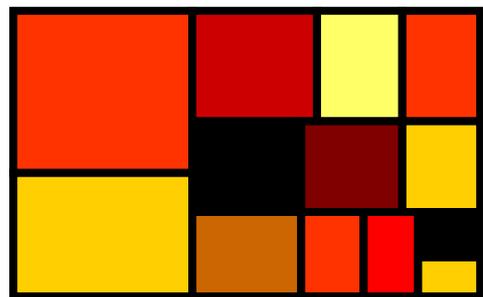
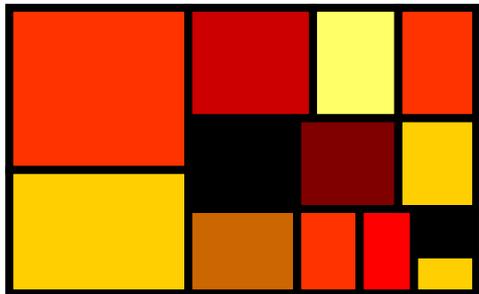


Initiation aux Bases de données



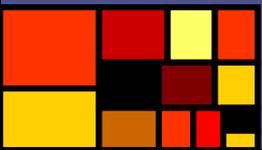
- I. Notions élémentaires
- II. Les Systèmes de Gestion de Bases de Données (SGBD)
- III. MySQL, SGBD populaire, par l'exemple

Notions élémentaires



- I. « Informatique »
- II. Représentation des données
- III. Formats de date
- IV. Formats de nombre
- V. La Tabulation et le texte tabulé
- VI. Noms de fichiers, extension, « ouvrir avec »

« Informatique » : aspects linguistique



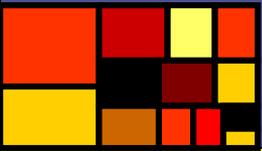
- Définition Petit Robert :
 - Science de l'information, ensemble des techniques de la collecte, du tri, de la mise en mémoire, de la transmission et de l'utilisation de l'information.
- Matériel : « *computer* », « *ordinateur* »
- Exemples :
 - Je tape une lettre avec *Word*[®]. Je joue à *Street Fighter*[®]. Je corrige une image avec *Photoshop*[®].
 - Fais-je de l'informatique ? Oui Non
 - Je conçois un formulaire papier, avec procédures de remplissage, stockage, MAJ, tri et réédition
 - Fais-je de l'informatique ? Oui Non

Représentation des données (1)

- Exercice - Soient 10 véhicules ainsi répartis :
 - 4 camions dont 1 rouge et 3 jaunes
 - 6 voitures dont 3 blanches, 2 rouges et 1 jaune
- Représentez ces données à l'aide d'un tableau. Essayez plusieurs possibilités et tentez de leur donner un nom.

Représentation des données

- option 1 -



Nature	Couleur
Camion	Rouge
Camion	Jaune
Camion	Jaune
Camion	Jaune
Voiture	Blanc
Voiture	Blanc
Voiture	Blanc
Voiture	Rouge
Voiture	Rouge
Voiture	Jaune

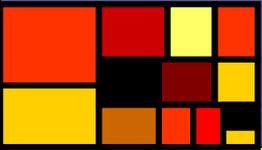
Nature	Couleur	Count(*)
Camion	Rouge	1
Camion	Jaune	3
Voiture	Blanc	3
Voiture	Rouge	2
Voiture	Jaune	1

Coul\Nat	Camion	Voiture
Rouge	1	2
Blanc	0	3
Jaune	3	1

Table standard :

- Pas joli
- Universellement utilisable
- 1 ligne = 1 élément
- Les entêtes de colonnes sont des noms de variables

Représentation des données – option 2 –



Nature	Couleur
Camion	Rouge
Camion	Jaune
Camion	Jaune
Camion	Jaune
Voiture	Blanc
Voiture	Blanc
Voiture	Rouge
Voiture	Rouge
Voiture	Jaune

Nature	Couleur	Count(*)
Camion	Rouge	1
Camion	Jaune	3
Voiture	Blanc	3
Voiture	Rouge	2
Voiture	Jaune	1

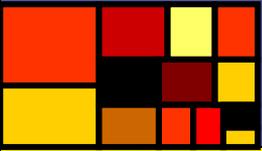
Coul\Nat	Camion	Voiture
Rouge	1	2
Blanc	0	3
Jaune	3	1

Table agrégée :

- peu joli
- Assez utilisable
- 1 ligne = 1 groupe d'éléments semblables
- Fonctions alors utilisable : nombre de, nombre de différents, somme, moyenne, variance, min, max
- Les entêtes de colonnes sont des noms de variables

Représentation des données

- option 3 -



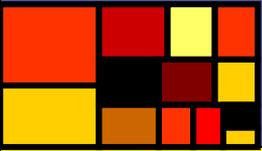
Nature	Couleur
Camion	Rouge
Camion	Jaune
Camion	Jaune
Camion	Jaune
Voiture	Blanc
Voiture	Blanc
Voiture	Blanc
Voiture	Rouge
Voiture	Rouge
Voiture	Jaune

