Introduction aux bases de données

- Se familiariser avec les bases de données (rôle, fonctionnement)
- Découvrir les origines et comprendre les concepts clefs du modèle relationnel
- Découvrir le paysage actuel des implémentations du standard SQL

Qu'est-ce qu'une base de données ?

Sert à **stocker**, **retrouver** des données plus ou moins **structurées** de manière **fiable**. Permet de les interroger et d'exécuter des traitements dessus. Fournit également un environnement multi-utilisateur (droits, accès)

Une base de données est le **cœur** du Système d'Information (SI) :

- **générer** de l'information
- **mémoriser** l'information
- **communiquer** et **diffuser** l'information
- **exécuter** des traitements

Propriétés désirables

- adaptable (*scalability*)
- maintenable (*maintainability*)
- sécurisée (*security*)
  - fiable (*accuracy*)
  - disponible (*availability*)
  - secret et confidentialité (*secrecy*)

Concrètement

Une base de données, ou Système de Gestion de Base de Données (SGBD) est composée d'un ensemble de *fichiers* et de *programmes* manipulant ces fichiers, répartis sur une ou plusieurs machines. Les SGBD sont souvent implémentés suivant l'architecture client/serveur.

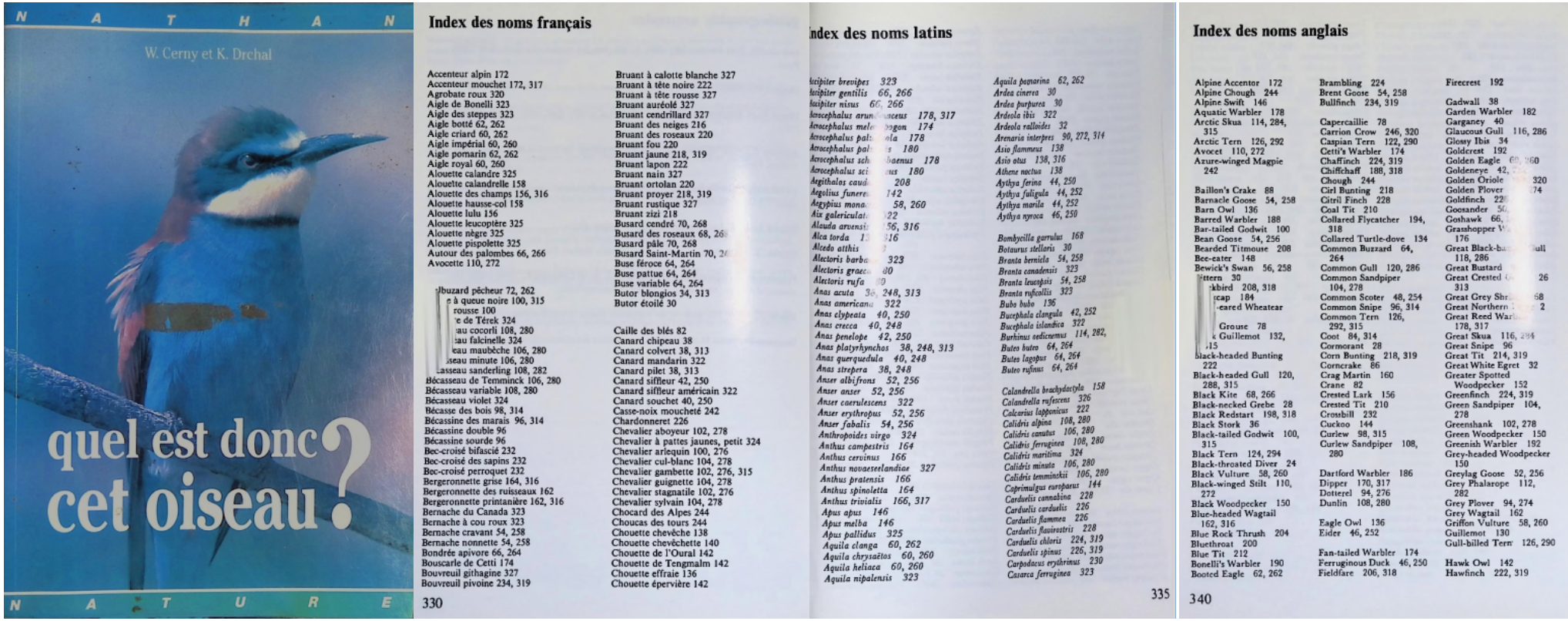
Atelier: *plain text database*

Par exemple, le fichier `/etc/passwd` sur les systèmes Unix est *une base de données*. Il est modifiable via le programme `passwd`.

