

Modélisation de données

Perspectives

AAA_01b
v201b

2022-01-04



Christina.Khnaisser@USherbrooke.ca

Luc.Lavoie@USherbrooke.ca

© 2018-2021, Μηττις (<http://info.usherbrooke.ca/llavoie>)
CC BY-NC-SA 4.0 (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

Plan

- Tour d'horizon
- Historique
- SGBD, un deuxième coup d'oeil
- Vocabulaire
- Références

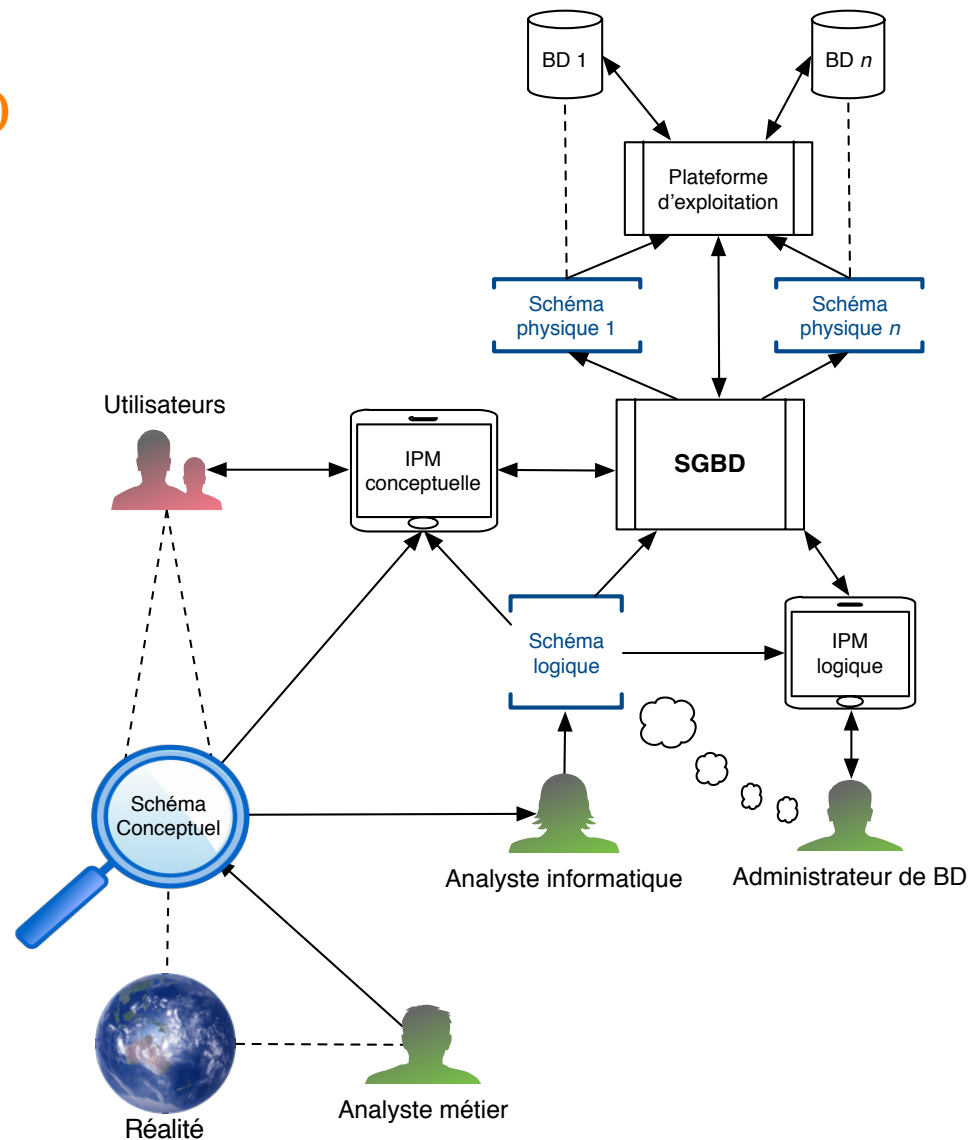


Tour d'horizon

- Aperçu d'un SGBD
- Besoins et propriétés
- Fonctions et services
- Rôles et utilisateurs

Tour d'horizon — Aperçu d'un SGBD (système de gestion de bases de données)

- L'utilisateur exploite les données.
- L'analyste du domaine (analyste-métier) élabore un schéma conceptuel (SC) des informations.
- L'analyste informatique traduit le SC en schéma logique de données.
- L'administrateur de bases de données voit à l'intendance et l'exploitation de la base de données (BD).



