

# SGBD-R : Les transactions

## Contrôle de concurrence

---

Guillaume Raschia — Nantes Université

Dernière mise-à-jour : 2 mars 2023

originaux de Philippe Rigaux, CNAM

# Plan de la session

Isolation par versionnement (S10.1)

Sérialisabilité (S10.2)

Contrôle de concurrence multi-versions (S10.3)

2PL (S10.4)

## Isolation par versionnement (S10.1)

---

Toute transaction  $T$  en cours a deux choix à chaque instant :

- valider les maj effectuées avec **commit**
- les annuler avec **rollback**.

Le SGBD maintient, pendant l'exécution de  $T$ , deux versions des nuplets mis à jour :


- une version du nuplet **après** la mise à jour;
- une version du nuplet **avant** la mise à jour.

On parle **d'image après** et **d'image avant** pour ces deux versions.

# Isolation fondée sur les versions

Soit deux transactions  $T$  et  $T'$ .

- Chaque fois que  $T$  met à jour un nuplet, la version courante est copiée dans **l'image avant**, puis remplacée par la valeur de **l'image après** fournie par  $T$ .
- Quand  $T$  lit des nuplets, le système doit lire dans l'image après pour assurer une vision cohérente de la base
- En revanche, une autre transaction  $T'$  doit lire dans **l'image avant** pour éviter les effets de lectures sales.

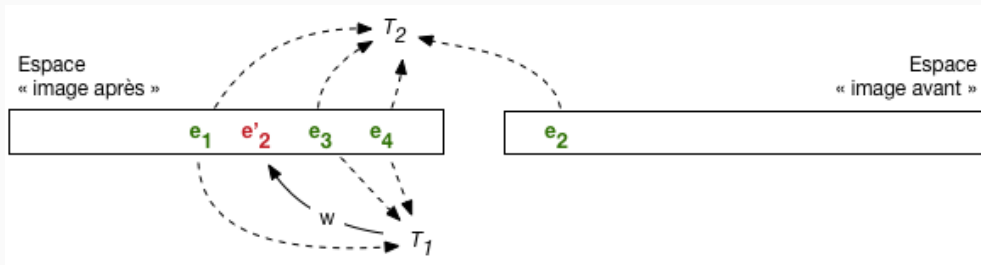
 **Combien d'images avant / après?** Une seule mise à jour est autorisée à la fois<sup>1</sup>, donc une seule paire (image avant, image après).

---

1. pas d'écriture sale.

# Illustration

Le nuplet  $e_2$  est en cours de modification.



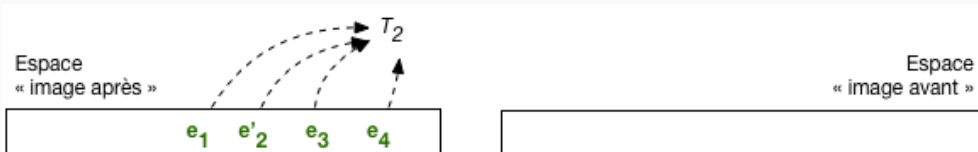
$T_1$  lit  $e'_2$  après sa maj, et  $T_2$  lit  $e_2$  : elles ne lisent pas la même version.

L'image après peut être interprétée comme l'état courant de la base, éventuellement non approuvé.

# Les lectures ne sont pas répétables

Le mode précédent est correct pour le niveau `read committed`.

Mais les lectures ne sont pas répétables : une fois que  $T_1$  est approuvée,  $T_2$  lit la dernière version validée de  $e_2$ .



## Image avant et **repeatable read**

Les images avant constituent un « cliché » de la base pris à un moment donné.

**Principe des lectures répétables** : toute transaction ne lit que dans le cliché valide au moment où elle a débuté.

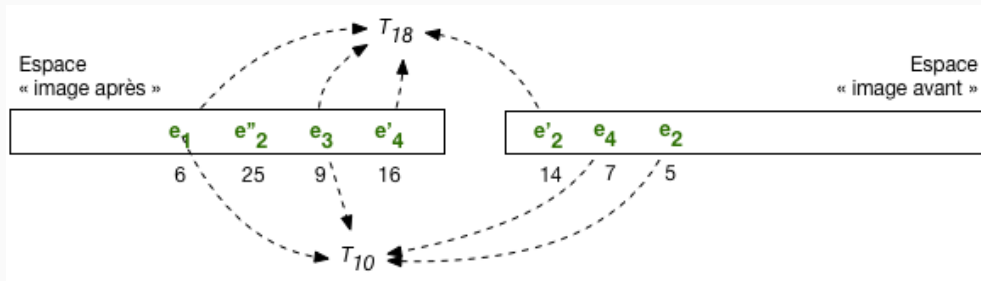
**Horodatage** :

- Toute transaction  $T_i$  est « estampillée » par le moment  $\tau_i$  où elle a commencé.
- Chaque nuplet  $e$  est marqué par le moment  $\tau_e$  de sa validation (**commit**).
- $T_i$  ne lit que les nuplets dont l'estampille  $\tau_e < \tau_i$  est immédiatement antérieure à la sienne.

