

IFT-6800, Automne 2016

Cours #4—Introduction aux bases de données

Louis Salvail

André-Aisenstadt, #3369

salvail@iro.umontreal.ca

Pourquoi les bases de données

- ❖ Des applications nécessitent l'organisation de quantité massive de données.
- ❖ Ces données sont souvent organisées selon une structure régulière et fixe.
Exemple: Nom, adresse, numéro de téléphone pour chaque entrée.
- ❖ Ces données doivent être consultées de diverses façons.
 - ❖ Une base de données:
 - ❖ permet de structurer les données selon un schéma fixe.
 - ❖ permet d'abstraire les données et de fournir une vue d'ensemble uniforme de celles-ci.
 - ❖ permet de traiter les données d'une façon efficace et standard: consultation, mise à jour, organisation.
 - ❖ permet de garantir la cohérence des données même si des situations anormales surviennent.

Définitions

- ❖ **Base de données:** Un ensemble organisé de données. Une collection de schémas, de tables, de requêtes, de vues (les résultats d'une requête mémorisée sur les données) et autres. Les données sont habituellement organisées pour représenter des aspects de la réalité d'une façon qui permette de l'interroger.
 - ❖ Exemple: Modéliser les vols des compagnies aériennes pour en trouver un sur lequel des sièges sont toujours disponibles.
- ❖ **DBMS (*database management system*):** Une application (un programme) qui interagit avec l'utilisateur, d'autres applications et la base de données pour représenter et analyser les données. Un DBMS est conçu pour permettre la définition, la création, le questionnement, la mise à jour et l'administration des bases de données.

Exemple

- ❖ Données de vente d'articles dans un magasin:
 - ❖ Date de vente,
 - ❖ No. de l'article,
 - ❖ Nom de l'article,
 - ❖ Prix,
 - ❖ Vendeur.
- ❖ Cette structure est fixe pour chaque article du magasin.
- ❖ Il y a une grande quantité de données qui sont représentées par la même structure.

Exemple: Représentation par une table

Date	no. article	Nom	Prix	Vendeur
2015/01/01	A1243	Table	356.75	Paul
2015/01/02	B1782	Chaise longue	750.50	Julie
...
2015/09/21	B7233	Tapis	1240.00	Martine
...

Traitements possibles

- ❖ Statistique:
 - ❖ volume de vente d'un jour,
 - ❖ volume de vente d'un vendeur,
 - ❖ les articles populaires (les plus vendus).
- ❖ Extraction d'une partie de données:
 - ❖ les ventes faites par Martine,
 - ❖ les ventes de plus de 70\$,
 - ❖ les vente de plus de 100\$ le jeudi.

Plus d'information

no. article	Fabricant	Pays	Stock
A1243	Plastic Tables inc.	USA	23
B1782	Chaises Laides inc.	Canada	234556
B7233	Tapis Perses inc.	Iran	2
D7823	Tables de plastique inc.	USA	12

Plus d'opérations

- ❖ Il est possible de combiner de l'information provenant de plusieurs tables:
 - ❖ Les ventes d'articles de la compagnie Plastic Tables inc.
 - ❖ Le nombre de tables en stock.
- ❖ On peut joindre 2 tables:

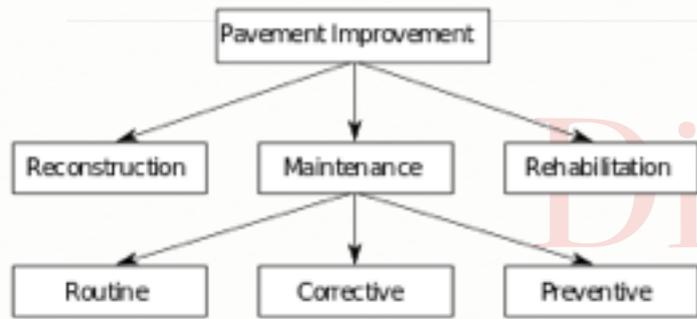


Date	no.	Nom	Prix	Vendeur	Fabricant	Pays	Stock
2015/01/	A1243	Table	356.75	Paul	Plastic Tables inc	USA	23

Organisation des données

- ❖ Le type de base de données le plus connu est nommé *base de données relationnelles*:
 - ❖ Ce type de BD représente les données par des tables. Une table est un tableau avec une propriété associée à chaque colonne. Une colonne correspond à un champs d'une table.
 - ❖ Une table définit une relation entre les champs qui la compose.

