INTÉGRITÉ

DES DONNÉES DANS NOS SYSTÈMES

par Nadine Giasson St-Amand

Présentation des SUJETS

- Contraintes d'intégrité internes
  - Data Definition Language (DDL)
  - Contraintes d'intégrité
  - Déclencheurs
  - Procédures stoquées
- Encodage & casse
- Administration externe
  - Sauvegarde & restauration
  - Versionnement des bases de données
  - Checksum & fingerprint
- Object Relational Mapping (ORM)

CONTRAINTES D'INTÉGRITÉ

INTERNES DE LA BASE DE DONNÉES

Première partie

Le DDL est le language SQL qui permet de créer/modifier/détruire des tables, des colonnes et des clés.

	Table	Colonne	Clés
Créer	CREATE TABLE matable (nomcolonne type, nomcolonne, type);	ALTER TABLE matable ADD colonne type;	ALTER TABLE matable ADD PRIMARY KEY (unetelle)
Modifier	ALTER TABLE matable ...	ALTER TABLE matable ALTER/MODIFY colonne type;	
Détruire	DROP TABLE matable;	ALTER TABLE matable DROP COLUMN colonne;	ALTER TABLE matable DROP PRIMARY KEY;

Data Definition Language

Contraintes d'intégrité

Les contraintes du SQL couvrent la majorité des besoins

Contraintes de colonnes

- NOT NULL - Empêche les valeurs d'être nulles dans cette colonne
- DEFAULT - Spécifie la valeur par défaut
- CHECK - Condition à respecter pour toutes les valeurs de la colonne

Contraintes de table

- UNIQUE - Oblige toutes les valeurs de la colonne à être différentes
- PRIMARY KEY - Valeur unique qui sert d'identifiant à chaque rangée
- FOREIGN KEY - Désigne toujours une rangée d'une autre table
- INDEX - La colonne sera indexée dans la mémoire pour l'accès rapide

Les déclencheurs s'utilisent pour déclencher des ajustements au reste de la base de données lorsque des données sont modifiées. Le but est de maintenir l'intégrité.

- Pour les contraintes d'intégrité :

  (la contrainte) ON INSERT/UPDATE/DELETE CASCADE/SET NULL

- Pour les procédures stockées :

CREATE TRIGGER untel BEFORE/ON/AFTER INSERT/UPDATE/DELETE

Déclencheurs

Les procédures stockées sont de petits bouts de code qui sont programmés directement dans la base de données.

Certaines procédures non-nommées sont directement attachées à leur événement déclencheur (trigger) et ne peuvent être appelées d'un autre point d'exécution. Par exemple :
CREATE TRIGGER apres_ajout_jouet AFTER INSERT ON jouet
BEGIN … END;

Les procédure stockées nommées sont des fonctions nommées programmées à l'intérieur de la base de données. Leur syntaxe n'est pas standard, elle diffère d'un SGBD à l'autre. Leur utilisation était importante dans le passé mais de nos jours elle est réservée au maintient de l'intégrité.

Procédures stoquées

MYSQL
https://dev.mysql.com/doc/refman/5.7/en/trigger-syntax.html
https://openclassrooms.com/courses/administrez-vos-bases-de-donnees-avec-mysql/triggers
(sauf partie sur business logic)

POSTGRESQL
https://www.postgresql.org/docs/9.1/static/sql-createtrigger.html
https://www.postgresql.org/docs/9.2/static/plpgsql-trigger.html
https://docs.postgresql.fr/9.3/sql-createtrigger.html

SQL SERVER
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-trigger-transact-sql
https://stackoverflow.com/questions/10443829/how-to-create-a-before-delete-trigger-in-sql-server
https://msdn.microsoft.com/fr-fr/library/ms189799(v=sql.120).aspx
http://codes-sources.commentcamarche.net/faq/449-ms-sql-server-debuter-avec-les-triggers

Trigger & procédures stoquées

Différents aspects de l'encodage et de la casse doivent être considérés:

- ISO 8859-1 correspondant à notre alphabet occidental versus UTF8
  - Ces encodages ont créé des problèmes d'incompatibilité.
- little indian versus big indian
- les opérations case-sensitive versus case-insensitive
- REGEX = Expressions régulières

Encodage et casse

Il existe plusieurs manières de sauvegarder et restaurer des bases de données.

- Avec les exports SQL et les imports SQL
  - Cette technique peut aussi permettre de faire des transferts de données entre les bases de données
- Avec les fichiers natifs de la base de données sur disque
  - Cette technique donne des résultats assez bons mais parfois des petits problèmes avec certains index surviennent
  - Elle est limitée à la même base de données et de préférence de même version - c'est une solution de secours !

On va aussi parler de sauvegarde à chaud et à froid. Une base de données en production ne peut être arrêtée facilement alors les fichiers sont en utilisation lorsqu'ils apparaissent sur les backup.

Sauvegarde et restauration

Un problème important dans le versionnement des logiciels est le versionnement correspondant des bases de données.

Si chaque révision du logiciel sauvegardée dans le dépôt de code ne présente pas sa structure de base de données correspondante, alors le code ne marchera pas.

- Il est donc important de versionner la structure de la base de données au fur et à mesure de ses changement, DANS LE MEME COMMIT que le code qui utilise les nouveaux éléments de celle-ci

De surcroit, il arrive qu'on ait déjà déployé des bases de données en 'production' et que celles-ci contiennent déjà des données de clients. On ne peut supprimer la base de données et la restaurer sans perdre les données

- On doit donc préparer des différenciels sous forme d'ALTER TABLE

Versionnement de base de données

Les logiciels et les données sont menacés dans leur intégrité par toutes sortes d'attaques. Cependant les données varient également de manière légitime.

- Un CHECKSUM est un calcul mathématique basé sur les bits d'un fichier groupe de fichiers ou système de fichiers qui a peu de chance de donner le même résultat si une erreur existe dans un fichier. Un checksum peut être une somme, une série de XOR, ou un calcul MD5.

- Un FINGERPRINT est la généralisation de l'utilisation de checksum dans la partie constante d'un système, par exemple dans les fichiers de configuration qui ne sont pas supposés changer entre les interventions admin.

Checksum et Fingerprint

Le Object Relational Mapping est un système qui va semi-automatiser les sauvegardes et mise-à-jour des rangées de tables correspondant à des objets en mémoire vive de l'application.

Chaque rangée est un objet métier. Les changements en mémoire se sauvegardent dans l'application, soit automatiquement, soit sur simple indication de se mettre à jour.

Des FrameWork ORM célèbres sont Hibernate en Java, Link en Dot Net et Doctrine en PHP.

Object Relational Mapping (ORM)