Tour d'horizon SGBDR

- Un SGBDR est une *machine* permettant de créer, modifier, détruire, gérer et interroger concurremment plusieurs bases de données **relationnelles**.
- Une machine est aussi appelée un système ou un service.

Tour d'horizon

Base de données relationnelle

- Une base de données relationnelle est un ensemble de *variables* respectant un schéma **relationnel**.
- Toute variable d'une base de données relationnelle est de « type relation », c'est-à-dire que la valeur référée par la variable est une relation.

Tour d'horizon

Schéma relationnel

- Un schéma relationnel est un ensemble de déclarations de **relations** soumises un ensemble de contraintes.
- Une déclaration de relation est la définition d'un « type relation » sur la base d'un entête de relation ET la déclaration d'une variable de ce type.

Tour d'horizon

Contrainte

- Une contrainte est une expression logique (un prédicat) dont les variables sont des variables de relation.
- Une contrainte est aussi appelée, selon le contexte, invariant.

Tour d'horizon

Besoins à satisfaire

- Acquisition
- Stockage (rémanent)
- Traitement
- Accès
 - concurrent
 - contrôlé
- Traçabilité
- Disponibilité

L'importance de ces besoins induit un système complexe de (très) grande envergure

Propriétés recherchées

- Exactitude
- Cohérence
- Intégrité
- Fiabilité
- Robustesse
- Performance
- Évolutivité dynamique
- ...

Le niveau de qualité recherché induit une réalisation soignée selon des méthodes rigoureuses mobilisant des experts de premier rang

Tour d'horizon

Fonctions requises

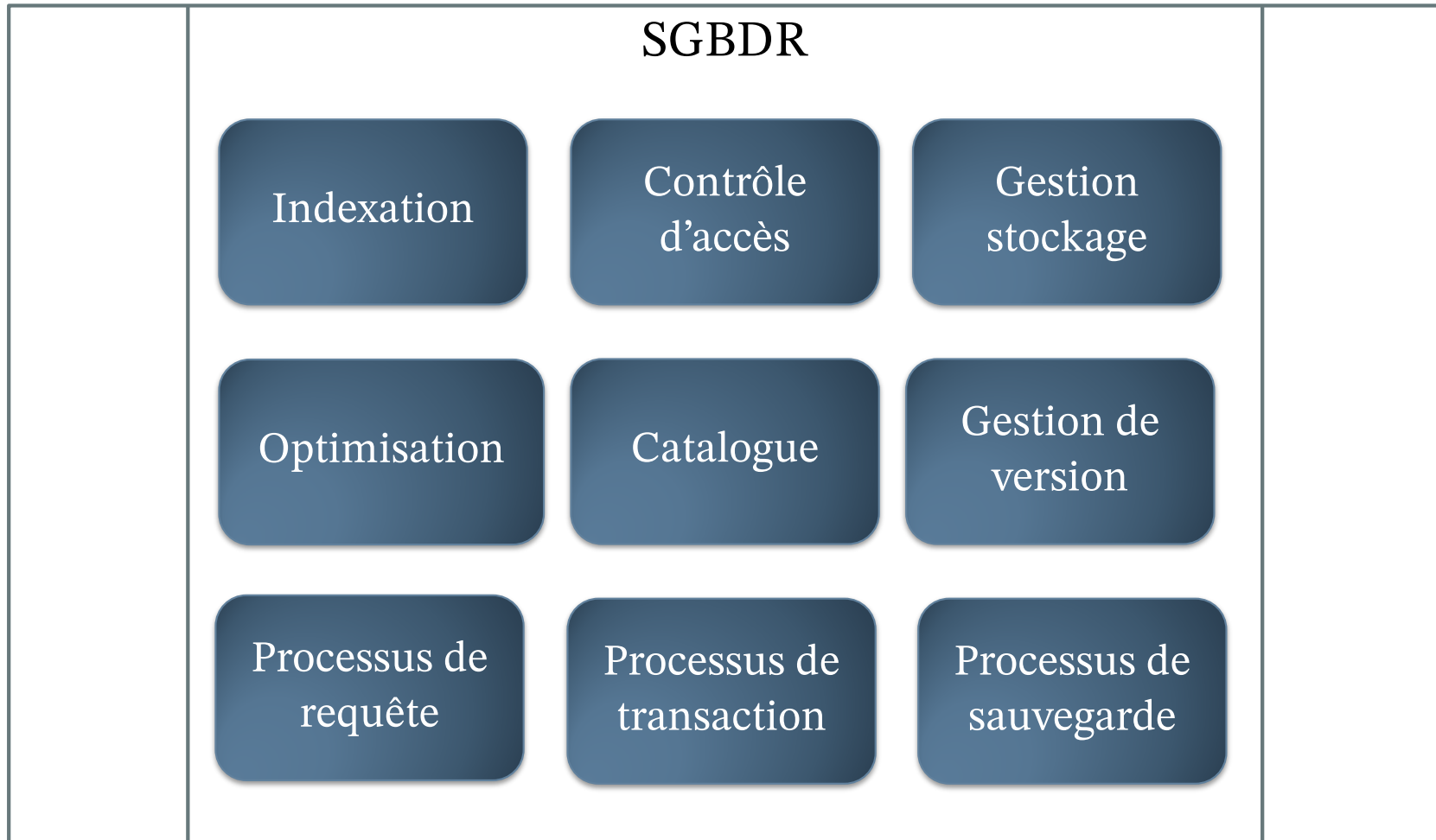
- Définition
- Construction
- Modification
- Interrogation
- Contrôle
- Évolution
- ...