> Voir le document td.pdf (fourni avec le module) sur l'atelier plain text database, où nous construisons et analysons les intérêts et limites d'une base de données très simple.

Recherche par index vs Recherche séquentielle

Exemple: un livre sur les oiseaux



Index des noms français

Accenteur alpin 172
Accenteur mouchet 172, 317
Agrobate roux 320
Aigle de Bonelli 323
Aigle des steppes 323
Aigle botté 62, 262
Aigle criard 60, 262
Aigle impérial 60, 260
Aigle pomarin 62, 262
Aigle royal 60, 260
Alouette calandre 325
Alouette calandrelle 158
Alouette des champs 156, 316
Alouette hausse-col 158
Alouette lulu 156
Alouette leucoptère 325
Alouette nègre 325
Alouette pispolette 325
Autour des palombes 66, 266
Avocette 110, 272

[illegible]buzard pêcheur 72, 262
[illegible] à queue noire 100, 315
[illegible] rousse 100
[illegible]e de Térek 324
[illegible]au cocorli 108, 280
[illegible]au falcinelle 324
[illegible]eau maubèche 106, 280
[illegible]seau minute 106, 280
[illegible]asseau sanderling 108, 282
Bécasseau de Temminck 106, 280
Bécasseau variable 108, 280
Bécasseau violet 324
Bécasse des bois 98, 314
Bécassine des marais 96, 314
Bécassine double 96
Bécassine sourde 96
Bec-croisé bifascié 232
Bec-croisé des sapins 232
Bec-croisé perroquet 232
Bergeronnette grise 164, 316
Bergeronnette des ruisseaux 162
Bergeronnette printanière 162, 316
Bernache du Canada 323
Bernache à cou roux 323
Bernache cravant 54, 258
Bernache nonnette 54, 258
Bondrée apivore 66, 264
Bouscarle de Cetti 174
Bouvreuil githagine 327
Bouvreuil pivoine 234, 319
Bruant à calotte blanche 327
Bruant à tête noire 222
Bruant à tête rousse 327
Bruant auréolé 327
Bruant cendrillard 327
Bruant des neiges 216
Bruant des roseaux 220
Bruant fou 220
Bruant jaune 218, 319
Bruant lapon 222
Bruant nain 327
Bruant ortolan 220
Bruant proyer 218, 319
Bruant rustique 327
Bruant zizi 218
Busard cendré 70, 268
Busard des roseaux 68, 26[illegible]
Busard pâle 70, 268
Busard Saint-Martin 70, 2[illegible]
Buse féroce 64, 264
Buse pattue 64, 264
Buse variable 64, 264
Butor blongios 34, 313
Butor étoilé 30

Caille des blés 82
Canard chipeau 38
Canard colvert 38, 313
Canard mandarin 322
Canard pilet 38, 313
Canard siffleur 42, 250
Canard siffleur américain 322
Canard souchet 40, 250
Casse-noix moucheté 242
Chardonneret 226
Chevalier aboyeur 102, 278
Chevalier à pattes jaunes, petit 324
Chevalier arlequin 100, 276
Chevalier cul-blanc 104, 278
Chevalier gambette 102, 276, 315
Chevalier guignette 104, 278
Chevalier stagnatile 102, 276
Chevalier sylvain 104, 278
Chocard des Alpes 244
Choucas des tours 244
Chouette chevêche 138
Chouette chevêchette 140
Chouette de l'Oural 142
Chouette de Tengmalm 142
Chouette effraie 136
Chouette épervière 142

