Systèmes d'information

Fabien Coelho, Claire Medrala

Mines Paris – PSL, MESR

Janvier 2025

Objectifs

- comprendre *savoir*
- manipuler *savoir-faire*
- construire *faire*

Évaluation

1/2 TP personnels

1/2 Examen

Participation

Hiérarchie DIKW — *Russel Ackoff 1988*

Data	données élémentaires	*symbole*
Information	connexions entre les données	*quoi/qui/où/quand*
Knowledge	connaissance, compréhension	*pourquoi/comment*
Wisdom	synthèse	*décision*

Promenade

D

I

K

W

Concept d'information DIKW

SI
FC/CM
Intro
Information
CRUD
Papier
Def
Archi
Fonctions
Modèles
Usages
Data/App
Market

Hiérarchie DIKW — *Russel Ackoff 1988*

Data	données élémentaires	*symbole*
Information	connexions entre les données	*quoi/qui/où/quand*
Knowledge	connaissance, compréhension	*pourquoi/comment*
Wisdom	synthèse	*décision*

Promenade

D 130

I

K

W

Concept d'information — DIKW

SI
FC/CM
Intro
Information
CRUD
Papier
Def
Archi
Fonctions
Modèles
Usages
Data/App
Market

Hiérarchie DIKW		*Russel Ackoff 1988*
Data	données élémentaires	*symbole*
Information	connexions entre les données	*quoi/qui/où/quand*
Knowledge	connaissance, compréhension	*pourquoi/comment*
Wisdom	synthèse	*décision*

Promenade

- D 130
- I vitesse du véhicule de 130 km/h
- K
- W

Concept d'information — DIKW

SI
FC/CM
Intro
Information
CRUD
Papier
Def
Archi
Fonctions
Modèles
Usages
Data/App
Market

Hiérarchie DIKW		*Russel Ackoff 1988*
Data	données élémentaires	*symbole*
Information	connexions entre les données	*quoi/qui/où/quand*
Knowledge	connaissance, compréhension	*pourquoi/comment*
Wisdom	synthèse	*décision*

Promenade

- D 130
- I vitesse du véhicule de 130 km/h
- K route de montagne sous la pluie avant un virage...
- W J'ÉCRASE LA PÉDALE DE FREIN !

Concept d'information — DIKW

SI
FC/CM
Intro
Information
CRUD
Papier
Def
Archi
Fonctions
Modèles
Usages
Data/App
Market

Hiérarchie DIKW		*Russel Ackoff 1988*
Data	données élémentaires	*symbole*
Information	connexions entre les données	*quoi/qui/où/quand*
Knowledge	connaissance, compréhension	*pourquoi/comment*
Wisdom	synthèse	*décision*

Promenade

- D 130
- I vitesse du véhicule de 130 km/h
- K route de montagne sous la pluie avant un virage...
- W

Cycle de vie d'une donnée — CRUD(S)

SI
FC/CM
Intro
Information
CRUD
Papier
Def
Archi
Fonctions
Modèles
Usages
Data/App
Market

Create Read Update Delete (Search)

- création – C initiale, capture, manuel...
- utilisation – R directe et indirecte, sauvegarde, archivage
- édition – U mise à jour en continue ?
- destruction – D contraintes légales...

Agenda papier — Moleskine semainier 18 mois

informations = données + liens sur le monde réel

géographie	cartes, été de l'hémisphère nord, fêtes européennes
histoire	fêtes nationales, 2*12h, 60mn
astronomie	année, solstices et équinoxe, lune, fuseaux
religion	chrétienne, fêtes

caractéristiques

sémantique	ambiguïtés, informations implicites, cohérence non garantie
quantité	faible, limitée à 18 mois, recopies ?
pérennité	papier, +100 ans ?
fiabilité	destruction, perte ?

Exploitation d'un agenda

Structure calendaire — adaptée à certaines questions !

- occupation le 29 juillet ? — Ok!
- quand ai-je cours Systèmes d'Information ? — KO!

partage — accès simultané

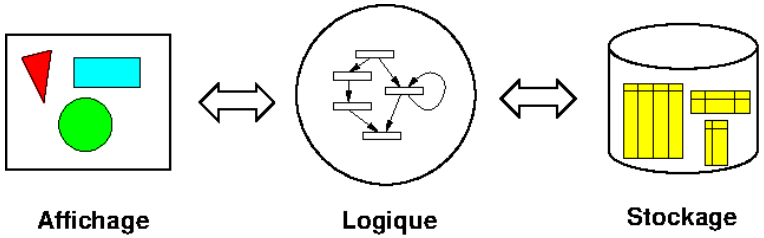
- copies : mises à jour multiples ? cohérence ?
- confidentialité : séparation perso/pro ?