Nature	Couleur	Count(*)
Camion	Rouge	1
Camion	Jaune	3
Voiture	Blanc	3
Voiture	Rouge	2
Voiture	Jaune	1

Nature Couleur	Camion	Voiture
Rouge	1	2
Blanc	0	3
Jaune	3	1

- Tableau croisé :
- Intuitif, humain
 - Totalement inutilisable !!
 - Les entêtes de lignes et colonnes sont des *modalités* de variables

Représentation des données – modèle entité association –



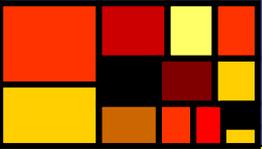
Nature	Couleur
Camion	Rouge
Camion	Jaune
Camion	Jaune
Camion	Jaune
Voiture	Blanc
Voiture	Blanc
Voiture	Blanc
Voiture	Rouge
Voiture	Rouge
Voiture	Jaune

Nous venons d'appliquer la méthode entité association :

- Quelle est l'entité de base ?
 - Le véhicule => une ligne par véhicule
- Quelle sont les propriétés/attributs de chaque véhicule ?
 - La nature et la couleur => une colonne par attribut

Autres représentations

- binarisation d'un variable qualitative -

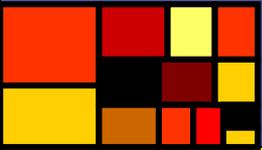


Est_voiture	Est_camion	Est_rouge	Est_jaune	Est_blanc
0	1	1	0	0
0	1	0	1	0
0	1	0	1	0
0	1	0	1	0
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	1	0	0
0	0	1	0	0
0	0	0	1	0

- Utilisé en Statistiques : transformation d'une variable qualitative en autant de variables binaires que de modalités
- Si le nombre de modalités est défini, $k-1$ variables binaires suffisent à représenter une variable à k modalités

Autres représentations

- modèle entité attribut valeur -



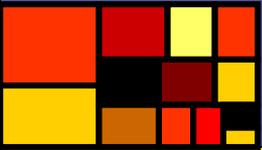
Individu	Attribut	Valeur
1	Nature	Camion
1	Couleur	Rouge
2	Nature	Camion
2	Couleur	Jaune
3	Nature	Camion
3	Couleur	Jaune
4	Nature	Camion
4	Couleur	Jaune
5	Nature	Voiture
5	Couleur	Blanche
...

- Utilisé en Informatique (biologie, eCommerce)
- permet un stockage de paires attribut-valeur par individu
- aucun a priori sur la liste des attributs

Formats de date (1)

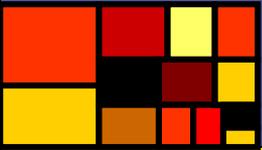
- Les formats administratifs :
 - Format français : jj/mm/aaaa
 - Format américain : m/j/aaaa
- Problèmes :
 - Ambiguïté si $jj \leq 12$
 - Microsoft Excel : conversions automatiques
 - Pas de classement alphabétique
 - Longueur variable
- Le format standard SQL : aaaa-mm-jj (avec des zéros initiaux au besoin)

Formats de date (2)



- Exemple de nombre : **1284**
 $= 1*10^3 + 2*10^2 + 8*10^1 + 4*10^0$ unités
- Exemple d'heure : **06h 24' 36''**
 $= 6*60^2 + 24*60^1 + 36*60^0$ secondes
- Exemple de date SQL : **2006-06-05**
 $= 2006*365.25 + 06*30.4 + 05*1$ jours
(Ceci est une approximation didactique)

Formats de date (3)

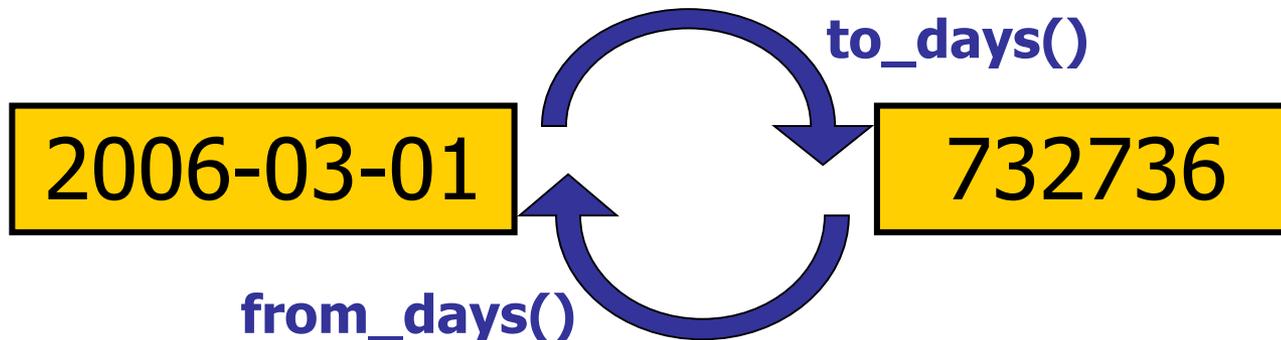


- Utilisez donc toujours le format universel : **aaaa-mm-jj**
- Avantages de ce format :
 - univoque : le seul avec des tirets
 - universel en informatique
 - classement alphabétique (votre courrier...ex ci-contre)