330

[illegible]ndex des noms latins

[illegible]ccipiter brevipes 323
[illegible]ccipiter gentilis 66, 266
[illegible]ccipiter nisus 66, 266
Acrocephalus arun[illegible]aceus 178, 317
Acrocephalus mele[illegible]ogon 174
Acrocephalus pal[illegible]ola 178
Acrocephalus pal[illegible]s 180
Acrocephalus sch[illegible]baenus 178
Acrocephalus sci[illegible]us 180
Aegithalos caud[illegible] 208
Aegolius funere[illegible] 142
Aegypius mona[illegible] 58, 260
Aix galericulat[illegible]22
Alauda arvensi[illegible]56, 316
Alca torda 1[illegible]16
Alcedo atthis
Alectoris barba[illegible] 323
Alectoris graec[illegible]30
Alectoris rufa
Anas acuta 3[illegible] 248, 313
Anas americana 322
Anas clypeata 40, 250
Anas crecca 40, 248
Anas penelope 42, 250
Anas platyrhynchos 38, 248, 313
Anas querquedula 40, 248
Anas strepera 38, 248
Anser albifrons 52, 256
Anser anser 52, 256
Anser coerulescens 322
Anser erythropus 52, 256
Anser fabalis 54, 256
Anthropoides virgo 324
Anthus campestris 164
Anthus cervinus 166
Anthus novaeseelandiae 327
Anthus pratensis 166
Anthus spinoletta 164
Anthus trivialis 166, 317
Apus apus 146
Apus melba 146
Apus pallidus 325
Aquila clanga 60, 262
Aquila chrysaëtos 60, 260
Aquila heliaca 60, 260
Aquila nipalensis 323
Aquila pomarina 62, 262
Ardea cinerea 30
Ardea purpurea 30
Ardeola ibis 322
Ardeola ralloides 32
Arenaria interpres 90, 272, 314
Asio flammeus 138
Asio otus 138, 316
Athene noctua 138
Aythya ferina 44, 250
Aythya fuligula 44, 252
Aythya marila 44, 252
Aythya nyroca 46, 250

Bombycilla garrulus 168
Botaurus stellaris 30
Branta bernicla 54, 258
Branta canadensis 323
Branta leucopsis 54, 258
Branta ruficollis 323
Bubo bubo 196
Bucephala clangula 42, 252
Bucephala islandica 322
Burhinus oedicnemus 114, 282,
Buteo buteo 64, 264
Buteo lagopus 64, 264
Buteo rufinus 64, 264

Calandrella brachydactyla 158
Calandrella rufescens 326
Calcarius lapponicus 222
Calidris alpina 108, 280
Calidris canutus 106, 280
Calidris ferruginea 108, 280
Calidris maritima 324
Calidris minuta 106, 280
Calidris temminckii 106, 280
Caprimulgus europaeus 144
Carduelis cannabina 228
Carduelis carduelis 226
Carduelis flammea 226
Carduelis flavirostris 228
Carduelis chloris 224, 319
Carduelis spinus 226, 319
Carpodacus erythrinus 230
Casarca ferruginea 323

335

Index des noms anglais

Alpine Accentor 172
Alpine Chough 244
Alpine Swift 146
Aquatic Warbler 178
Arctic Skua 114, 284, 315
Arctic Tern 126, 292
Avocet 110, 272
Azure-winged Magpie 242

Baillon's Crake 88
Barnacle Goose 54, 258
Barn Owl 136
Barred Warbler 188
Bar-tailed Godwit 100
Bean Goose 54, 256
Bearded Titmouse 208
Bee-eater 148
Bewick's Swan 56, 258
[illegible]ittern 30
[illegible]kbird 208, 318
[illegible]tcap 184
[illegible]-eared Wheatear
[illegible] Grouse 78
[illegible]k Guillemot 132, [illegible]15
Black-headed Bunting 222
Black-headed Gull 120, 288, 315
Black Kite 68, 266
Black-necked Grebe 28
Black Redstart 198, 318
Black Stork 36
Black-tailed Godwit 100, 315
Black Tern 124, 294
Black-throated Diver 24
Black Vulture 58, 260
Black-winged Stilt 110, 272
Black Woodpecker 150
Blue-headed Wagtail 162, 316
Blue Rock Thrush 204
Bluethroat 200
Blue Tit 212
Bonelli's Warbler 190
Booted Eagle 62, 262
Brambling 224
Brent Goose 54, 258
Bullfinch 234, 319

Capercaillie 78
Carrion Crow 246, 320
Caspian Tern 122, 290
Cetti's Warbler 174
Chaffinch 224, 319
Chiffchaff 188, 318
Chough 244
Cirl Bunting 218
Citril Finch 228
Coal Tit 210
Collared Flycatcher 194, 318
Collared Turtle-dove 134
Common Buzzard 64, 264
Common Gull 120, 286
Common Sandpiper 104, 278
Common Scoter 48, 254
Common Snipe 96, 314
Common Tern 126, 292, 315
Coot 84, 314
Cormorant 28
Corn Bunting 218, 319
Corncrake 86
Crag Martin 160
Crane 82
Crested Lark 156
Crested Tit 210
Crossbill 232
Cuckoo 144
Curlew 98, 315
Curlew Sandpiper 108, 280