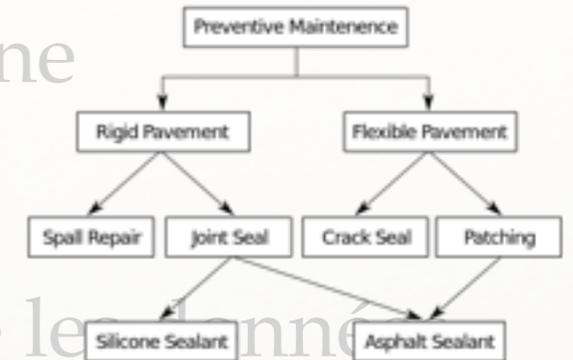
Hierarchical Model



Différents types de BD

- ❖ **Hiérarchique:** Les données sont représentées par une arborescence, un arbre.
- ❖ **Réseau:** Toutes les connexions sont autorisées entre les données. Les données sont représentées par un graphe.
- ❖ **Relationnelle:** E. F. Codd introduit le modèle relationnel dans "A Relational Model of Data for Large Shared Data Bank", CACM 13, No. 6, June 1970 .
- ❖ **À objet:** Les données sont représentées par des objets programmés par objets.
- ❖ **XML:** Système dont les données sont représentées en format XML. XML est un langage de balisage générique qui permet à des textes d'être compréhensibles à l'humain et à l'ordinateur.

Network Model



Object-Oriented Model

Object 1: Maintenance Report Object 1 Instance

Date	01-12-01
Activity Code	24
Route No.	I-95
Daily Production	2.5
Equipment Hours	6.0
Labor Hours	6.0