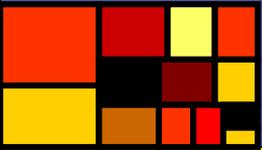


Réaliser des additions et soustractions sur des dates (1)

- on utilise une fonction du calendrier (grégorien ou julien)
- retourne le nombre de jours écoulés depuis une date de référence très ancienne.
- Mécanisme de calcul universel
 - Timestamp UNIX (MySQL) : nombre de jours depuis le 01/01/1970
 - Excel : nombre de jours depuis le 01/01/1900
 - PHP : nombre de jours depuis le début du calendrier julien



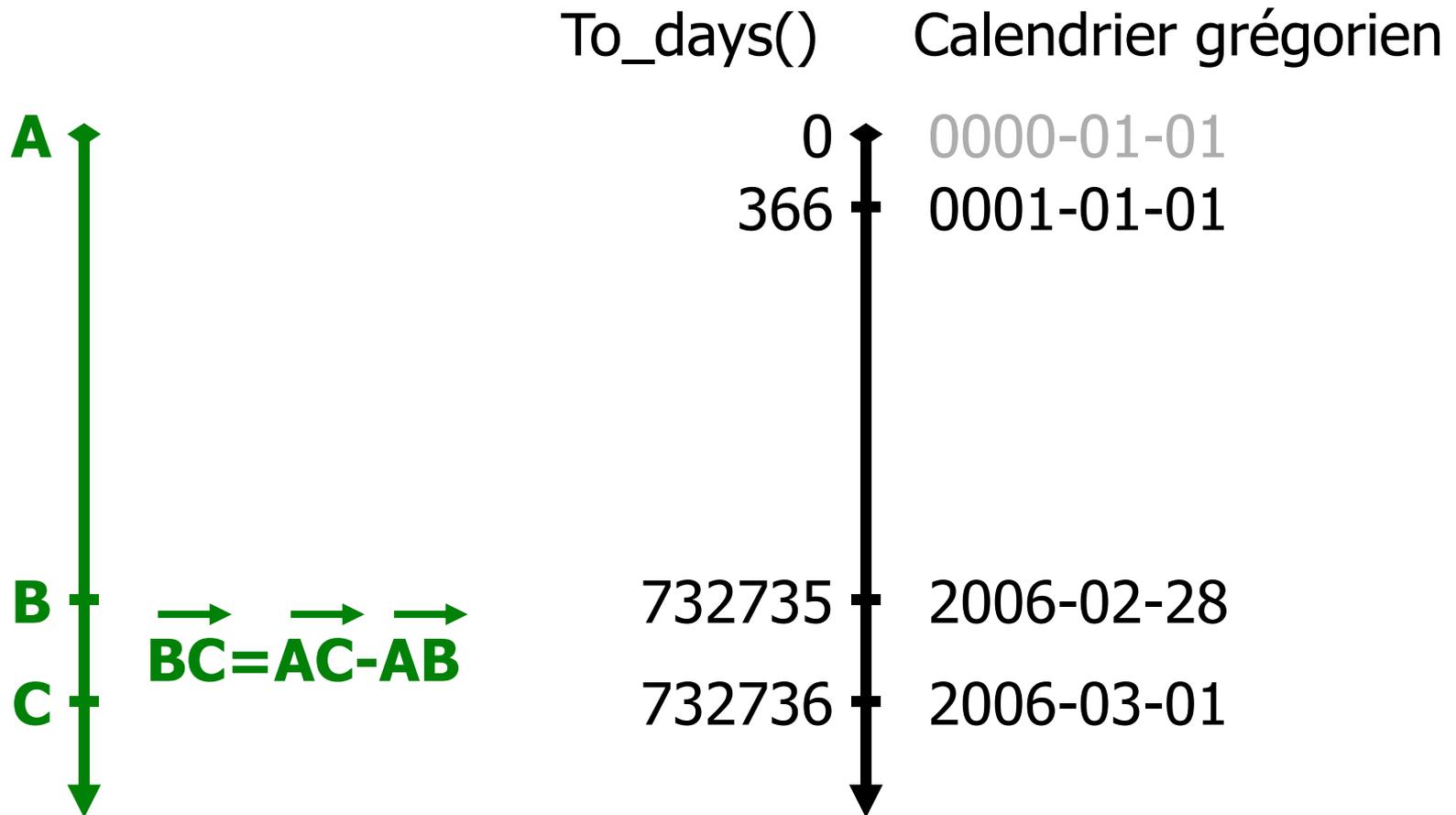
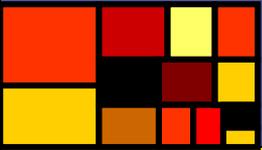
Réaliser des additions et soustractions sur des dates (2)



■ Exercice :

- Lancer le serveur MySQL
- Lancer la console d'interrogation shell
- Calculer les résultats suivants en associant une explication textuelle à chaque calcul :
 - `Select to_days("2006-03-01") ;`
 - `Select from_days(732736) ;`
 - `Select to_days("2006-03-01") - to_days("2006-02-28") ;`
 - `Select from_days(to_days("2006-02-28") + 1) ;`