Dartford Warbler 186
Dipper 170, 317
Dotterel 94, 276
Dunlin 108, 280

Eagle Owl 136
Eider 46, 252

Fan-tailed Warbler 174
Ferruginous Duck 46, 250
Fieldfare 206, 318
Firecrest 192

Gadwall 38
Garden Warbler 182
Garganey 40
Glaucous Gull 116, 286
Glossy Ibis 34
Goldcrest 192
Golden Eagle [illegible]60
Goldeneye 42, [illegible]
Golden Oriole [illegible]320
Golden Plover [illegible]74
Goldfinch 22[illegible]
Goosander 5[illegible]
Goshawk 66, [illegible]
Grasshopper W[illegible] 176
Great Black-ba[illegible]ull 118, 286
Great Bustard
Great Crested [illegible] 26 313
Great Grey Shr[illegible]68
Great Northern [illegible]2
Great Reed War[illegible] 178, 317
Great Skua 116, [illegible]
Great Snipe 96
Great Tit 214, 319
Great White Egret 32
Greater Spotted Woodpecker 152
Greenfinch 224, 319
Green Sandpiper 104, 278
Greenshank 102, 278
Green Woodpecker 150
Greenish Warbler 192
Grey-headed Woodpecker 150
Greylag Goose 52, 256
Grey Phalarope 112, 282
Grey Plover 94, 274
Grey Wagtail 162
Griffon Vulture 58, 260
Guillemot 130
Gull-billed Tern 126, 290

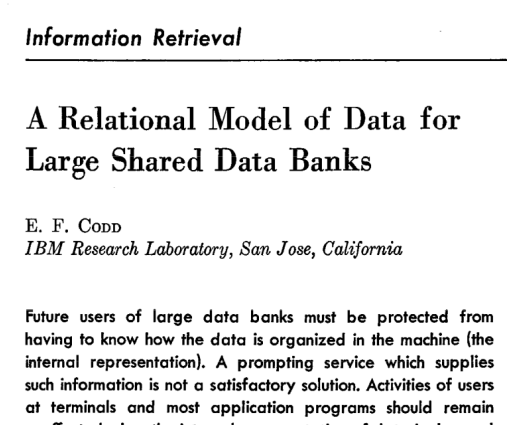
Hawk Owl 142
Hawfinch 222, 319

340

Pour résumer

- un SGBD sert à stocker, retrouver, interroger et traiter des données. C'est le cœur du SI.
- un SBGD est constitué de *programmes* qui manipulent des *fichiers*. Il est souvent (mais pas toujours) implémenté suivant l'architecture client/serveur (via le protocole TCP/IP)
- un SGBD a la responsabilité de beaucoup de choses différentes: *techniques* (optimisation, gestion de la mémoire, des pannes, des accès, etc.) et *conceptuelles* (modèle de données). Chaque SGBD est *optimisé* en interne.
- il existe toujours un *compromis* entre lecture et écriture
- un index permet de parcourir plus efficacement des données qu'un accès séquentiel (comme dans un livre)