Ce mécanisme de **versionnement** assure la cohérence de toutes les lectures.



## Le niveau repeatable read



**Combien de versions?** Il faut maintenir le « cliché » de la base pris pour la transaction la plus ancienne en cours d'exécution.

Notion d'image avant et image après

- **Image après** : les dernières versions de chaque nuplet; certaines ne sont pas encore validées.
- **Image avant** : toutes les versions antérieures

Le multi-versions assure les niveaux **read committed** et **repeatable read**.

- **read committed** : il suffit de conserver une version dans l'image avant
- **repeatable read** : il faut conserver les versions immédiatement antérieures au début de la transaction (« cliché » = *snapshot*)

Niveau d'isolation plus élevé = performances affectées

## Sérialisabilité (S10.2)

---

## Sérialisabilité (à connaître et—surtout—comprendre)

Une exécution concurrente  $H$  de  $n$  transactions  $T_1, \dots, T_n$  est **sérialisable** si et seulement si le résultat de l'exécution de  $H$  est le même que celui d'**une** exécution **en série** de  $T_1, \dots, T_n$ .

Bien comprendre : si j'ai deux transactions  $T_1$  et  $T_2$ , leur exécution imbriquée est sérialisable ssi équivalente à  $T_1$  puis  $T_2$ , ou à  $T_2$  puis  $T_1$  (**en série**).

En d'autres termes, on compare  $H$  à une exécution **en isolation complète**.

## Caractérisation de la sérialisabilité : les conflits

La définition précédente est **déclarative** : elle dit ce qu'est la sérialisabilité, pas la manière de la vérifier en pratique.

Nous avons besoin d'une caractérisation plus opérationnelle.

### Opérations conflictuelles (à connaître)

Deux opérations  $p_i(x)$  et  $q_j(y)$  sont **en conflit** ssi  $x = y$ ,  $i \neq j$ ,  $p$  ou  $q$  est une écriture.

En clair : deux transactions distinctes accèdent au même nuplet, et—au moins—une veut le modifier.

## Exemple

Reprenons une nouvelle fois l'exemple des mises à jour perdues :

$r_1(s); r_1(c_1); r_2(s); r_2(c_2); w_2(s); w_2(c_2); w_1(s); w_1(c_1)$

- $r_1(s)$  et  $w_2(s)$  sont en conflit ;
- $r_2(s)$  et  $w_1(s)$  sont en conflit ;
- $w_2(s)$  et  $w_1(s)$  sont en conflit.

$r_1(s)$  et  $r_2(s)$  **ne sont pas** en conflit (lectures); pas de conflit sur  $c_1$  et  $c_2$ ; **pas de conflit** entre  $r_1(s)$  et  $w_1(s)$ .

 Quid des conflits ici ?

$r_1(c_1); r_1(c_2); r_2(s); r_2(c_2); w_2(s); w_2(c_2); r_1(s)$

## Relation $\triangleleft$ de préséance (à connaître)

Soit  $H$  une exécution concurrente des transactions  $\mathcal{T} = \{T_1, T_2, \dots, T_n\}$ ;

Il existe une relation  $\triangleleft$  sur cet ensemble, définie par :

$$T_i \triangleleft T_j \Leftrightarrow \exists p_i \in T_i, q_j \in T_j, p_i \text{ est en conflit avec } q_j \text{ et } p_i <_H q_j$$

où  $p_i <_H q_j$  indique que  $p_i$  apparaît avant  $q_j$  dans  $H$ .

Dans l'exemple précédent : on a  $T_1 \triangleleft T_2$ , ainsi que  $T_2 \triangleleft T_1$ .

## Condition de sérialisabilité

Une définition opérationnelle de la sérialisabilité s'exprime à partir du graphe de la relation  $(\mathcal{T}, \triangleleft)$ , dit **graphe de sérialisation**.