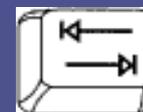
Réaliser des additions et soustractions sur des dates (3)



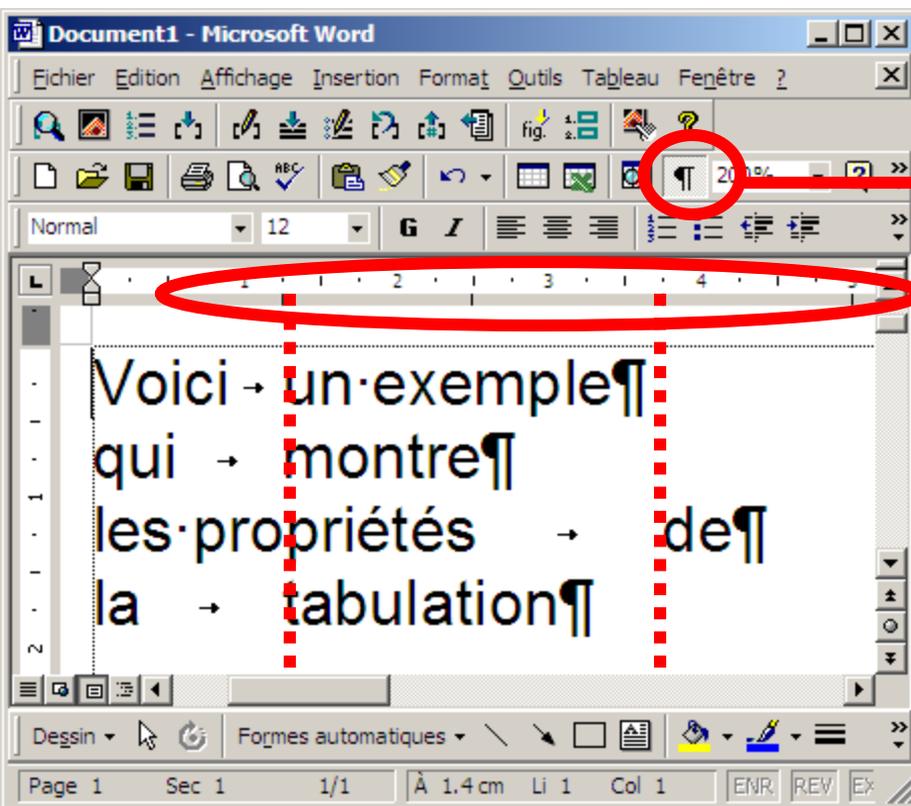
Formats de nombre

- Conventions universelles :
 - Un nombre est représenté par un ou plusieurs chiffres
0123456789
 - Comprenant éventuellement un ou plusieurs de ces symboles (maximum une fois chacun) : « - », « E », « . »
 - Sont en particulier illicites : l'espace, la virgule
- Exemples :
 - Ceci est un nombre : [1023] [0.2456]
 - Ceci est souvent compris : [.2456] [1.23E-06]
 - Ceci n'est pas un nombre : [1 023] [123.145,4]
- [démonstration]

La TABULATION



- C'est un caractère d'espacement de longueur variable
- Invisible par définition



La tabulation fera au moins 1 espace de large, mais le logiciel l'élargira jusqu'à ce qu'elle atteigne le taquet de tabulation suivant.



Le format « texte tabulé »

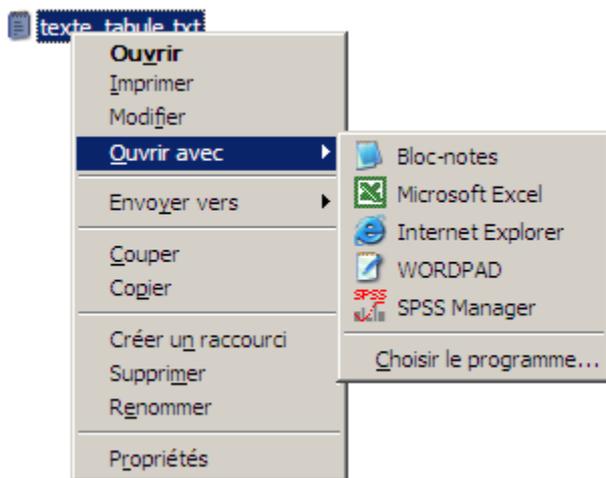
- Texte tabulé = format universel de représentation des tables simples :
 - Les lignes sont séparées par des sauts de ligne
 - Les champs sont séparés par une tabulation
- Un autre format très fréquent : le CSV
 - Les champs sont séparés par des virgules « coma separated text », et parfois encadrés par des guillemets doubles.
 - Cette représentation permet d'embarquer des sauts de ligne et des tabulations dans les valeurs
- Exercice : copier le contenu du fichier `texte_tabule.txt` vers Excel, et vice-versa.