Le modèle relationnel et le standard SQL



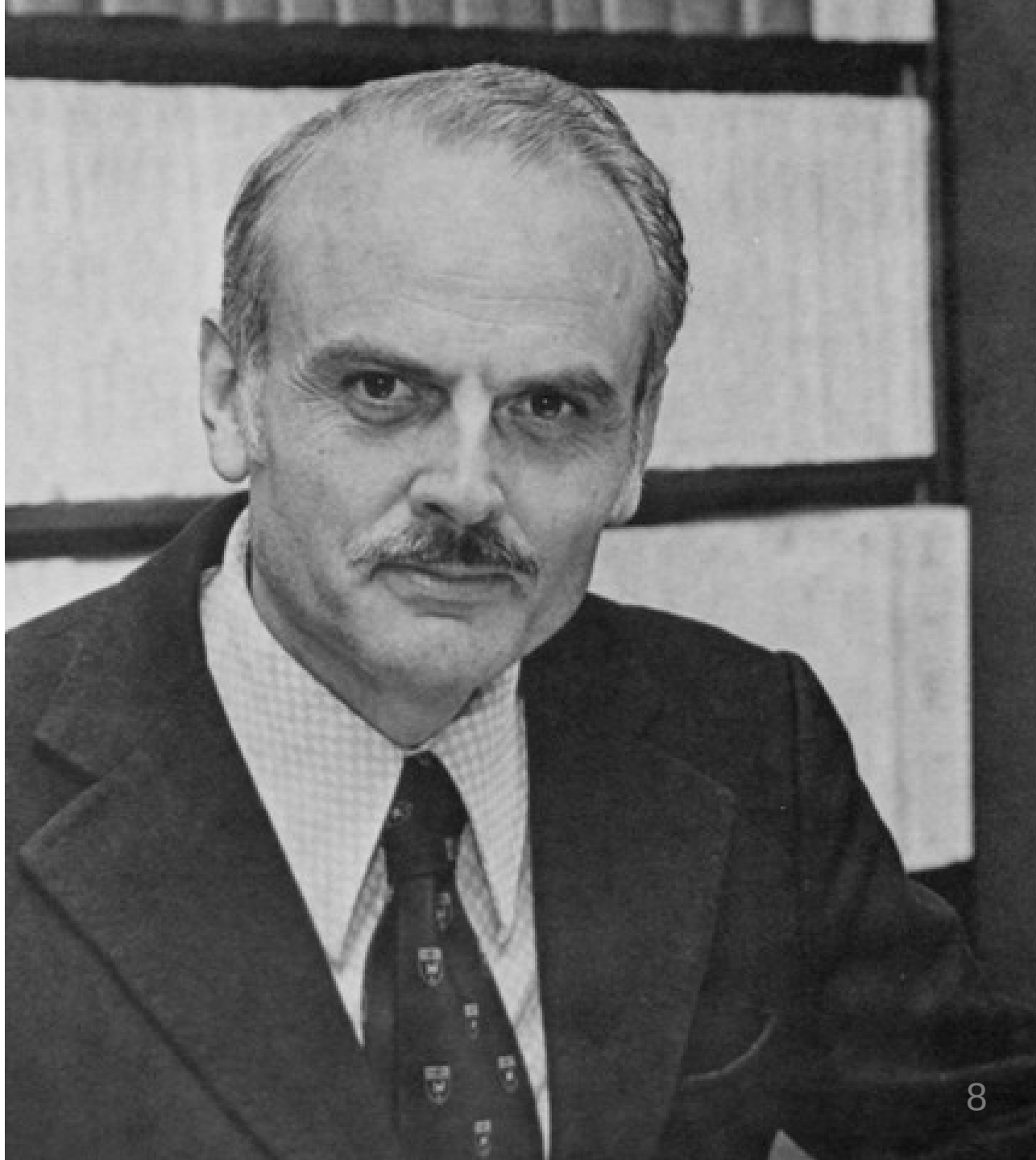
Information Retrieval

A Relational Model of Data for Large Shared Data Banks

E. F. Codd
IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain

Le modèle relationnel est basé sur les travaux d'Edgard Codd (1970, Prix Turing en 1981)



Les principes du modèle relationnel

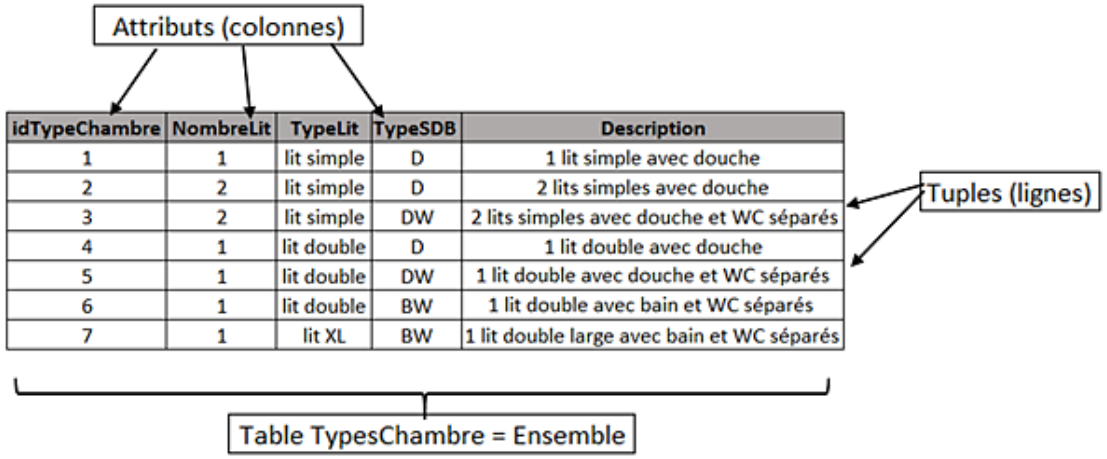
Organiser les données en *relations* (*ensemble* en mathématiques, *table* en SQL).

Une relation est définie par un ou plusieurs attributs valués et par une *clef*.

Chaque relation est une collection non ordonnée de *tuples* (*lignes* en SQL).

Par exemple, `(1, 1, lit simple, D, 1 lit simple avec une douche)` est un tuple de la relation `TypesChambre`.

