Systèmes d'Information

Information + Réflexion + Maths + Automatisation

Fabien Coelho
Mines Paris – PSL

Id: intro.tex 4370 2022-11-29 13:04:38Z fabien

Objectif du cours

— comprendre (savoir)
— manipuler (savoir-faire)
— construire (faire)
— faire-savoir ? faire-faire ? un peu plus tard !

Évaluation de l'enseignement

1/2 TP et participation

1/2 examen écrit

Cours interactif !

Hiérarchie D-I-K-W *(Russel Ackoff 1988)*

Data données élémentaires *symbole*
1234567.89

Information connexions entre les données *qui/quoi/où/quand*
Your bank balance has jumped 8087% to $1234567.89

Knowledge connaissance, compréhension *comment/pourquoi*
Nobody owes me that much money

Wisdom synthèse *décision*
I'd better talk to the bank before I spend it,
because of what has happened to other people

Zeleny (1987) : know-nothing, know-what, know-how, know-why
donnée, méta-donnée, analyse, décision

Cycle de vie d'une donnée

(S)CRUD : *(Search) Create Read Update Delete*

C – création initiale, capture

SR – utilisation directe et indirecte

U – édition mise à jour continue ?

R – sauvegarde stockage, pertes ?

SR – archivage donnée fixée, contraintes légales, recherche…

D – destruction contraintes légales…

Agenda papier Moleskine semainier 18 mois

informations = données + liens sur le monde réel
espace, temps, matière, humanité…
— astronomie (année, solstices et équinoxe, lune, fuseaux)
— religion chrétienne (fêtes)
— histoire (fêtes nationales, 2*12h, 60mn)
— géographie (cartes, été de l'hémisphère nord, fêtes européennes)

sémantique très souple
— ambiguïtés, informations implicites
— cohérence non garantie

quantité de données conservées
— papier vs mémoire humaine…
— limité à 18 mois

pérennité du support
— papiers conservables quelques années

fiabilité pas de panne, ni effacement accidentel ?
— mais risques de pertes, destruction par l'eau et le feu…
— redondance sur 6 mois ?

exploitation selon les questions
— occupation le vendredi 29 février ?
— quand ai-je cours Systèmes d'Information ?

structure de l'agenda adaptées à certaines questions !
— agenda calendaire
— calendriers annuels, mensuels
— informations pratiques (vins, web, téléphone, conversions)

partage accès simultané difficile entre personnes
— photocopies : mises à jour multiples ? cohérence ?
— confidentialité : séparation personnel/professionnel ? privé/public ?

Définition d'une base de donnée

gros par rapport à la mémoire humaine ! petit pour l'informatique ?

ensemble notion mathématique. . . formalisme ?

persistant dans le temps, pannes

de données *i.e.* d'informations, typées ? multimédia ?

structurées liées, organisées

et cohérentes contraintes d'intégrité déclarées et forcées

exploitable interrogation, modifications, évolution

simultanément parallélisme des accès, partage

Solution informatique = information + automatique

— stockage des information dans des fichiers (disque, bande)

— traitement automatique : répondre à une question !

Solution fichier + programme

format spécifique. . .

— compatibilité entre applications ? standards ?

— accès et mise à jour simultanée ? distribution ?

— maintien de la cohérence ?

manipulations fastidieuses. . .

— une question implique un programme !

— modification implique réécriture complète du fichier ?

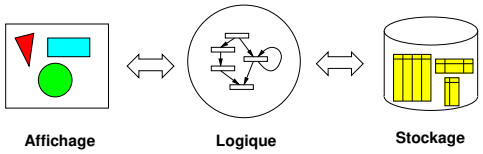
Architecture *3-tiers* d'une application

affichage navigateur, application graphique, mobile. . .

logique programmation *java/servlet python/flask ruby/rails PHP*. . .

stockage persistant avec base de donnée

liens entre parties : protocoles réseaux `http` et librairies



Fonctionnalités d'une base de donnée

intégrité physique en cas de panne — WAL
reprise sur panne, distribution, duplication
problèmatique système, bas niveau

intégrité sémantique maintien de la cohérence logique — RI
vérification des contraintes applicatives
approche déclarative ? facilité d'évolution ?

simplicité et optimisation des accès aux données — SQL
interopérabilité, interfaces standards
approche déclarative ?

partage et confidentialité des données — MVCC
transactions, droits. . .

Différents modèles. . .

relationnel le grand gagnant !
basé sur un modèle mathématique clair

hiérarchique contenu organisé en arbre
utilisé par LDAP, MS AD (répertoires partagés)

réseau contenu organisé en graphe
document XML avec références ?

orienté-objet le retour du modèle réseau ?
relationnel objet ? !

misc big data map-reduce. . .
clef-valeur, document, colonnes larges, graph. . .

Données non-relationnelles (ou mauvais relationnel)

corruption — *stockage manuel (et bogué)*

incohérence — *vs intégrité référentielle*
ou cohérence non fiable. . . *laissé à l'application*

duplication — *vs modèle normalisé*
laquelle est la bonne ? deux informations identiques ou une seule ?

modèle instable — *lié à l'application*

Mais parfois utile !

