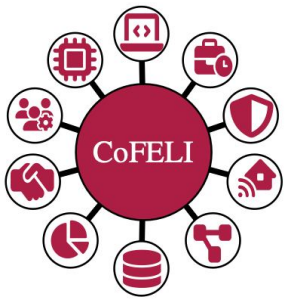


Collectif francophone pour l'enseignement libre de l'informatique

Systèmes de gestion de bases de données

Stockage

CoFELI:SGBD_01

Christina KHNAISSER *(christina.khnaisser@usherbrooke.ca)*

Luc LAVOIE *(luc.lavoie@usherbrooke.ca)*

(les auteurs sont cités en ordre alphabétique nominal)

—

CoFELI/Scriptorum/SGBD_01-Stockage, version 1.0.1.a, en date du 2025-09-15

— document de travail, ne pas citer —

Sommaire

Introduction aux techniques de stockage de données massives utilisées dans les systèmes de gestion de bases de données.

Mise en garde

Le présent document est en cours d'élaboration ; en conséquence, il est incomplet et peut contenir des erreurs.

Historique

diffusion	resp.	description
2025-09-15	LL	Revue préalable aux activités 2025-1.
2024-09-16	CK	Récupération de notes diverses.
2024-08-16	LL	Ébauche initiale à partir de documents antérieurs produits entre 2004 et 2023.

Table des matières

Introduction

Le présent document a pour but de présenter une synthèse des techniques de stockage de données utilisées par les SGBD. La présentation repose sur une connaissance des bases de fonctionnement des systèmes de fichier contemporains.

Contenu des sections

- La section 1 expose la problématique générale du stockage des données en faisant ressortir le grand nombre de caractéristiques à considérer et la variété des solutions, ce qui justifie son abstraction dans une couche spécifique (la couche physique) au sein de l'architecture trischématique.
- La section 2 présente un survol de plusieurs technologies contemporaines, confirmant la problématique exposée à la section 1 et utilisées dans les exemples des hiérarchies de stockage classique (section 3) et contemporaine (section 7). Elle justifie également la nécessité de développer un modèle de stockage indépendant de l'interface de la couche physique.
- La section 3 présente les hiérarchies classiques des systèmes établis et courants.
- La section 4 rappelle certains principes organisationnels des fichiers qui seront utilisés pour présenter les interfaces d'accès et les mécanismes de palliation aux défaillances à la section 6 ainsi que les mécanismes de pagination à la section 5.
- La section 5 (à venir) présentera les fonctionnalités de la pagination et certaines techniques courantes.
- La section 6 (à venir) présentera les techniques de palliation aux défaillances.
- La section 7 (à venir) présentera les hiérarchies de stockage utilisées par certains systèmes de pointe.
- La section 8 (à venir) reprendra les différents critères de décision utilisés dans les sections précédentes et en dégagera les principales mesures utiles.

Le sujet de l'indexation, étroitement lié aux différents sujets traités dans le présent module, est traité dans un module distinct `SGBD_02-Indexation`.

Évolution du document

La première version du document a été établie sur la base des travaux publiés par Elmasri, Navathe, Snodgrass et Ullman.

1. Présentation

1.1. Mise ne contexte

L'organisation du stockage des données et l'ensemble des algorithmes d'accès aux données forment le modèle physique de données (selon l'architecture trischématique).

Le problème du stockage est celui de la conciliation de nombreuses caractéristiques des dispositifs à choisir, assembler et exploiter afin de garantir de façon exacte (cohérente, valide et efficace), pratique (suffisamment complète, efficiente et évolutive) et pérenne le stockage et l'accès à un volume de données déterminé.

1.2. Attentes

Entre autres caractéristiques déterminantes, on peut noter les suivantes :

- capacité de stockage,
- granularité de l'adressage,
- délai d'accès (lecture, écriture) aussi appelé latence,
- débit de transfert (lecture, écriture),
- rémanence (intrinsèque [durée], extrinsèque [consommation])
- taux d'erreur (au repos, en lecture, en écriture),
- cout (acquisition, exploitation, disposition),
- quantité d'énergie requise (en fonction d'un profil de séquences d'opérations données),
- quantité de chaleur dégagée (en fonction d'un profil de séquences d'opérations données),
- durée de vie (en fonction d'un profil de séquences d'opérations données),
- encombrement physique (volume),
- canal (type de mécanisme de transfert de données et protocole afférent, comprend entre autres l'adressage direct, le bus, réseau de communication)
- connexion (connecteur, standard, protocole),
- alimentation énergétique (voltage, ampérage, fréquence, variation de puissance).