L'attribut `idTypeChambre` sert de *clef*.



idTypeChambre	NombreLit	TypeLit	TypeSDB	Description
1	1	lit simple	D	1 lit simple avec douche
2	2	lit simple	D	2 lits simples avec douche
3	2	lit simple	DW	2 lits simples avec douche et WC séparés
4	1	lit double	D	1 lit double avec douche
5	1	lit double	DW	1 lit double avec douche et WC séparés
6	1	lit double	BW	1 lit double avec bain et WC séparés
7	1	lit XL	BW	1 lit double large avec bain et WC séparés

Le standard SQL

SQL (*Structured Query Language*) est un standard.

> *Par abus de langage*, désigne un langage *déclaratif* pour manipuler des bases de données relationnelles.

Ce standard définit une interface pour:

- *définir* les données (le schéma: la représentation et le type des données)
- *manipuler* les données (mettre à jour, insérer, supprimer, etc.)
- réaliser des *transactions*
- *contrôler* l'accès aux données

Le standard SQL est éprouvé et en constante évolution

Implémenté en 1974 chez IBM, puis première implémentation commercialisée en 1979 par Oracle.

Normalisé ANSI (*American National Standards Institute*) en 1986.

Norme internationale ISO en 1987.

Un standard en perpétuelle évolution.

> Version actuelle du standard: 2016. **Version 2023 en cours de validation**.

The Mother of all Query Languages: SQL in Modern Times, par Markus Winand

Conférence de Markus Winand, 2009 (Data natives)

Les origines du modèle

- un SGBD qui masque les détails d'implémentation
- **traitement des données des activités économiques** exécutées sur un ordinateur central (ventes, transactions et fonds bancaires, réservation de billets, inventaires, et.)
- **traitements par lots** (*batch processing*): édition de factures, de fiches de paie, de rapports

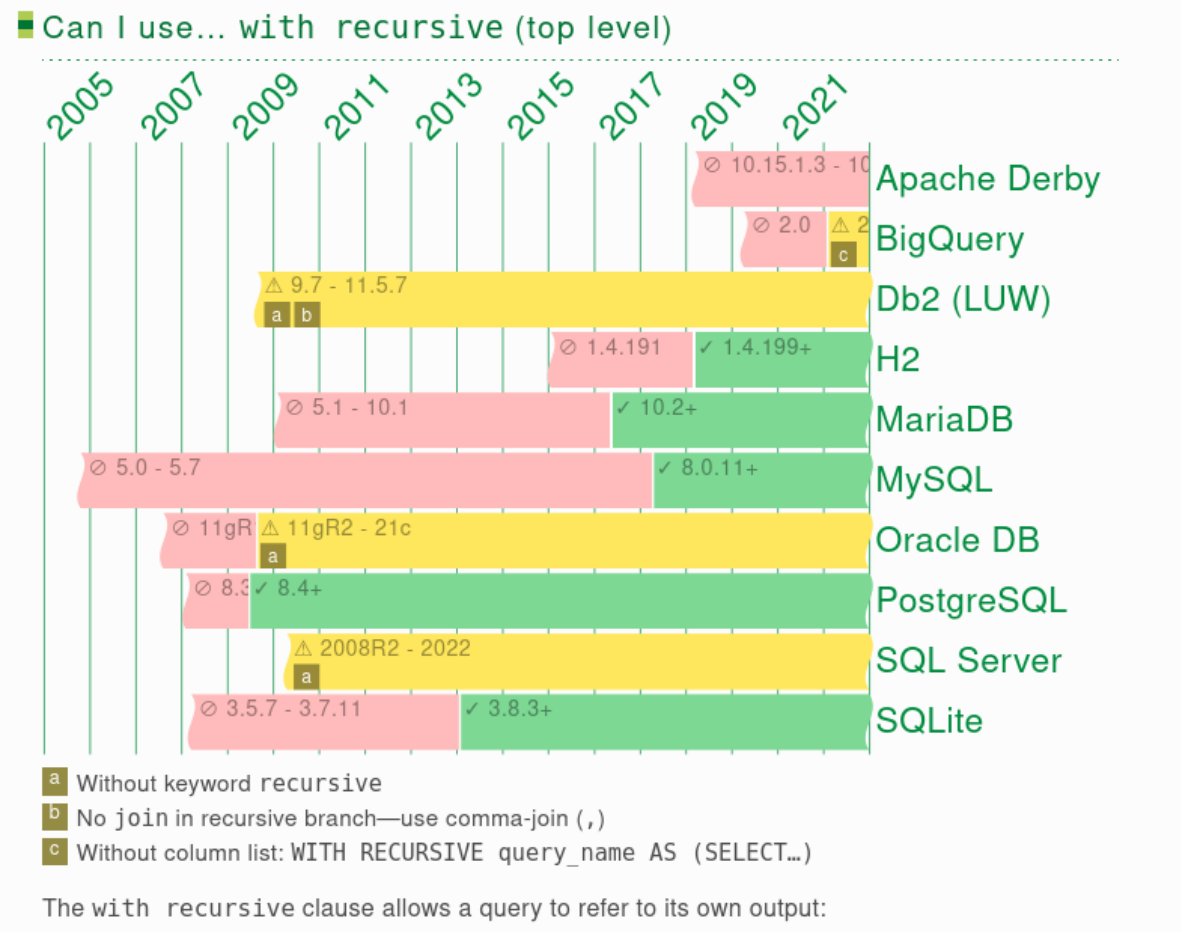
> Légende: l'ordinateur Mainframe d'IBM (1964)