Noms de fichiers

- Les noms de fichiers sont moins normés depuis 1995 (sortie de Windows 95)
- Seuls caractères interdits `/:\\?*<>"|`
- Norme toutefois conseillée :
 - Deux parties séparées par un seul point
 - Le nom de base : utilisez uniquement az09_- et évitez les majuscules, les caractères accentués, les espaces, les ponctuations
 - L'extension : 2 à 5 lettres ou chiffres az09

Forcer l'ouverture d'un fichier par un programme (1)

- Windows XP propose d'associer plusieurs actions à un type de fichiers
 - Le clic droit vous affiche les actions possibles
 - L'action par défaut est en gras : elle est associée au double-clic
 - Le menu « ouvrir avec » vous permet d'afficher les programmes utilisables



Clic droit sur un fichier...

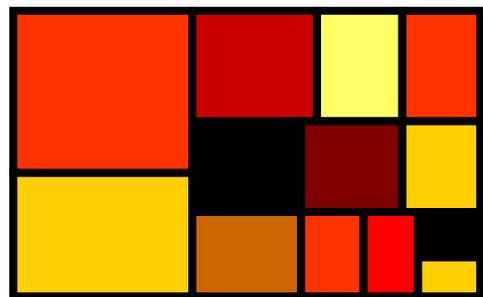
Forcer l'ouverture d'un fichier par un programme (2)

- Exercice : ouvrez le fichier *texte_tabule.txt* avec Microsoft Excel en ajoutant ce programme à la liste « ouvrir avec », à l'aide d'un clic droit sur le fichier.
- Autre astuce très pratique : ajouter le bloc-notes à « envoyer vers » (disponible pour tout type de fichier !)
 - En utilisant l'explorateur de fichiers
 - activer l'affichage des fichiers et dossiers cachés : outils > options des dossiers > affichage > afficher les fichiers et dossiers cachés
 - trouver le dossier caché *C:\Documents and Settings\[utilisateur]\SendTo* et ajouter un raccourci vers le bloc-notes

L'extension du fichier

- Vous devriez demander à votre explorateur de toujours l'afficher
 - Outil > option des dossiers > affichage > décocher la case « masquer les extensions des fichiers dont le type est connu »
- L'extension indique au système quelle icône et quelles actions associer au fichier
- Exercice : après avoir forcé l'affichage des extensions, renommez le fichier *texte_tabule.txt* en *texte_tabule.xls* puis double-cliquez dessus. Rendez-lui son nom initial puis double-cliquez dessus.

Les Systèmes de Gestion de Bases de Données



- I. Motivations de ce cours
- II. Architectures
- III. Logiciels connus
- IV. Schéma relationnel de données

Motivations de ce cours (1)

- Vous utiliserez peut-être des SGBD :
 - Microsoft Access
 - SGBD divers interrogés par Business Objects®
 - Conception de sites web avec PHP & MySQL
- Vous mènerez des travaux sur des données importantes (thèse) :
 - Microsoft Excel ne permet ni traçabilité ni réversibilité, calculs limités, pas de schéma relationnel, max 64000 lignes
 - Les logiciels de statistiques émulent tous des pseudo-SGBD... avec des langages propriétaires peu puissants, à réapprendre à chaque fois

Motivations de ce cours (2)

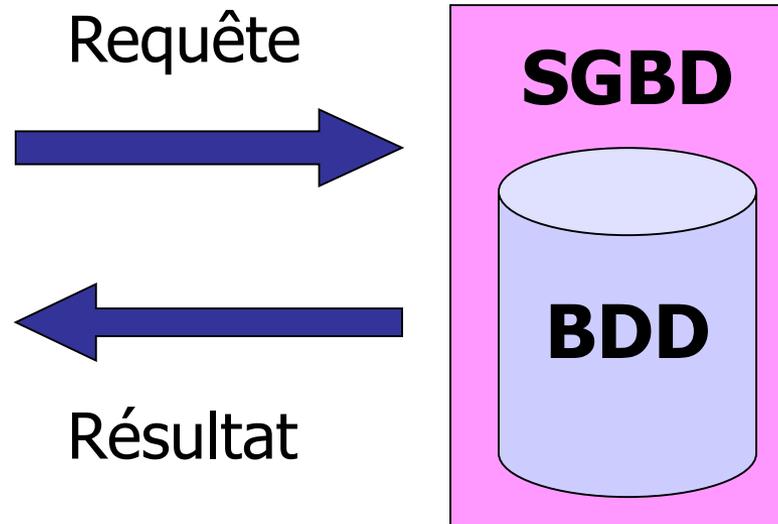
- En apprenant les rudiments du langage SQL :
 - Vous apprendrez un langage utile toute votre vie car quasiment universel
 - Vous comprendrez mieux les erreurs faites en utilisant Business Objects
 - Vous serez un utilisateur de SAS moins malheureux
- En utilisant directement un serveur SQL :
 - Vous ferez des traitements complexes, en toute traçabilité et réversibilité
 - Vous pourrez importer et exporter vos données facilement en texte tabulé
 - Votre logiciel de statistiques pourra interroger directement le serveur ! Des adaptateurs (packages) existent dans tous les « grands » logiciels de statistiques