Les nœuds du graphe sont les transactions  $T_i$  et chaque arc  $T_i \rightarrow T_j$  dessine une relation  $T_i \triangleleft T_j$ .

**Sérialisabilité par conflit** : une condition suffisante de sérialisabilité

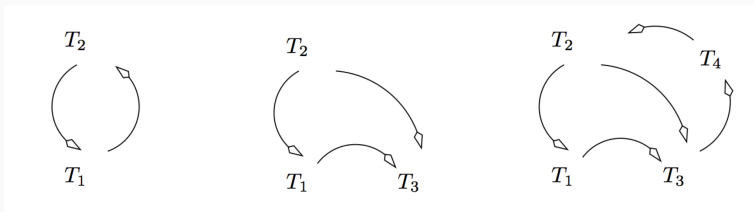
Soit  $H$  une exécution concurrente d'un ensemble de transactions  $\mathcal{T}$ . Si le graphe de  $(\mathcal{T}, \triangleleft)$  est **acyclique**, alors  $H$  est **sérialisable**.

Autrement dit, la sérialisabilité par conflit impose que  $(\mathcal{T}, \triangleleft^*)$  soit une relation d'ordre partiel (antisymétrie).



# Graphes de sérialisabilité

✎ Lesquels correspondent à une exécution sérialisable ?



**Important** : un algorithme de contrôle de concurrence doit s'assurer qu'aucun cycle ne puisse apparaître dans ce graphe.

✎ Trouver une exécution à la fois sérialisable et non sérialisable par conflit !

Caractérisation « sémantique » de la sérialisabilité : équivalence avec une exécution en série, pour au moins un ordonnancement des transactions impliquées.

Caractérisation « syntaxique » de la sérialisabilité : pas de cycle dans le graphe de sérialisation construit sur les conflits entre opérations.

**Objectif d'un algorithme de contrôle** s'assurer que toutes les exécutions sont sérialisables.

- en surveillant le graphe de sérialisation et en rejetant une transaction si un cycle apparaît (variante « optimiste »)
- en tentant de prévenir l'apparition de cycles (variante « pessimiste »)

## Contrôle de concurrence multi-versions (S10.3)

---

# Un algorithme de contrôle de concurrence multi-versions

Connu sous le nom de *Snapshot Isolation*. Très utilisé.

Tire parti du versionnement, qui limite les opportunités de conflits.

Très léger : aucun verrouillage en lecture, un contrôle sur les écritures.

**Mais** : ne garantit pas la sérialisabilité stricte...

## Souvenons-nous des versions

Un système qui fournit un niveau **repeatable read** s'appuie sur du versionnement par horodatage.

Dans ce cas les lectures se font sur l'état de la base **avant** le début de la transaction.

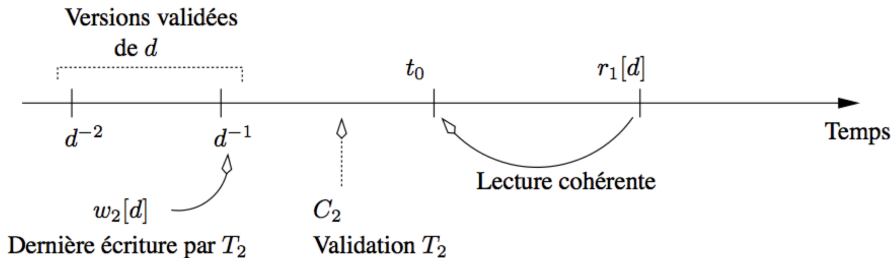
Deux possibilités de conflit entre une **lecture** de  $T_1$  et une tx  $T_2$ .

- $r_1(d)$  est en conflit avec une écriture  $w_2(d)$  qui a eu lieu **avant**  $\tau_1$ ;
- $r_1(d)$  est en conflit avec une écriture  $w_2(d)$  qui a eu lieu **après**  $\tau_1$ .

Approfondissons.

## Premier cas de conflit

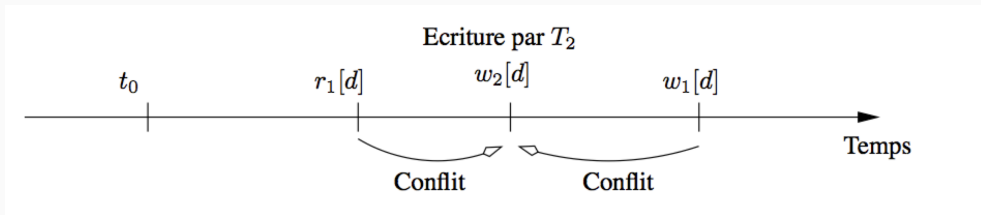
Cas 1 :  $r_1(d)$  en conflit avec  $w_2(d)$  avant  $\tau_1$ . Alors  $T_2$  est approuvée. Pas de risque.