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 ?
structurées	liées, organisées
et cohérentes	contraintes d'intégrité déclarées et forcées
exploitable	interrogation, modifications, évolution
simultanément	parallélisme des accès, partage

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

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...

Modèles...

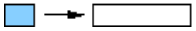
relationnel le grand gagnant !
basé sur un modèle mathématique clair

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

réseau contenu organisé en graphe
document XML avec références ?

objet le retour du modèle réseau ?
relationnel objet ?!

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



Deux types d'usages — *transactions vs décision*

Transactionnel — OLTP

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

Décisionnel — OLAP

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

Les applications passent, les données restent...

IBM 7094, 1962

- **données** spécifiques : intégrité, précision, sécurité...
- **applications** : locales, achats, remplacements...
- définir des **interfaces** !

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...

- croissance : +30% par an...
- nouveaux projets vs legacy...

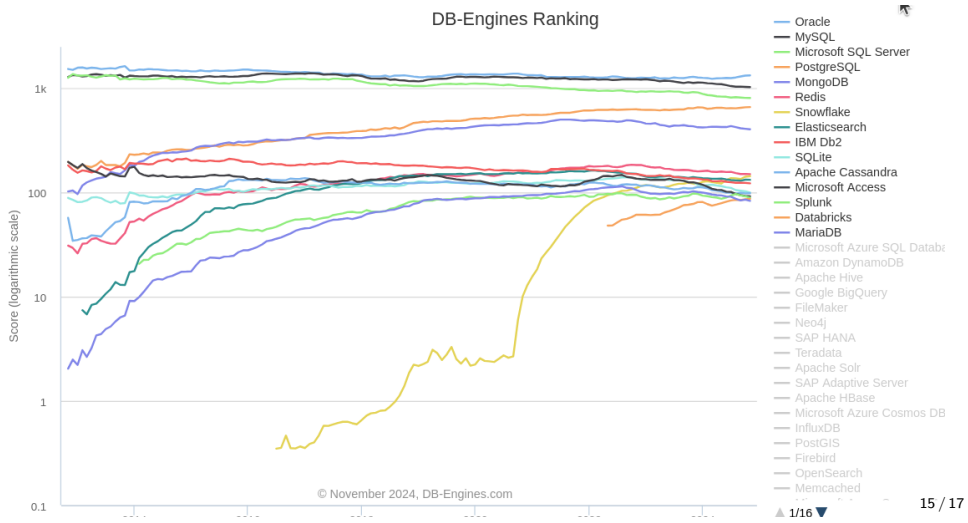
Popularité des BD `https://db-engines.com`, 11/2024

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

#	Origine	Base de donnée	Type	Indice
1	Oracle	DB	relation	1317
2	Oracle/*Libre*	MySQL	relation	1018
3	Microsoft	SQL Server	relation	800
4	*Libre*	Postgres	relation	654
5	*Libre*	MongoDB	document	401
6	*Libre*	Redis	key-value	149
7	Snowflake Co.	Snowflake	relation	143
8	*Libre*	ElasticSearch	search	132
9	IBM	DB2	relation	122
10	*Libre*	SQLite	relation	99

PostgreSQL is the DBMS of the year for 2017, 2018, 2020, 2023

Trends `https://db-engines.com`, 11/2024



Popularité écosystème Python `https://pypistats.org` 11/2024

Postgres	70.2 M/mo
pg8000	33.2 M/mo
psycopg2	17.2 M/mo
psycopg	10.3 M/mo
asyncpg	9.3 M/mo
pygresql	0.2 M/mo

MySQL	83.1 M/mo
pymysql	51.7 M/mo
mysql-conn...	31.4 M/mo

Enquête StackOverflow 2024 (65,000)

Popularité : admiré et désiré

https://survey.stackoverflow.co/2024/

47.1%	Postgres	21.1%	Redis
26.8%	SQLite	15.4%	SQL Server
24.6%	MySQL	12.2%	Elasticsearch
21.9%	MongoDB	11.6%	MariaDB

Par catégories

pro	*Postgres, MySQL, SQLite, SQL Server, MongoDB, Redis*
apprentis	*MySQL, SQLite, Postgres, MongoDB, MariaDB, SQL Server*
autres	*MySQL, Postgres, SQLite, SQL Server, MongoDB, MariaDB*