Les serveurs de bases de données – architectures (1)

```
SELECT clinique_id,  
       bio_enc, ecart_relatif  
FROM biologie  
LIMIT 0,10 ;
```

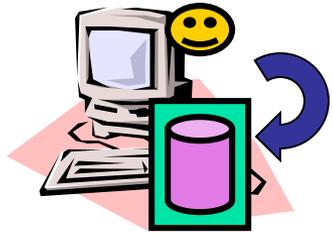
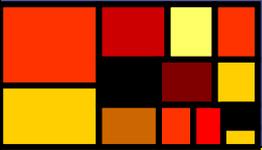
clinique_id	bio_enc	ecart_relatif
1350	276102	-0.4211
1650	445748	-0.0374
1660	364835	0.5562
1770	175604	-0.3615
ORL	351973	-0.5843
2070	251873	0.9863
2400	811461	0.0503
3420	77981	-0.6991
3740	123726	-0.4826
3800	1193553	0.491

10 rows in set (0.00 sec)

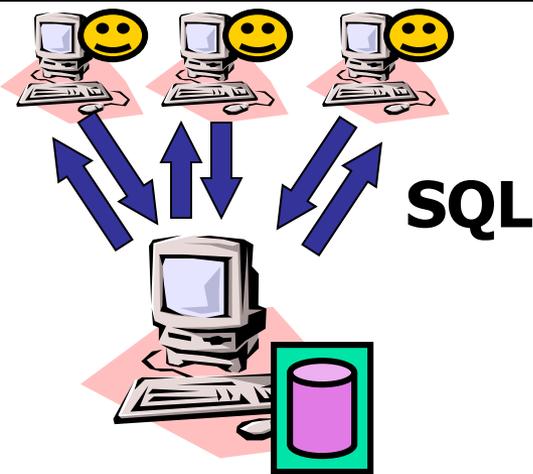


Types de requêtes : création de table, insertion et modification de données, interrogation et calcul.

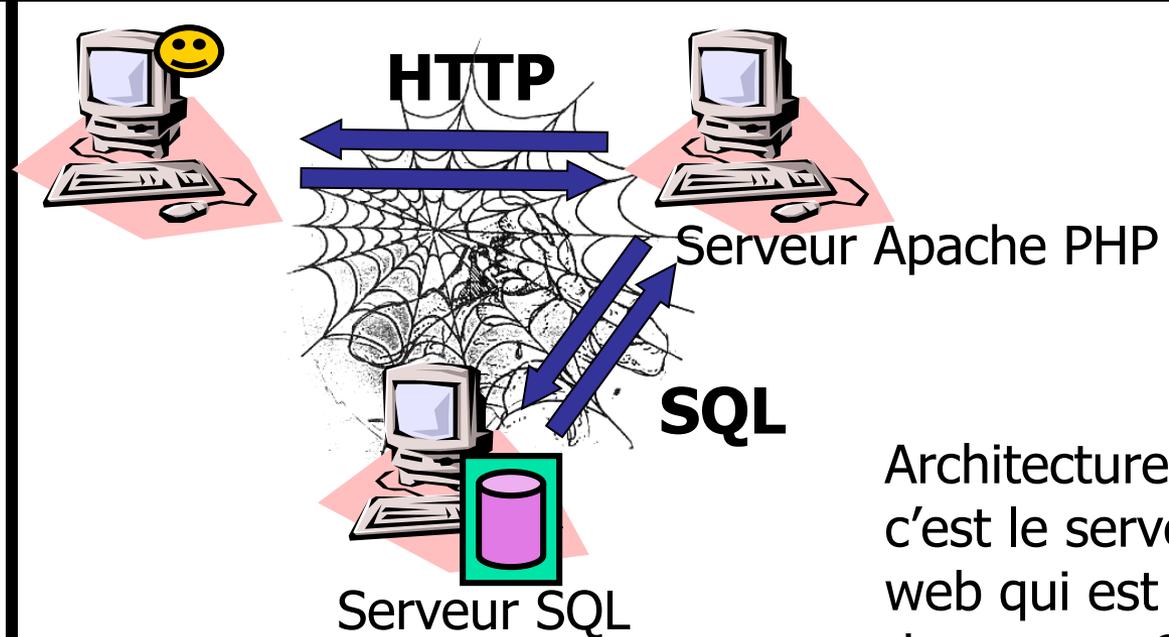
Les serveurs de bases de données – architectures (2)



Architecture basique : un seul poste est utilisé pour l'interrogation, il est à la fois client et serveur. Exemple : Microsoft Accès



Architecture classique : plusieurs postes clients interrogent un serveur.