Le nombre, l'ampleur et la complexité des fonctions et des services induisent des interfaces variées et complexes

Services offerts

- Autodescription (catalogue)
 - référence unique, cohérente, intègre
- Encapsulation et modularisation
 - abstraction
 - indépendance
 - entre logiciel et données
 - entre gestion et application
- Particularisation et spécialisation
 - spécificité (vues)
 - sécurité (règles)
- Traitement transactionnel
 - partage et concurrence
- ...

Tour d'horizon



Tour d'horizon

Rôles diversifiés

- Exploitation
 - fourniture
 - saisie
 - validation
 - analyse
 - gestion
- Conception
 - base de données
 - application
 - système de gestion (SGBD)
 - outils

Utilisateurs diversifiés

- Utilisateurs
 - occasionnels
 - experts
 - autonomes
- Préposés à l'exploitation
 - gestionnaire de BD,
data base administrator (DBA)
 - technicien de BD,
database operator
- Informaticiens
 - concepteur logiciel
 - analyste logiciel (de BD)
 - analyste d'affaires
- Gestionnaires
 - directeur informatique
 - directeur des TI
chief information officer (CIO)

Historique

- Montée en puissance
- Théorie relationnelle
- Crise 1985-2005

Historique — la montée en puissance

1960

- Modèles primitifs
 - fichiers
 - index
 - hiérarchies

1970

- Modèles à 2 niveaux
 - Modèle réseau (graphe)
 - CODASYL
 - **Début** du modèle relationnel
 - Alpha, R, SEQUEL
 - QBE, Ingres...

1980

- Modèles à 3 niveaux
- **Généralisation** relationnelle
 - Optimisation locale
 - Normalisation de SQL
 - SPARC, ANSI

1990

- **Domination** relationnelle
 - Optimisation globale
 - Normalisation de SQL
 - ISO
 - Répartition et distribution
- Apparition des BDOO

Historique — la théorie relationnelle


1969-1976

- Recherche
 - Travaux de Codd
 - Travaux de Date
 - ...
- Prototypes
 - Systèmes R (IBM)
 - INGRES
 - SEQUEL (IBM)
 - QBE (IBM)
 - SQL
 - Oracle
 - ...