Le standard SQL: au-delà du modèle relationnel

Depuis la spécification de 1999:

- nested tables (modèles/tables récursives)
- tableaux dans un attribut
- XML et JSON comme valeurs *recherchables* (comme les bases orientées Document)
- types composites (objets)
- etc.

> Légende: *with recursive* clause permet d'écrire une requête récursive (standard de 1999)



Le standard SQL: un *langage* déclaratif

SQL **ne dit pas *comment* le faire, il permet de *dire ce qu'on veut faire***.

```
#Je veux récupérer tous les animaux de la famille des requins
SELECT * FROM animals WHERE family='sharks';
```

Il existe d'autres modèles de données

- Le modèle de base de données réseau
- Le modèle hiérarchique
- Les bases de données XML
- etc.

Le modèle relationnel s'est parfaitement adapté au web:

- publication en ligne, discussion (BBC News, Wikipédia, Youtube, etc.)
- réseaux sociaux (Facebook, Twitter, LinkedIn, etc.)
- ecommerce (Booking.com, Airbnbn, eBay)
- jeux vidéo (World of Warcraft (Oracle), The Witcher 3, moteur Unity (SQLite), etc.)
- SaaS (GitHub, Netflix)
- etc.

La dernière remise en cause: NoSQL

Hashtag sur Twitter #NoSQL depuis 2010. Ne parle pas d'une technologie en particulier.

Volonté de remettre en cause le modèle relationnel:

- grands volumes de données (greater scalability)
- demande de SGBD libre et open source VS SGBRD commerciaux (ex: Oracle)
- frustration face à la restriction imposée par le schéma de données du modèle relationnel (schema *on-write*)

Autres modèles émergeant:

- Document Databases (ex: MongoDB, CouchDB): peu de relation entre documents
- Graph Databases (ex: Oracle, Neo4j): grande connectivité des données

> A chaque modèle son cas d'utilisation. Choisir le modèle *adapté à son besoin*.

Les différentes implémentations du standard SQL (SGBDR)

SQL est un standard. Il en existe de *nombreuses implémentations* (par différents éditeurs):

- Oracle: Oracle Database, MySQL
- Microsoft: Microsoft SQL Server
- Groupe de développement open source: PostgreSQL
- MariaDB Foundation: MariaDB
- SQLite (base de données intégrée, < 600Ko), souvent embarquée utilisé dans les navigateurs (Firefox), produits Adobe
- etc.

Chaque vendor/implémentation a :

- ses particularités d'implémentations (types, schema, etc.), son histoire et implémente **de manière plus ou moins avancée le standard**
- ses conventions de nommage
- son langage procédural (PL/SQL pour Oracle, etc.)

Le cas MySQL: la petite histoire

- **1995**: MySQL créée et distribuée par Micheal Widenius (gauche) et David Axmark (droite) au sein de la société MySQL AB (Suède).
- **2000**: MySQL passe sous licence GPL

Le cas MySQL: premier rachat

- **2008**: MySQL AB rachetée par Sun Microsystems pour la modique somme de...
- 1 milliard de dollars !
- Michal Widenius quitte Sun pour créer le SGBDR MariaDB (fork open source de MySQL)

Le cas MySQL: deuxième rachat

- **2009**: Sun Microsystems (et donc MySQL) est racheté par Oracle Corporation (218 milliards de dollars de capital) pour la modique somme de...
- 7,4 milliards de dollars !
- Oracle possède Oracle Database et MySQL !
- Le projet MySQL est distribué sous double licence: libre et propriétaire

Pourquoi (allons-nous encore) utiliser le modèle relationnel ?