Architecture web : c'est le serveur web qui est client du serveur SQL

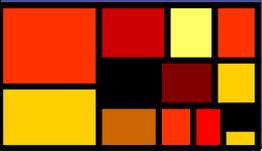
SGBD et grands noms (1)

- Microsoft Access :
 - SGBD sommaire qui ne fonctionne qu'en mode local
 - Toutes les opérations peuvent être faites à la souris. Très intuitif.
- Véritables serveurs de bases de données : Oracle, MySQL, Caché...
- SQLite, un SGBD très particulier :
 - Mode local uniquement, Un fichier = Une BDD
 - Pas de programme mais une DLL, utilisée dans différents langages
- Business Objects, interrogateur intuitif :
 - Ce n'est pas un SGBD, c'est **un interrogateur de SGBD**, utilisable avec tous les grands noms des serveurs SGBD.
 - BO est donc installé sur chacun des postes clients, et pas sur le serveur.
 - Permet d'écrire une requête avec la souris, d'après le schéma décrit par l'administrateur
 - Permet de rapatrier et présenter les résultats

SGBD et grands noms (2)

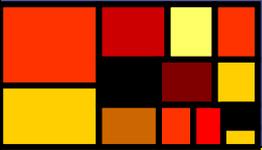
- Produits MySQL gratuits :
 - LE serveur MySQL (également dans EasyPHP)
 - Shell MySQL : interrogation en ligne de commande d'un serveur MySQL, sur la même machine uniquement
 - MySQL Query Browser : simple client d'interrogation d'un serveur MySQL distant
 - MySQL Control Center : logiciel d'administration d'un serveur MySQL distant : sauvegardes, utilisateurs...

SGBD et grands noms (3)



Action possible Nom produit	Design à la souris d'une requête SQL	Saisie d'une requête SQL	SGBD	Réception du résultat brut	Présentation du résultat	Formulaires de saisie et modif des données	Fonctionnement distant
Microsoft Accès	OUI	OUI	OUI	OUI	OUI	OUI	<i>non</i>
Grands serveurs : Oracle, MySQL, Caché	<i>non</i>	OUI	OUI	<i>non</i>	<i>non</i>	<i>non</i>	SERVEUR
SQLite	<i>non</i>	OUI	OUI	OUI	<i>non</i>	<i>non</i>	<i>non</i>
Business Objects	OUI !!	OUI (...)	<i>non</i>	OUI	OUI	<i>non</i>	CLIENT
MySQL Query Browser	<i>non</i>	OUI	<i>non</i>	OUI	<i>non</i>	<i>non</i>	CLIENT

Le Schéma relationnel (1)

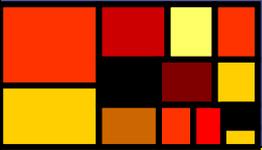


Exemple d'une table, avec une ligne par consultation :

- Problème évident **d'identité** des patients et des médecins

<i>Identité :nom</i>	<i>Identité :prénom</i>	<i>Date consultation</i>	<i>Médecin traitant</i>	<i>Poids</i>	<i>HbA1c</i>
Dupont	Jean	01/03/1998	Dr XXX	82	8.4
Dupont	Jean	08/09/1998	Dr XXX	82	9.2
Martin	Pierre	02/07/2001	Dr ZZZ	70	10.1
Léger	Paul	02/04/2003	Dr ZZZ	62	9.1
Léger	Paul	23/01/2004	Dr ZZZ	62	8.9
Léger	Paul	31/12/2004	Dr ZZZ	60	8.7

Le Schéma relationnel (2)



Mise en schéma relationnel :

Table des patients

<i>id</i>	<i>nom</i>	<i>prénom</i>
101	Dupont	Jean
102	Martin	Pierre
103	Léger	Paul

Table des consultations

Relation événements-patients



<i>id</i>	<i>id_Patient</i>	<i>Date consultation</i>	<i>id_médecin</i>	<i>Poids</i>	<i>HbA1c</i>
1	101	01/03/1998	501	82	8.4
2	101	08/09/1998	501	82	9.2
3	102	02/07/2001	502	70	10.1
4	103	02/04/2003	502	62	9.1
5	103	23/01/2004	502	62	8.9
6	103	31/12/2004	502	60	8.7

Relation événements-médecins

Table des médecins

<i>id</i>	<i>Médecin traitant</i>
501	Dr XXX
502	Dr ZZZ

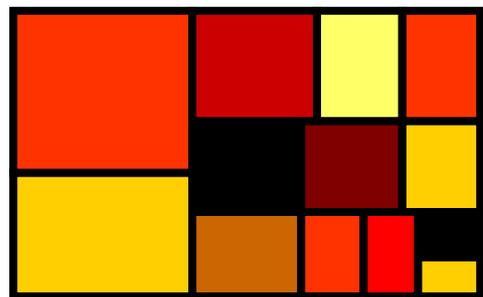
Le Schéma relationnel (3)

- Qualités majeures (dans cet exemple) :
 - Garantie de l'identité des patients
 - Un patient déjà connu n'a pas à être saisi, donc pas de différence d'orthographe
 - Garantie de l'intégrité des données
 - Un patient ne peut pas être supprimé tant que des consultations y font référence (si Foreign Key)
 - La modification d'un patient se fait en un seul endroit et affectera automatiquement toutes les consultations

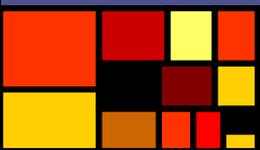
Le Schéma relationnel (4)

