Résistance aux pannes dans les Bases de Données

Illustration avec Oracle

J. Akoka - I. Wattiau

Objectifs

Le SGBD doit permettre de :

- minimiser le travail perdu
- assurer un retour à des données cohérentes

A quoi sont dues les pannes ?

- erreur humaine
- erreur de programmation
- défaillance matérielle

Principes

Le SGBD doit fournir un protocole aux applications permettant de :

- faire une transaction
- défaire une transaction
- refaire une transaction

3 moyens à conjuguer :

- la journalisation
- les sauvegardes
- la réplication

En cas de panne :

- on reprend l 'état sauvegardé de la base
- on ré-exécute toutes les actions du journal
- la réplication permet de limiter les interruptions de service

Les concepts : la transaction

- **Propriétés ACID**
  - **Atomicité :**
    - une transaction est une unité d 'exécution indivisible
  - **Cohérence :**
    - respecte les contraintes d 'intégrité
    - les contraintes de type différé sont vérifiées à la validation
  - **Isolation :**
    - l'exécution est indépendante de l'exécution des autres transactions
  - **Durabilité :**
    - le résultat d 'une transaction validée ne doit pas être perdu

Les concepts : la transaction (suite)

Atomicité	**Résistance aux pannes**
Cohérence	**Contrôle d'intégrité**
Isolation	**Contrôle de concurrence**
Durabilité	**Résistance aux pannes**

Les concepts : mémoire sûre et point de reprise

Mémoire sûre :

mémoire découpée en pages dans laquelle une écriture de page est soit correctement exécutée, soit non exécutée.

↖ ***doubles écritures***

***Point de reprise système* :**

état du système sauvegardé sur mémoires secondaires à partir duquel il est possible de repartir après un arrêt.

Méthodes de validation des transactions (1)

- ***Mise à jour différée :***
  - ***technique des pages d 'ombre***
  - ***basculement de la table des pages***
- ***Mise à jour immédiate (Oracle) :***
  - ***les modifications sont d'abord journalisées***
  - ***puis répercutées dans la BD***

Méthodes de validation des transactions (2)

- ***Technique des pages d'ombre***

  les écritures ne sont pas exécutées en place dans la base mais dans des pages nouvelles séparées propres à la transaction, appelées pages différentielles => avant toute lecture, le SGBD consulte les pages différentielles

  => problèmes de performances

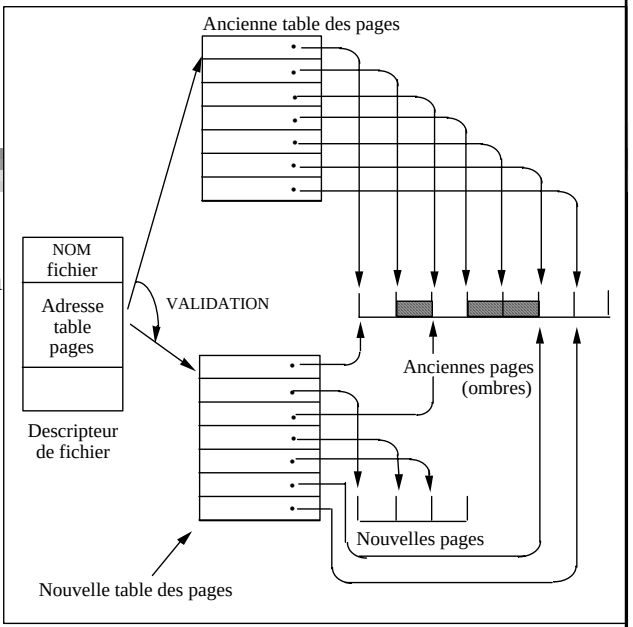
Méthodes de validation des transactions (3)

- ***Basculement des tables des pages***

  Le SGBD gère deux tables des pages qui pointent sur les deux versions des données.

  A la validation, on supprime l'ancienne table des pages.