1976-2015

- Évolution
 - Normalisation
 - OLAP (Codd)
 - Travaux d'Ullman
 - Travaux de Date
 - Typage (Cardelli)
 - Temporalité (Allen et Lorentzos)
 - Co-relationalité (noSQL)
 - Néo-relationalité (newSQL)

Historique — la crise 1985-2006

- Grandeur et décadence du modèle relationnel
 - ANSI SQL-89
 - ANSI SQL-92
 - ANSI SQL-99
 - Parallèlement, une approche objet qui ne s'impose pas... *et pour cause!*
 - CORBA
 - ODMG
 - ...
 - La souplesse et ses limites
 - XML
 - La poursuite d'une normalisation de moins en moins suivie
 - ISO 9075:1999
 - ISO 9075:2003
 - ISO 9075:2006
 - ISO 9075:2008
 - ISO 9075:2011
 - ISO 9075:2016
- 

Historique — les défis depuis 2006

- Intégrer le flou
 - noSQL
- Intégrer le temps
 - BCDM
 - DDLM
 - ...
- Intégrer l'intégration
 - MOLAP, ROLAP, HOLAP, DOLAP
 - Dimensionnalité (étoiles et flocons)
 - Entrepôts (ETL et médiateurs)

Allons-nous vers une renaissance du modèle relationnel avec le newSQL?

...
D
Vertica
Hanna
VoltDB
...

Historique — en guise de conclusion

- On assiste à un retour triomphal (?) du relationnel « pur » avec le « newSQL » en raison de :
 - sa capacité de maintenir les propriétés ACID (parce que, finalement, une réponse exacte... c'est souvent utile, voire nécessaire),
 - sa capacité de modéliser la temporalité (grâce à la logique des intervalles),
 - ses performances insurpassées (grâce aux représentations verticales et maillées),
 - aux capacités de plus en plus grandes des mémoires primaires (RAM).

Prospective – Que 2022 nous réserve-t-il ?

- Initialement prévu pour 2021, le nouveau standard ISO de SQL pourrait être officialisé en aout 2022.
- La principale innovation est l'inclusion du sous-langage spécialisé SQL/PGQ pour le traitement des graphes (un cas particulier des relations) aussi connu sous le nom de GQL.
- Voir
 - <https://www.gqlstandards.org/what-is-a-gql-standard>
 - <https://www.iso.org/standard/79473.html>

SQL, noSQL, newSQL

Le point de vue
de
Bharat Upadrasta
et
Austin Chungath

LL :

- (1) **Oui**
- (2) **Grand**
- (3) **Dév. grandissant.**

| Caractéristiques | SQL | noSQL | newSQL |
|--|--------|-------------|--------------------------|
| Propriétés ACID (intégrité des transactions) | Oui | Non | Oui |
| OLAP/OLTP | Oui | Non | Oui |
| Analyse des données | Oui | Non | Oui |
| Schéma défini (connu à l'avance) | Oui | Non | Peut-être ⁽¹⁾ |
| Flexibilité du format des données | Non | Oui | Peut-être |
| Calcul distribué | Oui | Oui | Oui |
| Passage à l'échelle (<i>Scale</i>) | Oui | Oui | Oui |
| Performance d'accès aux données volumineuses | Rapide | Rapide | Très rapide |
| Utilisation ressources | Grand | Moyen | Minimal ⁽²⁾ |
| Popularité / support de la communauté | Grand | Dév. rapide | Dév. lent ⁽³⁾ |

source :

<http://www.informationweek.com/big-data/big-data-analytics/nosql-newsqli-or-rdbms-how-to-choose/a/d-id/1297861>
consultée le 2014-08-24, re 2021-01-18

SGBD — un deuxième coup d'oeil

- Modèles
- Langages
- Architectures
- Plateformes et produits
- Avantages et contre-indications
- Conclusion

SGBD — modèles

Modèles de schémas

- **b**ischématique
 - externe (logique)
 - interne (physique)
- **t**rischématique
 - externe (conceptuel)
 - logique
 - interne (physique)
- **q**uadrischématique
 - connaissance
 - information (conceptuel)
 - données (logique)
 - représentation (physique)

Modèles d'application utilisant une BD

- monolithique
 - 1 niveau
- client-serveur
 - 2 niveaux
- MVC (modèle-vue-contrôleur)
 - 3 niveaux
- multi-niveaux
 - ≥ 3 niveaux

SGBD — langages

Portée

- externe (conceptuel)
 - LRD
 - requêtes
- logiques
 - LDD (définition)
 - définition des schémas
 - définitions des types
 - LMD (manipulation)
 - modification des données
 - interrogation des données
 - LCD (contrôle)
 - contrôle d'accès aux données
 - contrôle des transactions
- interne (physique)
 - LSD
 - stockage et représentation

Style

- Procédural, axiomatique ou fonctionnel
- Ensembliste ou algébrique
- Textuel ou graphique

À l'origine, SQL se voulait axiomatique, ensembliste et textuel...

