmation impérative type PHP, C, etc. à savoir : les variables, les affectations, les tests, les boucles et les fonctions. Ce langage s'appelle : le PL-SQL, c'est-à-dire : Programmation Language SQL.
- Le PL-SQL ne permet pas de développer une interface utilisateur très élaborée. Il ne remplace donc pas le développement d'une application cliente.
- Le PL-SQL permet de :
 - éviter aux applications clientes de répéter les mêmes séries d'opérations SQL : principe de modularité
 - mieux gérer l'intégrité des données (triggers et règles).
 - séparer les modules de calcul des modules d'affichage

SQL - Historique

- Années 70 : premiers prototypes de SQL à la suite de l'article de CODD.
- 1979 : première version de SQL, proposé par ORACLE.
- 1986 : SQL ANSI (American National Standard Institute)
- 1989 : SQL-1, ISO et ANSI (International Standard Organisation)
- 1992 : SQL-2, ISO et ANSI (jointures externes et « join »)
- 1999 : SQL-3, ISO et ANSI (SQL orienté objet)

SQL - Pérennité

- Le SQL existe depuis 1980 en pratique et 1970 en théorie : c'est un langage pérenne : il a vocation à durer ! Pourquoi ?
 - Qui dit SI dit BD : la BD est le cœur du SI.
 - L'algèbre relationnelle est une théorie mathématique simple et solide (et non pas une technologie).
 - L'implantation massive du SQL rend son élimination coûteuse.
 - Le SQL est un langage commun à tous les SGBD-R (normalisation ANSI et ISO)
 - De nombreux Framework de développement propose des outils qui permettent de « s'abstraire » du choix du SGBD et du SQL en général.

SQL - Pérennité

- Pour détrôner le SQL, il faudrait :
 - Une meilleure théorie : pas facile ! L'algèbre relationnelle c'est des maths pures !
 - Des outils aussi puissants qui garantissent la même intégrité: pas facile !
- Les évolutions correspondent à des usages particuliers :
 - Le XML est adapté aux données hiérarchisées avec peu de modifications et de requêtes complexes. Le JSON s'installe sur ce marché.
 - Le NoSQL est adapté aux données dynamiques. Le sigle NoSQL est passé de la signification : pas de SQL (Not SQL) à pas seulement du SQL : Not Only SQL !
 - Le SQL est adapté à tout le reste ! Et particulièrement aux données de l'entreprise en général.
- Aujourd'hui (années 2010), on peut parier sans grand risque que le SQL sera toujours présent d'ici à la fin de la carrière des étudiants !