Object 2: Maintenance Activity

Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

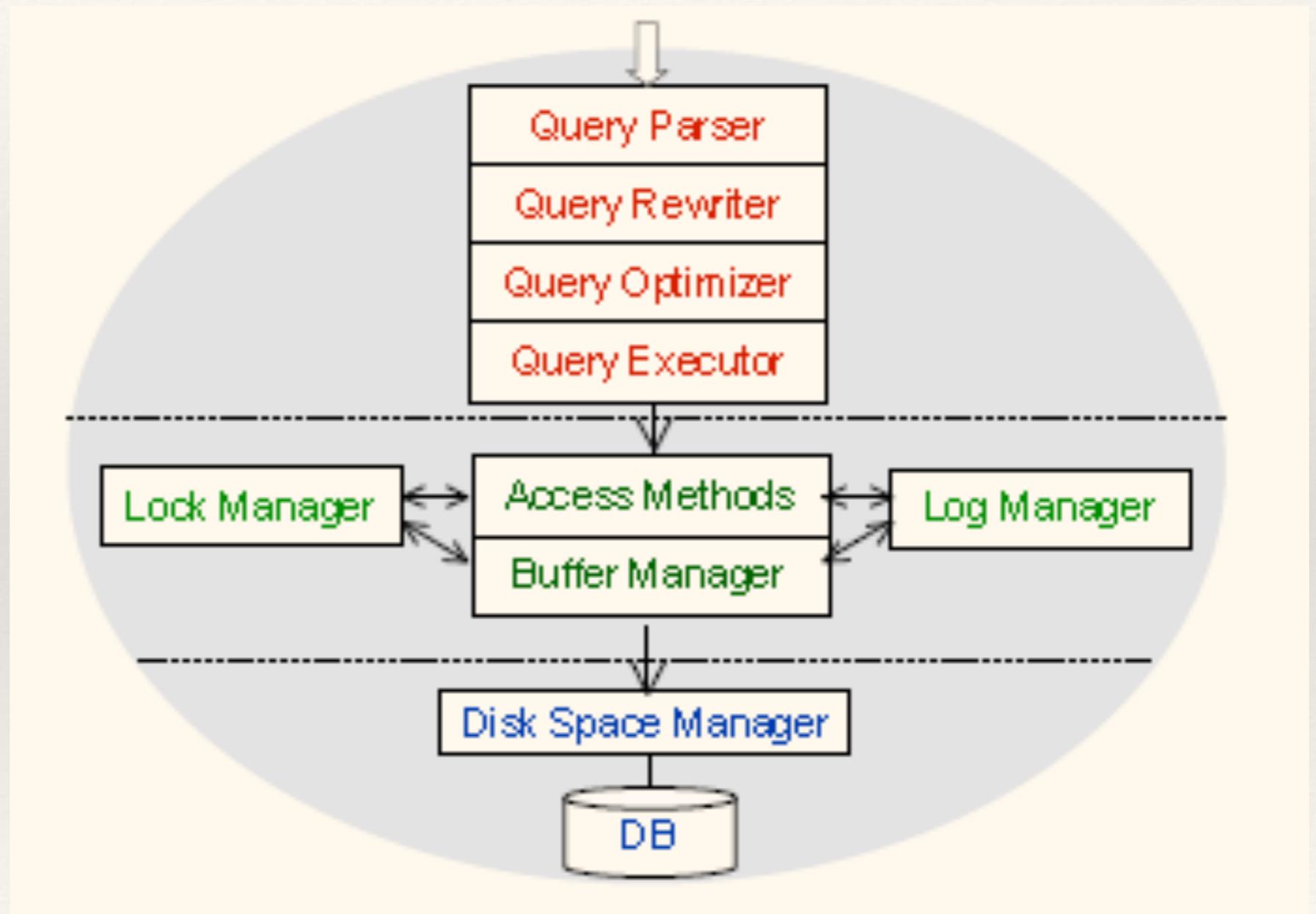
Propriétés ACID des BD

- ❖ **Atomicité:** Chaque transaction (opération) effectuée sur une BD est ou bien exécutée complètement ou bien ignorée.
- ❖ **Consistance:** Chaque transaction transforme une BD consistante en une BD consistante. Chaque valeur écrite dans la BD doit être consistante avec les règles définies par la BD.
- ❖ **Isolation:** L'exécution courante des transactions produit un état du système qui serait obtenu en exécutant chaque transaction en série.
- ❖ **Durabilité:** Lorsqu'une transaction est commise elle le demeurera même s'il y a une coupure de courant, un crash sérieux ou des erreurs du système.

Comment réaliser un DBMS

Architecture pas couche:

- ❖ Traitement des requêtes:
- ❖ Accès aux données:
- ❖ Gestion des disques:



Traitements des requêtes

- ❖ Séquence de traitement d'une requête:
 - ❖ Requête de l'utilisateur:
 - ❖ Déterminer les relations et les attributs de la requête.
 - ❖ Déterminer les fichiers qui contiennent ces éléments.
 - ❖ Accéder aux données dans les fichiers:
 - ❖ Blocage durant l'écriture (pour les propriétés ACID).
 - ❖ Gestion cache, est-ce que les données sont en cache?
 - ❖ Déterminer *les indices* (i.e. *les clés*) des données.
 - ❖ Déterminer à partir des indices les adresses physiques sur les disques qui contiennent ces indices.
 - ❖ Accéder aux données sur le disque:
 - ❖ Lire ou écrire les données sur le disque via les adresses physiques.
 - ❖ Gestion de l'espace sur les disques.