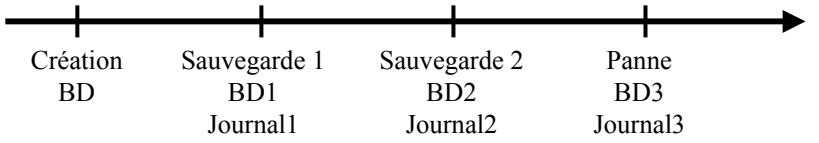
Permet de ne valider qu ’en fin de transaction
…
tout en visualisant les données modifiées



Les procédures de reprise

- normale
- après une panne du système : reprise à chaud
  - on a perdu la mémoire centrale mais pas la mémoire secondaire
- après une panne de mémoire secondaire : reprise à froid
  - perte de mémoire secondaire
  - principe :
    - reprendre les sauvegardes sur bande
    - utiliser le journal s'il est disponible
- panne catastrophique (!)

Exemple de situation



- Reprise à chaud : on réapplique le journal 3 sur BD3 pour défaire les transactions non validées
- Reprise à froid : si BD2 est endommagée, on reprend BD1 et on réapplique Journal2 et Journal3

La réplication

- Permet de synchroniser le contenu d'un espace de stockage avec un réplicat
- Modifications en double
- pour assurer la continuité de service par bascule

Conseils

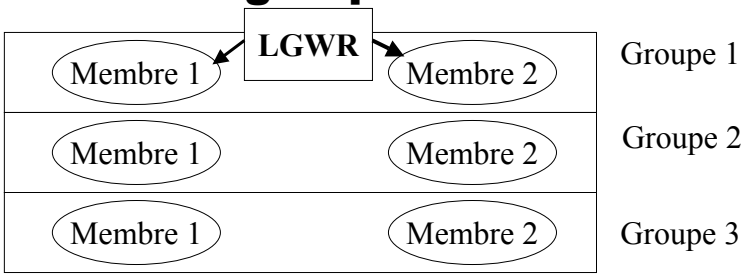
- Localiser les deux copies sur des disques différents
- répliquer au moins le journal des données sensibles
- mieux : répliquer les méta-données, les données et les journaux

La réplication dans Oracle

- Multiplexage : écriture simultanée
- Concerne principalement :
  - le fichier de contrôle de la BD
  - le journal (redo log)
- Principe : un ensemble de fichiers membres de groupes
  - au minimum deux groupes avec chacun un membre

Principe du multiplexage

- Le processus de journalisation écrit « en double » dans les membres d 'un même groupe



Fonctionnement des groupes

- Quand les fichiers du groupe i sont saturés :
  - ils peuvent être archivés par le processus d'archivage
  - la journalisation se poursuit dans un autre groupe, et ce de façon circulaire
- Si pb dans l 'écriture du membre i
  - message d 'erreur
  - les écritures se poursuivent dans les autres membres
  - si tous sont en panne, l 'instance est arrêtée

Contenu du journal

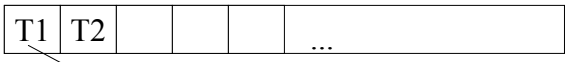
- Pour chaque transaction T
  - (id-transaction, début)
  - (id-transaction,écriture,donnée concernée,ancienne valeur, nouvelle valeur)
  - (id-transaction,lecture, donnée concernée)
  - (id-transaction,commit)
- + Points de contrôle : marquent les écritures physiques (cache -> disque)

Segments d'annulation

- **Contiennent les valeurs des données avant leur modification pour les transactions non validées**
- **permettent :**
  - **les lectures par les transactions concurrentes**
  - **l'annulation des effets d 'une transaction**
  - **une reprise sur la base**

Contenu d'un fichier d'annulation

Table des transactions



(idfichier,idbloc,donnée,valeur avant modif)
= des entrées d'annulation
chaînées entre elles par transaction

Types de segments d'annulation

- **Rollback segment SYSTEM**
  - **c 'est le minimum**
  - **ne peut être supprimé**
- **Privé ou public**
  - **l 'utilisation dépend de la stratégie**