En fait, cela traduit une exécution sérielle de  $T_2$  puis  $T_1$ .

## Second cas de conflit

Cas 2 :  $r_1(d)$  en conflit avec  $w_2(d)$  après  $\tau_1$ .



Si  $T_1$  cherche à écrire  $d$  après l'écriture  $w_2(d)$ , un conflit cyclique apparaît.  
L'algorithme doit surveiller ce cas.

# Le contrôle de concurrence multi-versions

## Le principe

On vérifie, au moment d'exécuter  $w_1(d)$ , qu'aucune transaction  $T_2$  n'a modifié  $d$  entre le début de  $T_1$  et l'instant présent.

En cas d'écriture  $w_1(d)$ <sup>2</sup>,

- si  $\tau_d \leq \tau_1$  et  $d$  non verrouillé :  $T_1$  verrouille (exclusif)  $d$ , et effectue  $w_1(d)$ ;
- si  $\tau_d \leq \tau_1$  et  $d$  verrouillé :  $T_1$  est mise en attente ;
- si  $\tau_d > \tau_1$ ,  $T_1$  est rejetée.

**Estampillage temporel.** Au moment de  $C_1$  tous les nuplets modifiés par  $T_1$  obtiennent une nouvelle version avec pour estampille l'instant du **commit**.

---

2. Règle du **First-Updater-Wins** (alt. : **First-Committer-Wins**).



## Exemple

Maj perdues :  $r_1(s); r_1(c_1); r_2(s); r_2(c_2); w_2(s); w_2(c_2); C_2; w_1(s); w_1(c_1); C_1$ .

On prend  $\tau_1 = 100, \tau_2 = 120$ .

- $T_1$  lit  $s$ ,  $T_1$  lit  $c_1$ ,  $T_2$  lit  $s$ ,  $T_2$  lit  $c_2$ ; aucun verrou.
- $T_2$  veut modifier  $s$  : l'estampille de  $s$  est inférieure à  $\tau_2 = 120$  :  $s$  n'a pas été modifié; on pose un verrou exclusif sur  $s$  et on effectue  $w_2(s)$  :
- $T_2$  modifie  $c_2$ , avec pose d'un verrou exclusif;
- $T_2$  valide, relâche les verrous; versions de  $s$  et  $c_2$  avec l'estampille 150;
- $T_1$  veut à son tour modifier  $s$  : version de  $s$  avec  $\tau_s > \tau_1 = 100$ .  $T_1$  est rejetée.

## Écriture biaisée : un cas de non-sérialisabilité

L'algorithme *Snapshot Isolation* dans certaines exécutions non sérialisables.

Exemple : une simple copie d'une ligne à une autre :

```
function copie (id1, id2)
begin
  val := select valeur from T where id = id1
  update T set valeur = :val where id = id2
  commit
end
```

Exécution concurrente de **copie**( $a, b$ ) et **copie**( $b, a$ ) :  $r_1(a); r_2(b); w_1(b); w_2(a)$

- L'exécution montre un conflit cyclique, elle n'est pas sérialisable;
  - une anomalie de type « écriture biaisée » (*Write Skew*) se produit.
- Elle est néanmoins autorisée par le contrôle multi-versions SI!

*Snapshot Isolation* est un algorithme très simple à mettre en œuvre, utilisé par exemple dans des contextes distribués (applications web).

Très fluide, très peu de verrouillage, contrôle bien le cas des mises à jour perdues.

Ne garantit pas la sérialisabilité stricte (des améliorations ont été proposées).

2PL (S10.4)

---

## Verrouillage à deux phases

Le principal (et le plus ancien) algorithme pour assurer la sérialisabilité.

Est réputé engendrer beaucoup de blocages et de rejets.

Repose sur le verrouillage (quelle surprise!) des nuplets.

Pose de **verrous** sur les nuplets.

- Une lecture ( $r$ ) requiert un **verrou partagé** (*shared lock* ou  $s$ ), compatible avec la pose d'autres verrous partagés sur le même nuplet.
- Une écriture ( $w$ ) requiert un **verrou exclusif** (*exclusive lock* ou  $x$ ), qui interdit la pose de tout autre verrou, exclusif ou partagé, et donc de toute lecture ou écriture par une autre transaction.