Auxquelles, en pratique, il faudrait ajouter celles-ci (entre autres) :

- disponibilité et variation de la disponibilité du dispositif,
- cout et variation du cout du dispositif,
- garantie et variation de garantie du dispositif,
- disponibilité du soutien technique,
- disponibilité des pièces,
- facilité de réparation.

2. Technologies

Trois catégories de dispositifs de stockage sont présentées :

- électro-mécaniques,
- électroniques,
- autres.

Les dispositifs de stockage les plus couramment utilisés se classent dans les deux premières catégories.

2.1. Dispositifs électro-mécaniques

C’est-à-dire tout dispositif dont le stockage et les fonctions d’accès ne dépendent que de composantes électroniques et mécaniques (au moins une électronique et une mécanique).

2.1.1. Exemples contemporains

Disque « magnétique »

Exploration autonome, voir références.

Disque « laser »

Exploration autonome, voir références.

Ruban « magnétique » (en cassette)

Exploration autonome, voir références.

2.1.2. Exemples historiques

Ruban perforé

Exploration autonome, voir références.

Carte perforée

Exploration autonome, voir références.

Tambour « magnétique »

Exploration autonome, voir références.

Ruban « magnétique » (en bobine)

Exploration autonome, voir références.

2.2. Dispositifs électroniques

C’est-à-dire tout dispositif dont le stockage et les fonctions d’accès ne dépendent que de composantes électroniques.

2.2.1. Exemples contemporains

MM0

Mémoire DDR SDRAM typiquement intégrée aux circuits de la puce hébergeant le processeur selon l’une des configurations suivantes :

- répartie entre les coeurs de calculs (donc sans contention),
- commune entre eux (donc avec contention possible),
- hybridée (généralement avec peu ou pas de contention).

Ce type de mémoire est caractérisée par un accès uniforme direct sans latence ni variance notable entre les emplacements.

MM1

Mémoire (SDR|DDR) (SDRAM|DDRAM) typiquement organisée en barrettes et accessible au processeur par l’entremise d’un bus **interne**.

Accès uniforme direct avec très faible latence et très faible variance entre les emplacements. Lorsque le bus est partagé, en fonction du trafic, la contention peut augmenter significativement la latence et la variabilité.

MM2 (memory module)

Mémoire DRAM ou NVRAM typiquement organisée en barrettes et accessible au processeur par l’entremise d’un bus **externe**.

Accès uniforme direct avec faible latence et faible variance entre les emplacements Lorsque le bus est partagé, en fonction du trafic, la contention peut augmenter significativement la latence et la variabilité.

SSD (solid state drive)

Mémoire NVRAM typiquement organisée en barrettes et accessible au processeur par l'entremise d'un bus **externe**.

Le SSD utilise les mêmes technologies de base que le MM2 doté de NVRAM, mais offre une rémanence intrinsèque. Il est généralement conditionné comme un dispositif indépendant, ayant sa propre alimentation électrique et utilisant les mêmes canaux et protocoles que les disques électro-mécaniques.

2.3. Autres dispositifs

Plusieurs autres technologies sont utilisées (mémoires optiques) ou en émergence (mémoires ADN), mais non présentées dans le cadre du présent module.

3. Hiérarchie classique des stockages

Les données d'une base de données sont stockées physiquement grâce à un dispositif de stockage. Il n'est toutefois généralement pas souhaitable, voire possible, de n'utiliser qu'un seul type de dispositif.

Pour les opérations en temps réel (ou quasi réel), on voudra privilégier la vitesse. Par contre, le cout plus élevé et les capacités limitées de ces dispositifs amènent le plus souvent à recourir à des dispositifs moins chers, à plus grande capacité. Par ailleurs, la rémanence est une propriété importante qui doit être assurée à long terme grâce au journal de transaction, aux copies de sécurité et aux archives (le « D » de ACID). Pour cela, d'autres dispositifs stables, à grande capacité et peu couteux sont requis.