Un peu plus de détails

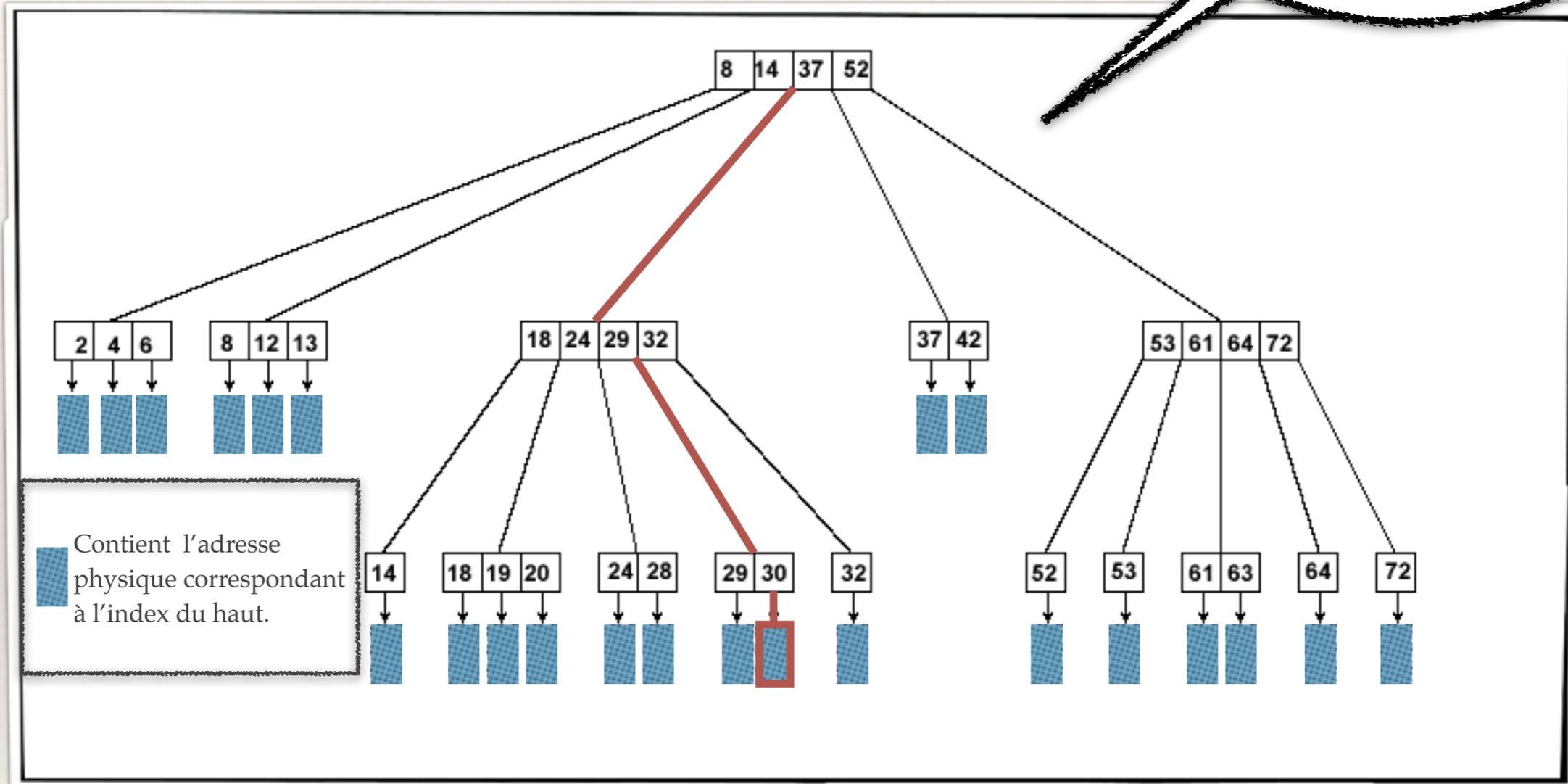
- ❖ La DBSM peut trouver les fichiers en relation à une requête en utilisant un dictionnaire sur le disque qui associe les relations avec les fichiers qui les contiennent.
- ❖ Pour trouver l'adresse sur le disque à partir de laquelle un n-tuplet d'une relation est rangé on procède de cette façon:
 - ❖ À partir de l'index (ou clé) on peut associer une adresse physique en consultant une table qui associe les deux.

AB	AC	AD	AL	...	QRS
0	8	10	2F		90AF

Si la table contient beaucoup d'indices alors la recherche des adresses physiques peut être très inefficace

B-Tree

Cherche en mémoire RAM des enregistrements de tailles 4 et 2 au lieu de charger en mémoire tous les indices en fouillant séquentiellement parmi ceux-ci...



Chercher l'adresse physique qui correspond à l'index 30:

- ❖ 30 est entre 14 et 37: le lien en 14 et 37 est emprunté à la racine,
- ❖ 30 est entre 29 et 32: le lien entre 29 et 32 est emprunté au 2e niveau,
- ❖ 30 est trouvé au 3e niveau et l'adresse physique qui y correspond est obtenue.