

La problématique de la Préservation des Bases de Données

Olivier ROUCHON – CINES

Groupe PIN – Jeudi 7 Octobre



Définition des bases de données

- **Terme du langage courant – base de données**
 - Utilisé pour désigner toute source importante de données.
 - Les bases de données sont classées en fonction de leur contenu : bibliographique, texte, numérique, et images.
- **Terme informatique – base de données**
 - Lot d'informations stockées dans un dispositif informatique.
- **Les technologies existantes permettent d'organiser et de structurer la base de données de manière à pouvoir facilement manipuler le contenu et stocker efficacement de très grandes quantités d'informations.**
- **Le logiciel qui manipule les bases de données est appelé système de gestion de base de données (SGBD).**
 - Il permet d'organiser, de contrôler, de consulter et de modifier la base de données.

Organisation logique des bases de données

- **L'organisation logique des données se fait selon un modèle de données.**

- Les données sont organisées en tableau (ou "matrice"), les colonnes représentant les différents "champs" d'informations
- Les index sont destinés à accélérer les opérations de recherche et de tri.

- **L'organisation logique est indépendante de la structure physique.**

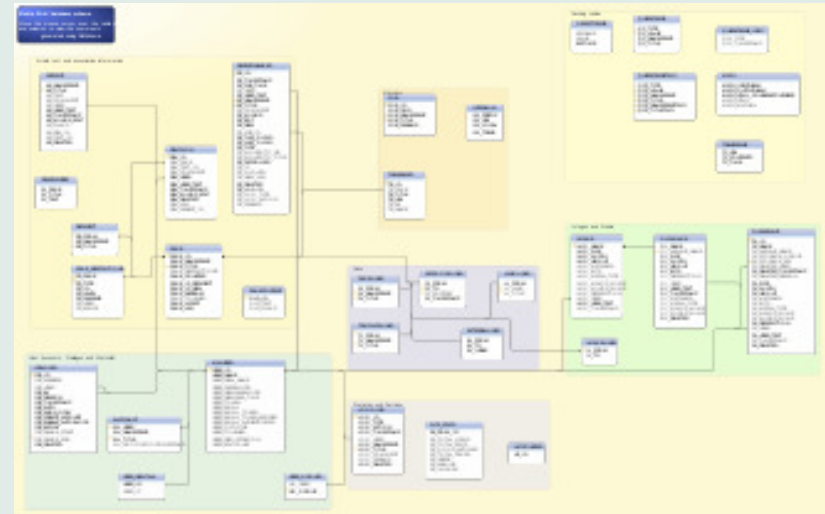
- Une modification de l'organisation logique ne provoque pas de modification de la structure physique et vice-versa.

- **La motivation fondamentale de l'organisation des données est de permettre d'effectuer des inférences (requêtes) sur celles-ci.**

- **Le modèle de données relationnel est aujourd'hui le plus utilisé parce que ce type de représentation permet (contrairement aux modèles hiérarchique et réseau)**

- Les relations multiples, complexes (n:m)

– la résolution de tous types de requêtes



- **Dans les fichiers de la base de données sont stockés**
 - Les informations (les données),
 - Les index (structures permettant d'accélérer la recherche),
 - Les clés primaires (permettant d'identifier une donnée de manière unique),
 - Les clés étrangères (permettant de lier des données entre elles),
 - Le détail du modèle de données (définition des structures de tables, etc.),
 - Les listes de contrôle d'accès,
 - Les procédures stockées (ensemble d'instructions précompilées, stockées et exécutées sur demande par le système qui manipule la base de données),
 - Les déclencheurs (provoquant un traitement particulier en fonction d'événements prédéfinis),
 - Les journaux
- **L'organisation physique d'une base de données diffère selon le matériel et le système d'exploitation pour lequel le système de gestion de base de données est conçu.**

Défis pour l'archivage des bases de données

1. L'organisation physique change selon les éditeurs de système de gestion de bases de données

- Cette organisation n'est pas forcément publiée (SGBD propriétaires)
- La seule solution est d'extraire et préserver l'organisation logique (les tables)

2. Les éléments à préserver sont multiples

- Informations (données).
- Droits d'accès aux données
- Méthodes d'accès aux données

3. Les volumes des bases de données sont potentiellement très grands

Principaux systèmes de gestion de bases de données relationnels

Nom SGBD	Année	Editeur	Utilisation type	Licence
Pick	1968	pick system	serveur	Licence propriétaire
Ingres	1974	Computer Associates^[3]	serveur	GPL
MaxDB	1977	SAP AG	composant logiciel	GPL
dBase	1978	Asht on-Tate	L4G	Licence propriétaire
Oracle Database	1979	Oracle Corporation	serveur	Licence propriétaire
Informix	1981	IBM	serveur	Licence propriétaire
DB2	1983	IBM	serveur	Licence propriétaire
4ème Dimension	1985	4ème Dimension	L4G	Licence propriétaire
PostgreSQL	1985	Michael Stonebraker	serveur	Licence BSD
FileMaker Pro	1985	FileMaker	logiciel applicatif	Licence propriétaire
Berkeley DB	1986	Université de Berkeley	Composant logiciel	Licence BSD
Paradox	1987 ^[4]	Corel^[5]	logiciel applicatif	Licence propriétaire
SQL Server	1989	Microsoft^[6]	serveur	Licence propriétaire
Access	1992	Microsoft	L4G	Licence propriétaire
HyperFile	1993	PCSoft	composant logiciel^[7]	Licence propriétaire
MySQL	1995	Oracle Corporation^[8]	serveur	GPL
Derby	1996	Apache Software Foundation	Composant logiciel	Licence Apache
InterSystems	1997	InterSystems	serveur	Licence propriétaire
Firebird	2000	Firebird Foundation	serveur	Interbase Public Licence
SQLite	2000	D. Richard Hipp	composant logiciel	Domaine public
OpenOffice.org Base	2002	Oracle Corporation^[9]	Logiciel applicatif	LGPL
MariaDB	2009	Monty Program Ab	serveur	GPL