Discussion

Quelles sont les caractéristiques distinguant les journaux, les copies de sécurité et les archives ?

Organisation

Typiquement, au moins trois niveaux de dispositifs seront mis à contribution, parfois quatre ou cinq. Il faut donc prévoir des mécanismes de transfert entre ces niveaux (typiquement la pagination, la journalisation, la duplication et la réplication).

Les trois prochaines sous-sections décrivent une hiérarchie classique en trois niveaux.

3.1. Stockage primaire

C'est le dispositif devant pouvoir être exploité directement par l'unité de traitement (*central processing unit CPU*) de l'ordinateur. On distingue deux sous-niveaux :

- La **mémoire cache** (historiquement SRAM, désormais MM0) requise pour l'exécution des instructions et la composition des résultats.
- La **mémoire principale** (historiquement SRAM, désormais MM1) pour sauvegarder les instructions elles-mêmes et les données requises par une transaction.

Le stockage primaire permet généralement un accès rapide aux données, mais sa capacité de stockage est limitée et le contenu peut être perdu en cas de coupure de courant ou de panne du système.

3.2. Stockage secondaire

C'est le dispositif qui permet de stocker en ligne l'ensemble des données de l'état courant de la base de données.

Les technologies le plus souvent utilisées sont le disque magnétique et le SSD.

3.3. Stockage tertiaire

On distingue deux sous-niveaux de stockage tertiaire selon qu'il doit impérativement être hors site ou en ligne. On distingue deux sous-niveaux :

Le *stockage hors site est requis pour les copies de sécurité et les archives.

Le **stockage proximal** est principalement destiné au journal principal (dont une copie pourra toutefois être externalisée hors site et dès lors traité comme une copie de sécurité).

3.4. Tendances

De nos jours, plusieurs ordinateurs peuvent être dotés d'une grande capacité en mémoire primaire, de sorte qu'il devient possible d'y conserver la totalité de la base de données. On parle alors de base de données « en mémoire » (*in memory database*).

Dans ce cas, des copies de sécurité plus fréquentes, appelées instantanés (*snapshot*), sont requise afin de pallier les pannes sans avoir à « rejouer » le journal sur une trop longue durée. Ces instantanés sont généralement conservés en mémoire secondaire.

3.5. La suite

La grande variété des technologies et des niveaux de stockage requis par le SGBD est susceptible d'entrainer une grande complexité de mise en oeuvre (comportant un paramétrage important variant dynamiquement en fonction des conditions d'exploitation) au risque de transparaitre dans les modèles logiques et conceptuels.