Stratégie Oracle

- **Règle 1 : si une instance est seule à accéder à la BD, elle acquiert le segment SYSTEM, sinon elle acquiert SYSTEM et au moins un autre segment d'annulation**
- **Règle 2 : une instance essaie d 'obtenir « son dû » :**

$$\frac{\text{transactions}}{\text{transactions\_per\_rollback\_segments}}$$

Nb max de transactions concurrentes

Nb max de transactions à associer à un segment d'annulation

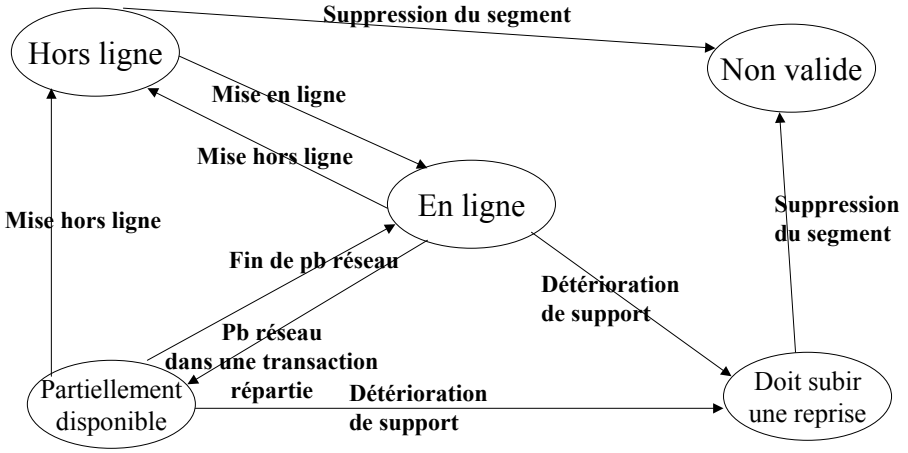
Stratégie Oracle (suite)

- **Règle 3 : après acquisition du segment d'annulation SYSTEM, une instance essaie d'obtenir tous les segments d 'annulation privés spécifiés**
- **Règle 4 : en cas de besoin, elle essaie d'acquérir des segments publics**
- **Règle 5 : un segment détenu par une instance n'est plus disponible pour d'autres sauf si :**
  - **il est mis hors ligne**
  - **l'instance qui le détient est arrêtée**

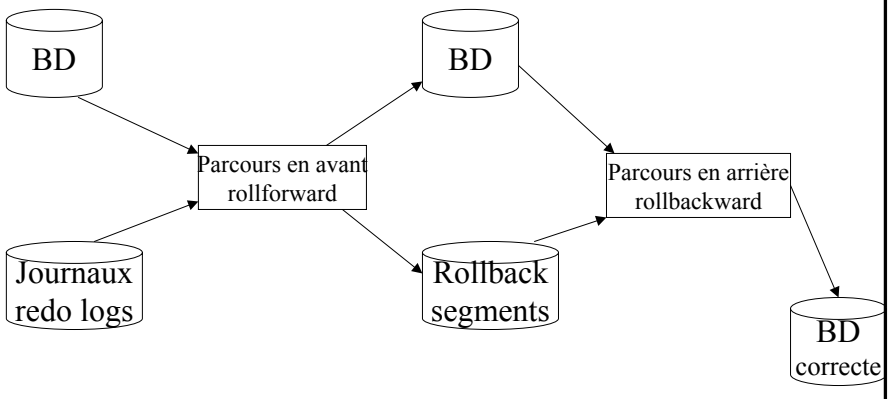
Affectation d'une transaction à un segment d'annulation

- **Établie en début de transaction**
  - **lors de la 1e instruction de m.a.j ou de LDD**
  - **soit par le système (par défaut) soit par le programmeur**
- **Oracle essaie de distribuer équitablement les transactions sur les segments**

Diagramme d ’états d ’un segment



Processus de reprise (recovery)



Reprise automatique

- À partir des journaux en ligne
- en cas de :
  - échec d'un processus utilisateur
  - échec d'une instance Oracle
  - panne du système hôte

Reprise sur support (media recovery)

- A partir des sauvegardes
  - en cas de détérioration d 'un support

Conclusion

- **Définir une « bonne » politique de sauvegarde**
- **Aux solutions du SGBD, s'ajoutent les solutions du système hôte**