> If you use SQL for CRUD operations only, you are doing it wrong (Winand)

- standard SQL **avancé**, **mature** et **continuellement mis à jour**
- le modèle relationnel est basé sur une **algèbre solide et éprouvée** (bon modèle)
- implémentations (SGBDR) développées et améliorées depuis plus de 40 ans (R&D): **optimisées**, performantes, *scalable*
- les SGBDR sont ACID (on y reviendra)
  - Atomicity
  - Coherence
  - Isolation
  - Durable

Pourquoi (allons-nous encore) utiliser le modèle relationnel ?

- SQL est un langage **déclaratif** (expressif, optimisation automatique, pas liée à une implémentation, parallélisable)
- accès à de **nombreuses implémentations** en fonction des besoins
- supporte les documents (JSON, XML)
- permet de réaliser du **code applicatif plus simple** (données connectées, procédures stockées, déclencheurs, moins de requêtes, moins d'I/O, etc.)
- etc.

Installation de MySQL (licence libre)

Installer MySQL Community Edition (licence libre) sur votre machine

Références

- Performance Optimization - Fabien Potencier - PHP Tour 2016, conférence de Fabien Potencier (Créateur des composants Symfony, Twig et bien d'autres choses) sur les performances des applications webs

MySQL (logiciel)

- MySQL Community Edition, lien pour télécharger MySQL Community Edition. Sous Windows, suivez ce lien, ou ici pour Debian/Ubuntu ou macOS
- Tutoriel d'installation de MySQL Community sur Windows
- Tutoriel d'installation sous Debian, debian vient avec mariadb par défaut. Les dépôts de MySQL Community ne sont pas accessibles par défaut, il faut les ajouter explicitement.

References

Standard SQL

- SQL (wiki)
- Datalog (wiki), le langage de requête qui a inspiré SQL
- Apprendre le SQL, bonne documentation en ligne sur le SQL et ses différentes implémentations
- Use the index, Luke !, site maintenu par Markus Winand, une référence des bases de données relationnelles. Vous trouverez sur son site une explication complète de l'optimisation et des performances des bases de données relationnelles. Le contenu est disponible en ligne ou en PDF (traduit en français)
- Modern SQL, a lot has changed since SQL-92, le site cousin du site Use the index, Luke !, de Markus Winand. Aborde les évolutions les plus récentes du standard SQL
- Le site de Markus Winand, Markus Winand est un expert du standard SQL.

References

Standard SQL

- PostgreSQL, le choix à faire lorsqu'on part sur une base de données relationnelle. Projet open source maintenu depuis 30ans, le SGBDR qui va le plus loin sur l'implémentation du standard SQL
- The Mother of all Query Languages: SQL in Modern Times | Markus Winand (Conférence), conférence de Markus Winand (Data Natives Berlin, 2018) sur l'état actuel du standard SQL, de l'évolution de ses différentes implémentations (Oracle, PostgreSQL, MySQL, etc.)
- A Relational Model of Data for Large Shared Data Banks, Edgard Codd (PDF), la publication originale de Edgard Codd de 1970, retenu par l'histoire comme élément fondateur du développement du modèle relationnel

Ouvrages

- Designing data intensive applications, de Martin Kleppmann, 2017, publié par O'Reilly, une référence absolue sur les systèmes d'information et sur les SGBD
- Programmer avec MySQL: SQL - Transactions - PHP - Java - Optimisations - XML - JSON - Avec 40 exercices corrigés, 6e édition, de Christian Soutou, 2021, par Eyrolles. Très bon livre et complet sur le modèle relationnel et le SGBD MySQL, avec beaucoup d'exercices corrigés. Bon ouvrage pour débutants (et confirmés) pour apprendre les bases et pratiquer.
- Modélisation des bases de données, UML et les modèles entité-association, 4e édition, de Christian Soutou, 2017, publié aux éditions Eyrolles. *Très bonne référence*. Principalement sur le passage du recueil des besoins à la conception d'une base de données (du modèle conceptuel au modèle physique) en suivant les meilleurs pratiques (techniques et de modélisation). Casse le bras aux idées reçues sur le modèle relationnel, et aborde les points importants pour utiliser à pleine puissance ce modèle (faites des jointures !). À étudier/parcourir *après avoir acquis les bases* et pratiqué un peu. À garder auprès de soi en cas de besoin.
- Les meilleurs livres MySQL, une liste de livres publiés en français sur MySQL, préparée par le site developpez.com.