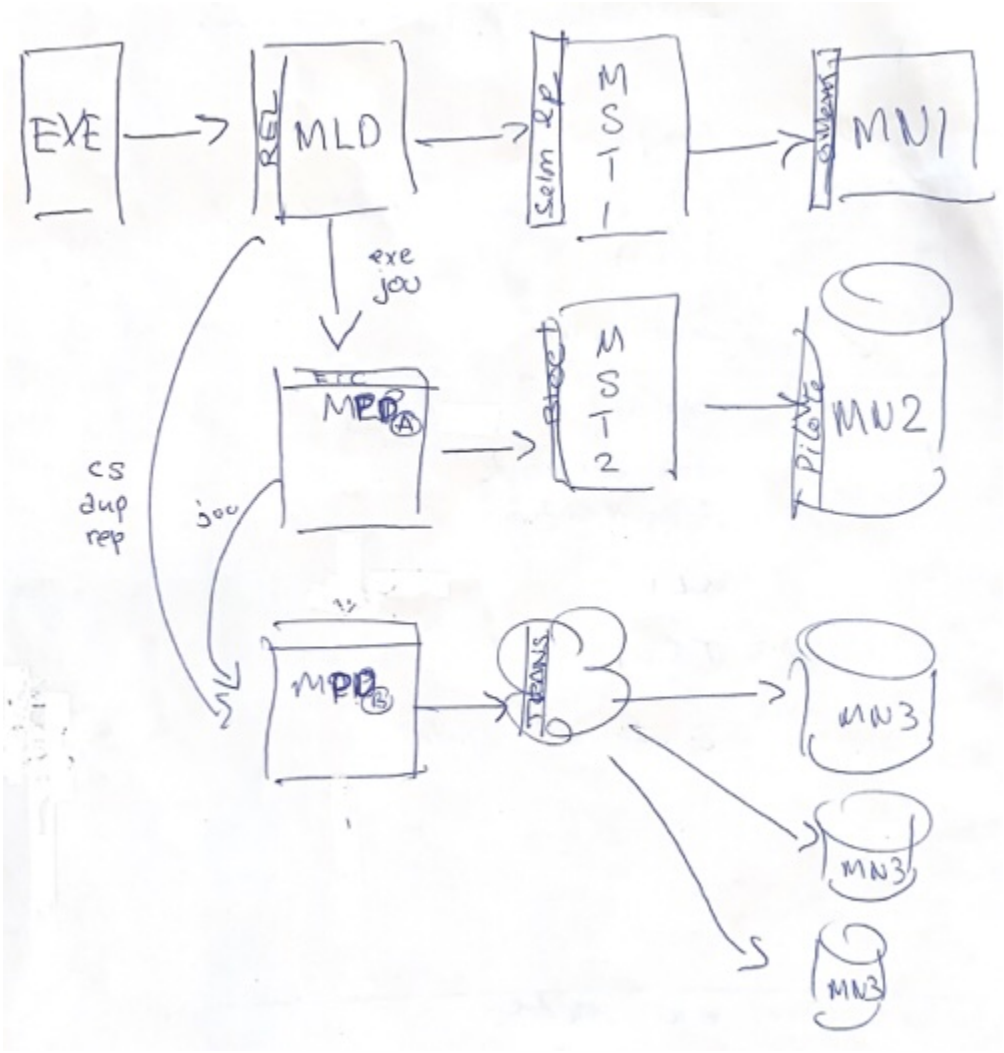
Pour éviter cela, un modèle physique, reposant sur un modèle de stockage, doit être développé et sera présenté dans les deux prochaines sections.

4. Modèle physique

Le modèle physique a pour but de découpler le modèle logique (un modèle relationnel, un modèle hiérarchique, etc.) des modèles du stockage. Ainsi, pour chaque classe d'entités du modèle logique, un fichier sera associé, et, pour chacun chaque instance d'entité, un enregistrement dans ce fichier.

La mémoire secondaire est structurée en blocs. Une page est composée d'une suite de blocs consécutifs, de façon à permettre d'établir un compromis entre le temps d'accès et le volume d'accès.

La couche physique traduit donc les commandes de la couche logique (exprimée en termes d'opérations sur les classes d'entités) en commandes destinées aux dispositifs de stockage (exprimées en termes de blocs ou de pages). Les opérations offertes par le modèle physique au modèle logique varieront selon ce dernier. Typiquement, pour le modèle relationnel, il comprendra minimalement les opérateurs de l'algèbre relationnelle.



Typiquement, le modèle de stockage doit permettre minimalement

- la définition et allocation d'un fichier sur la base du type d'enregistrement
- la définition et allocation d'index sur un fichier
- l'ajout d'un ensemble d'enregistrements à un fichier
- le retrait d'un ensemble d'enregistrements à un fichier (identifiés par leurs clés)
- le parcours des enregistrements d'un fichier (éventuellement selon un ordre spécifié en termes des clés)
- l'accès à un enregistrement identifié par une de ses clés
- la modification des attributs d'un enregistrement identifié par une de ses clés

La mise en oeuvre du modèle de stockage nécessite, notamment pour des raisons d'efficience, le recours à l'indexation (traitée dans le module suivant) et de la pagination (intermédiaire incontournable entre deux niveaux de mémoires).

Dans le modèle physique, les données sont organisées sous la forme d'un ensemble de fichiers. Les fichiers sont eux-mêmes une liste d'enregistrements. Typiquement chaque fichier représente une classe d'entités. Dans un SGBDR, chaque fichier représentera une variable de relation.