- Défauts du schéma relationnel
 - Structure figée à définir au début
 - Le nombre de combinaisons possibles ralentit les requêtes
 - placer des index UNIQUE et PRIMARY KEY sur les tables
 - Erreurs possibles : trop ou pas assez de lignes
 - Connaître le schéma de la base
 - Faire des tests sur des extraits
 - Ne pas méconnaître la jointure « left outer join »

MySQL, SGBD populaire, par l'exemple



- I. Installation et utilisation
- II. Introduction au SQL
- III. Trucs et astuces



Installation et utilisation

- Pour installer le serveur, copier simplement le répertoire fourni sur une clef USB ou un disque dur
- Lancer le serveur avec *MYSQL_allumer.bat*
- Interroger le serveur avec un client comme *MySQL Query Browser*, ou avec la ligne de commande
- Éteindre le serveur avec *MYSQL_eteindre.bat*

Introduction au SQL (1)

- ▶  ci
- ▶  mysql
- ▶  recherche
- ▼  simulation
 - ▶  activite_2005
 - ▶  dmi_chru_2005
 - ▶  lpp_2005
 - ▶  lpp_2006
 - ▶  lpp_radies_2006
 - ▶  lpp_total
 - ▶  molon_chru_2005
 - ▶  tarifs_2005
 - ▶  tarifs_2006
- ▶  testdb
- ▶  these

- Un SGBD peut utiliser en même temps plusieurs *databases*
- Chaque *database* peut contenir plusieurs *tables*
- Nb : la structure des tables ne contient pas les relations, elles sont explicitées dans chaque requête d'interrogation (contrairement à MS Access)

Introduction au SQL (2)

- Dans les exemples suivants, nous utiliserons la ligne de commande SQL.
- La ligne de commande permet aussi d'utiliser des fichiers texte :
 - En entrée : au lieu de saisir les commandes SQL, on peut les stocker dans un fichier de script
 - En sortie : au lieu de lire les sorties à l'écran, on peut les enregistrer dans un fichier texte

Introduction au SQL (3)

- Créer une *database* :
 - CREATE DATABASE essai ;
 - SHOW DATABASES ;
 - USE essai ;
- Création et remplissage d'une table de patients
 - Lire le fichier patients.sql puis l'exécuter :
 - SOURCE fichiers/patients.sql ;
 - Vérifier que ça a marché :
 - SHOW TABLES ;
 - DESC patients ;
 - SELECT * FROM patients ;
 - SELECT count(*) FROM patients ;

Introduction au SQL (4)

- Restreindre le nombre de colonnes :
 - Exemple :

```
SELECT nature, couleur
FROM véhicules ;
```
 - Exercice :
 - Afficher seulement le nom et le prénom
 - Que signifiait « * » auparavant ?
- Ajouter d'autres colonnes calculées
 - Ajouter YEAR(date_naiss) AS an_naiss

Introduction au SQL (5)

- Restreindre le nombre de lignes par une clause de sélection WHERE
 - Exemple :

```
SELECT nature, couleur
FROM véhicules
WHERE nature = "camion"
```
 - Exercice
 - Afficher le nom et le prénom des personnes nées avant 1950
 - Utiliser le mot clef *DISTINCT* pour limiter les réponses identiques

Introduction au SQL (6)

La philosophie du « group by » :

- « GROUP BY nom, prénom » agrège les lignes dont les champs *nom* et *prénom* sont identiques
- Dès lors on ne peut pas afficher les champs exclus de la clause *group by*
- En revanche on peut utiliser sur ces autres champs des fonctions applicables aux séries :
 - count(*), count(distinct ...)
 - sum(...), avg(...), var(...)
 - min(...), max(...)
 - group_concat(...)

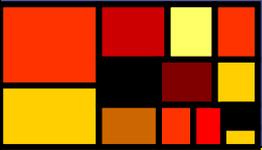
Introduction au SQL (7)

- Exercice avec group by :
 - Afficher pour chaque personne (triplet nom & prénom & date_naiss) :
 - Son nom
 - Son prénom
 - Sa date de naissance
 - Son hba1c moyenne
 - Le nombre de consultations (nombre de lignes)

Connexion à distance

- Pour tenter une connexion à distance :
 - L'ordinateur « serveur » doit connaître son adresse IP
 - Soit Démarrer > exécuter > ipconfig
 - Soit exécuter le fichier *adresse_ip.bat* fourni ici
 - L'ordinateur « client » peut se connecter en spécifiant cette adresse IP au lieu de « localhost », à l'aide d'un client comme MySQL Query Browser

Trucs et astuces



Écrire du code SQL depuis Excel :

- Exemple 1 : insérer des individus dans une table.
 - Exercice : écrire la formule sur le fichier excel.
- Exemple 2 : corriger des identités de patients faussement différentes.
 - Exercice : utiliser simplement la feuille

Annexe : Ressources fournies avec ce cours

- Le présent cours format PPT
- Divers fichiers XLS, TXT, SQL pour exemples
- Un serveur MySQL copiable
- Un interrogateur : MySQL Query Browser
- Un gestionnaire : MySQL Control Center
- Notice de MySQL 5.0 - lire *uniquement* :
 - Tutoriels d'introduction
 - Syntaxe des commandes SQL
 - Fonctions à utiliser dans les commandes SELECT et WHERE