SGBD — Architecture (interfaces)

- Externe
 - utilisateur → (formule des) requêtes
 - programmeur → (élabore des) logiciels
 - administrateur → (gère) schémas et données
- Interne (système d'exploitation)
 - gestionnaire de fichiers (mémoire rémanente)
 - méthodes d'accès
 - gestionnaire de tampons systèmes
 - pilotes
 - ...

SGBD — Architecture (composants)

○ SGBD

- Compilateur de LDD
- Analyseur de LMD
- Analyseur de LCD
- Logithèque (types et opérateurs)
- Gestionnaire de catalogue
- Gestionnaire de BD
 - ... (détail sur la diapositive suivante)...

SGBD — Architecture (sous-composants)

○ SGBD

- Gestionnaire de BD
 - Gestionnaire d'accès (sécurité)
 - Processeur de commandes
 - Vérificateur d'intégrité
 - Optimisateur de commandes
 - Gestionnaire de transactions
 - Planificateur
 - Gestionnaire de restauration
 - Gestionnaire de tampons
 - ...

SGBD — plateformes et produits

- Super-calculateur
 - Serveur
 - Équipement spécialisé (*appliance*)
 - Poste autonome
 - Assistant mobile
- Adabase
 - IDMS
 - Oracle
 - DB2
 - MS-SQL
 - INGRES
 - Sybase
 - Informix
 - PostgreSQL
 - MariaSQL
 - SQL lite
 - Berkeley DB
 - VoltDB
 - FileMaker
 - Access
 - ...

SGBD

Avantages des SGBD comparativement aux solutions *ad hoc*

| Avantages SGBD | Contre-indication |
|---------------------------|---|
| Normalisation | Problèmes hors norme. (on cherche encore lesquels 😊) |
| Évolutivité | Usage unique (par un seul client, dans un environnement figé!) |
| Multi-accès | Accès unique en tout temps (sur et depuis un seul ordinateur) |
| Économie d'échelle | Cout de développement initial faible et garanti (qui peut garantir cela?) |
| Mise à disposition rapide | Temps de développement initial court et garanti (qui peut garantir cela?) |
| Très haute fiabilité | Spécification formelle et programmation automatisée du système |
| Auto-documentation | Documentation complète pré-existante |
| Traçabilité | Il faut choisir entre performance et traçabilité |
| ... | Application de type temps réel critique |

SGBD – en guise de conclusion

2015

Figure 1. Magic Quadrant for Operational Database Management Systems



2017

Figure 1. Magic Quadrant for Operational Database Management Systems



SGBD – en guise de conclusion

2018



2019

Figure 1. Magic Quadrant for Operational Database Management Systems



Références

- Bases de données (Databases) :
 - [Date 2004], chapitres 1 et 2;
 - [Elmasri and Navathe 2004], chapitres 1 et 2;
 - [Elmasri and Navathe 2016], chapitres 1 et 2;
 - [Ullman and Widom 2008], chapitre 1;
 - https://fr.wikipedia.org/wiki/Base_de_données
 - <https://en.wikipedia.org/wiki/Database>
- Mégadonnées (Big Data) :
 - https://fr.wikipedia.org/wiki/Big_data
 - https://en.wikipedia.org/wiki/Big_data

Les colles du prof

- Qu'est-ce qu'un SGBDR ?
- Qui sont les principaux utilisateurs d'un SGBD ?
- Quelles sont les fonctionnalités d'un SGBDR ?
- Quel est le modèle prédominant dans les bases de données ?
- Quel modèle doit-on choisir pour développer un logiciel de gestion de prescriptions de médicament ?
- Quel modèle doit-on choisir pour développer pour un blogue sur les technologies ?
- Quel modèle doit-on choisir pour développer pour un site web d'achat en ligne ?