4.1. Fichier

Un fichier est une séquence d'enregistrements. Les enregistrements d'un fichier sont souvent de même type, mais peuvent être de même taille (*fixed-length record*) ou de taille variable (*variable-lenght records*).

Pour les enregistrements de longueur fixe, tous les enregistrements sont de même type et tous les champs sont de tailles fixes. Ainsi le système peut identifier le début de chaque champ par rapport à la position de départ de l'enregistrement.

Pour les enregistrements de longueur variable, la taille des champs est variable. Le système détermine la position d'un champ en utilisant des caractères de séparations à la fin de chaque champ ou en enregistrant la taille effective du champ après sa valeur.

Les champs optionnels peuvent prendre différents formats : la valeur « NULL » est enregistrée. Pour généraliser la structure, une paire <nom du champ = valeur du champ> est enregistrée.

Pour les fichiers qui contiennent différents types d'enregistrements, chaque enregistrement est précédé par un indicateur de type.

4.2. Type d'enregistrement (*record type*)

Une liste d'attributs (champ, *field*) chacun identifié par un identifiant (nom, *name*) et défini par un type. Le nombre d'octets requis pour chaque valeur (la taille) est déterminé par le type, mais peut varier en fonction du SGBD.

4.3. Enregistrement (*record*)

Chaque enregistrement est formé d'une liste de valeurs. Chaque valeur est formée d'un ou de plusieurs octets qui correspondent à un champ de l'enregistrement (*record field*). Dans un SGBD, les enregistrements correspondent aux instances d'entités. Dans un SGBDR, ils correspondent aux tuplets des variables de relation.

Lorsque la taille des valeurs d'un même attribut varie, une partie fixe est conservée dans l'enregistrement et le complément est relocalisé dans une zone de débordement commune à plusieurs enregistrements.

Dans certaines applications de base de données, il peut s'avérer nécessaire de stocker des éléments de données constitués de grands objets non structurés, qui représentent des images, des vidéos et des sons. On parle alors d'objets binaires de grande taille (BLOB). Un BLOB est généralement stocké séparément de son enregistrement dans un groupe (pool) de blocs de disque, et un pointeur vers le BLOB est inclus dans l'enregistrement.

En fait, le BLOB est un cas particulier de valeur de taille variable souvent traité de façon distincte des autres valeurs de taille variable en raison de sa très grande taille. La différenciation s'opère généralement en fonction de la taille moyenne de la valeur :

- moins d'une page, les différents types partagent une même espace de débordement utilisant une même fonction de compression ;
- plus d'une page, chaque type a un espace dédié avec sa propre fonction de compression.

4.4. Bloc

Un bloc est une unité de données. Un bloc est utilisé lors du transfert de données entre le disque et la mémoire. Les enregistrements d'un fichier sont organisés en bloc (ou page) sur le disque. La taille du bloc est configurable et varie selon les dispositifs.

Pour affranchir le mécanisme de gestion des fichiers de la variabilité de la taille des blocs en fonction des dispositifs de stockage, une page (représentant un nombre entier de blocs consécutifs) est généralement définie.

Avec une organisation non étendue (*unspanned organization*), chaque enregistrement se trouve dans une page. Lorsque la taille de la page est supérieure à la taille d'un enregistrement, la page peut contenir plusieurs enregistrements. Cependant, la page peut avoir de l'espace non utilisé.

Avec une organisation étendue (*spanned organization*), un enregistrement peut se trouver réparti en plusieurs pages. Dans ce cas, à la fin de la première page, un pointeur est ajouté pour indiquer la page contenant la suite de l'enregistrement.

Exemple

En PostgreSQL, la taille par défaut d'une page est de 8192 octets. Cette valeur est paramétrable au moment du déploiement du SGBD.

Allocation de blocs sur le disque

Il existe plusieurs techniques d'allocation de blocs de fichiers sur un disque :

- Une allocation contigüe (*contiguous allocation*), les blocs de fichiers sont alloués à des blocs de disque consécutifs. Un fichier occupe un ensemble contigu de blocs de disque Cette technique rend la lecture d'un fichier très rapide, mais l'augmentation de la taille du fichier devient plus difficile.
- Une allocation chainée (*linked allocation*), chaque bloc de fichier contient un pointeur vers le bloc de disque suivant. Un fichier est une liste chainée de blocs de disque. Cette technique rend l'augmentation de la taille du fichier plus facile, mais la lecture plus lente.
- Une allocation indexée (*indexed allocation*), un bloc d'index contient les pointeurs par les blocs de fichiers.