Usage de bases de données

transactionnel données partagées, acteurs multiples — *OLTP*
— reflet de la situation réelle (stocks, comptabilité, commandes)
— mise à jour en **continue** : cohérence essentielle

décisionnel cubes d'analyses stats multicritères — *OLAP*
— informations aggrégées (nombre de ventes / magasin / semaine)
géographie, temps, produits, segmentations clients. . .
— mise à jour périodique (semaine, mois, année)

individuel application **bureautique** non intégrée
— informations partielles extraites d'une base principale
— accessible à un utilisateur, partage difficile

embarqué (caché) dans une application
— structuration du stockage des données

Données vs Applications ?

les applications passent, les données restent. . .

IBM 7094, 1962

— **données** spécifiques : intégrité, précision, sécurité. . .

— **applications** : locales, achats, remplacements. . .

Jean-Paul Figer (ancien CTO CapGemini)

Une entreprise ne devrait s'occuper QUE de ses données...

— **modélisation** du (cœur de) métier
— **stockage**, protection
— **interfaces** : import, export, accès...

Marché des bases de données

— b$78.7 (BRC, 2020), largement relationnel
— Commercial : Oracle, Microsoft, IBM, SAP, Teradata
croissance depuis 2012 : +10% par an
— Libre : MySQL, Postgres, SQLite, MongoDB, Hadoop...
croissance : +30% par an...
nouveaux projets vs legacy...

ORACLE Microsoft IBM SYBASE TERADATA

MySQL SQLite HSQL database engine

Implémentations relationel libres...

MySQL/MariaDB la plus populaire

Postgres plus sérieux...

SQLite embarqué, la plus utilisée dans le monde ?

Firefox Chrome Skype IOS MacAfee Acrobat Symbian Solaris MacOS...

Ingres ex-commerciale, libre depuis 2004

HSQLDB en java

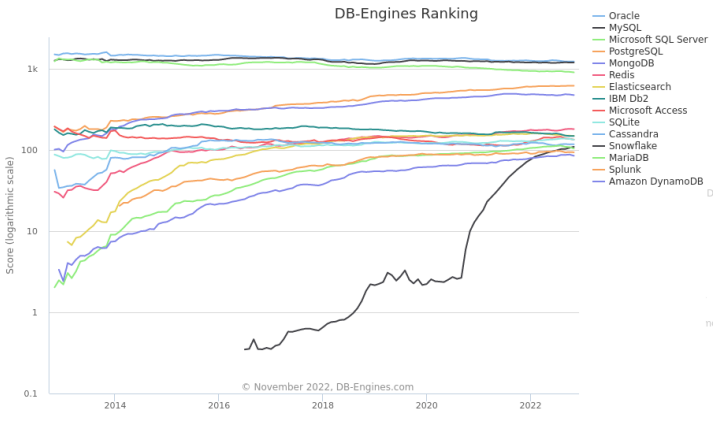
Firebird basé sur Interbase/Borland

...

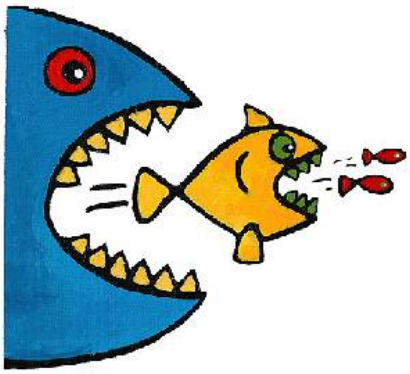
Popularité des BD – `db-engines.com`, novembre 2022

Méthodologie : occurrences web, search, jobs, articles, CV...

#	Origine	Base de donnée	Type	Indice
1	Oracle	DB	relation	1241
2	Oracle/*Libre*	MySQL	relation	1205
3	Microsoft	SQL Server	relation	912
4	*Libre*	Postgres	relation	623
5	*Libre*	MongoDB	document	477
6	*Libre*	Redis	key-value	182
7	*Libre*	Elasticsearch	search	150
8	IBM	DB2	relation	149
9	Microsoft	Access	relation	135
10	*Libre*	SQLite	relation.	134



Oracle vs MySQL



Timeline

2005-10 rachat *Innobase* par *Oracle* InnoDB
2006-02 rachat *Sleepycat Software* par *Oracle* BDB
2008-01 rachat *MySQL AB* par *Sun Microsystems*
2008-12 fork *Percona Server*
2009-01 fork *MySQL/MariaDB* vs projet de rachat par Oracle
2009-04 rachat *Sun Microsystems* par *Oracle* MySQL
autorisé par la Commission Européenne le 21 janvier 2010,
en citant l'alternative offerte par Postgres...
2013-04 licence doc MySQL changée
2013-06 license BDB changée vers AGPL (toujours libre)
2014-03 fork WebScaleSQL (Fb, Google, LinkedIn, Twitter)

List of Slides

1 Systèmes d'Information
1 Information + Réflexion + Maths + Automatisation
2 Objectif du cours
2 Évaluation de l'enseignement
2 Cours interactif !
4 Hiérarchie D-I-K-W *(Russel Ackoff 1988)*
5 Cycle de vie d'une donnée
6 Agenda papier Moleskine semainier 18 mois
9 Définition d'une base de donnée
10 Solution informatique = information + automatique
10 Solution fichier + programme

11 Architecture *3-tiers* d'une application

12 Fonctionnalités d'une base de donnée

13 Différents modèles…

14 Données non-relationnelles (ou mauvais relationnel)

15 Usage de bases de données

16 Données vs Applications ?

17 Jean-Paul Figer (ancien CTO CapGemini)

18 Marché des bases de données

19 Implémentations relationel libres…

20 Popularité des BD – `db-engines.com`, novembre 2022

22 Oracle vs MySQL

23 Timeline