4.5. Entête d'un fichier

Un entête de fichier contient des informations qui sont nécessaires au programme qui accède aux enregistrements du fichier. L'entête comprend des informations permettant de déterminer les adresses de disque des blocs de fichier ainsi que des descriptions de format des enregistrements incluant : la longueur des champs, l'ordre des champs, les codes de type de champs, les caractères de séparation et les codes de type d'enregistrements.

4.6. Catalogue

Manipulation du catalogue

La liste des fonctionnalités varie beaucoup selon la structure du catalogue (en particulier selon qu'il est hiérarchisé ou non) minimalement, on retrouve :

- liste critériée de fichiers
- parcours des fichiers
- obtention de l'entête d'un fichier

4.6.1. Hiérarchisation (ou non) du catalogue

La hiérarchisation du catalogue (par l'entremise de dossiers, sous-dossiers, etc.) n'est pas requise par le stockage des données du SGBD. Elle peut cependant être prise en charge, notamment lorsque le système de gestion de fichiers du système d'exploitation sous-jacent est utilisé. Lorsqu'un système propre au SGBD est privilégié, elle est souvent omise.

4.6.2. Indirection (ou non) des fichiers

La double indirection au niveau du catalogue des fichiers est un impératif de performance. L'absence de redirection est souvent un motif pour lequel les systèmes de fichiers de certains systèmes d'exploitation ne seront pas retenus.

Voici quelques exemples :

- Poursuite du traitement sans perturbation en cas de renommage.

- Poursuite du traitement en cas de relocalisation.
- Préparation d'une mise à jour d'un fichier en concurrence avec l'exploitation de la valeur précédente du fichier. Au moment opportun (déterminé par le gestionnaire transactionnel), le deuxième pointeur est simplement modifié, sans impact sur les références conservées en mémoire primaire.

4.7. Organisation des dispositifs matériels

Bien qu'un seul dispositif matériel puisse être utilisé (voir un petit nombre), le plus souvent, plusieurs seront associés afin de satisfaire aux critères de cout, de capacité ou de délai. L'utilisation de plusieurs dispositifs est aussi un élément de palliation déterminant.

4.7.1. Dispositif unique

4.7.2. NAS

4.7.3. SAN

4.8. Palliation

4.8.1. Miroir

4.8.2. RAID

4.8.3. autres

5. Pagination

- En direct, en présence, au tableau !

Conclusion

Références

[Elmasri2016]
Ramez ELMASRI et Shamkant B. NAVATHE;
Fundamentals of database systems;
7th Edition, Pearson, Hoboken (NJ, US), 2016;
ISBN 978-0-13-397077-7.

[Lelarge2023a]
Guillaume LELARGE, Juilen ROUOHAUD;
PostgreSQL : architecture et notions avancées;
5e édition, Éditions D_Booker, 2023;
ISBN 978-2-8227-1124-1.

[Sciore2020a]
Edward SCIORE;
Database design and implementation;
Second edition, Springer, 2020;
ISBN 978-3-030-33855-0.

Définitions

```
Sources consultées de juin 2023 à juillet 2024

* Antidote : Antidote 11 v4.2 (2023), voir https://www.antidote.info
* Le Larousse : https://www.larousse.fr/dictionnaires/francais
* Le Robert : https://dictionnaire.lerobert.com
* Wikipédia : https://fr.wikipedia.org/wiki
```

Modèle physique de données

Le modèle physique de données (MPD) décrit la représentation des données (structure de données et méthodes d'accès) [DoDAF-DOD-DIV-3]. Au féminim, l'acronyme MPD désigne la modélisation physique de données.

Sigles

DRAM

…

SRAM

…

NVRAM

…

DDRAM

…

SDRAM

…

Produit le 2025-09-19 07:04:57 -0400

Collectif francophone pour l'enseignement libre de l'informatique