Verrous = **blocages**. Il faut donc en poser :

- **le moins possible**, surtout pour les verrous exclusifs;
- **pour la durée la plus courte possible**.

# Transactions et verrous

Le **contrôleur** (ou programmeur) pose des verrous pour une transaction  $T_i$ , à réception de chacune des opérations  $r_i(a)$  et  $w_i(a)$ .

- On ne peut poser un **verrou partagé**  $s_i(a)$  sur un nuplet  $a$  que s'il n'y a que des verrous partagés sur ce nuplet.
- On ne peut poser un **verrou exclusif**  $x_i(a)$  que si
  - il n'y a aucun autre verrou, **ou**,
  - il y a un verrou partagé  $s_i(a)$  déjà posé par  $T_i$  elle-même (*upgrade* ou *promotion*)

Une transaction qui n'obtient pas un verrou, **est mise en attente**.

# Un protocole de verrouillage

On pose des verrous pour empêcher l'apparition de cycle dans le graphe de sérialisation.

- **Lecture** sur  $a$  par  $T$  : on regarde s'il y a un verrou **exclusif** sur  $a$  ;
  - si oui la transaction  $T$  est mise en attente.
  - si non, la transaction pose un verrou partagé et l'opération est exécutée.
- **Ecriture** sur  $a$  par  $T$  : on regarde s'il y a un verrou **quelconque** sur  $a$ , autre qu'un verrou partagé de la tx  $T$  elle-même ;
  - si oui la transaction  $T$  est mise en attente.
  - si non, la transaction pose un verrou exclusif (ou promet son verrou partagé en verrou exclusif) et l'opération est exécutée.

 Est-ce suffisant pour garantir la sérialisabilité ?



## Vers le verrouillage à deux phases : 2-Phase Locking ou 2PL

Le contrôleur reçoit l'histoire  $H = r_1(a); w_2(a); w_2(b); w_1(b)$  non sérialisable.

Au fur et à mesure, il pose et libère (action  $\ell$ ) les verrous :

$s_1(a); r_1(a); \ell_1(a);$

$x_2(a); w_2(a); \ell_2(a);$

$x_2(b); w_2(b); \ell_2(b);$

$x_1(b); w_1(b); \ell_1(b)$

$H$  n'est cependant toujours pas sérialisable !

La règle 2PL

Chaque tx pose tous ses verrous d'abord (phase 1), et les libère ensuite (phase 2).

## Deux variantes de 2PL

### 2PL strict (ou S2PL)

Les verrous exclusifs ne sont relâchés qu'au `commit` ou `rollback` de la tx.

### 2PL rigoureux (ou SS2PL)

Tous les verrous (exclusifs et partagés) ne sont relâchés qu'au `commit` ou `rollback` de la tx.

**En pratique** : C'est le protocole 2PL rigoureux (*Strong Strict 2PL* ou *Rigorous 2PL*) qui est mis en œuvre.

## Exemple d'exécution avec SS2PL

Prenons l'exécution concurrente :  $H = r_1(a); w_2(a); w_2(b); C_2; w_1(b); C_1$

- $T_1$  pose un verrou partagé sur  $a$ , lit  $a$  mais ne relâche pas le verrou;
- $T_2$  tente de poser un verrou exclusif sur  $a$  : impossible puisque  $T_1$  détient un verrou partagé, *donc  $T_2$  est mise en attente*;
- $T_1$  pose un verrou exclusif sur  $b$ , modifie  $b$ , valide; ses verrous sont relâchés;
- $T_2$  est libérée : elle pose un verrou exclusif sur  $a$ , et le modifie;
- $T_2$  pose un verrou exclusif sur  $b$ , et modifie  $b$ ;
- $T_2$  valide, ce qui relâche les verrous sur  $a$  et  $b$ .

Exécution **après réordonnancement** :  $H' = r_1(a); w_1(b); C_1; w_2(a); w_2(b); C_2$

# Un gros problème : les deadlock

Reprenons notre exemple des mises à jour perdues.

$$r_1(s); r_1(c_1); r_2(s); r_2(c_2); w_2(s); w_2(c_2); C_2; w_1(s); w_1(c_1); C_1$$

- $T_1$  lit  $s$  et  $c_1$ , qui sont verrouillés en lecture.
- $T_2$  lit  $s$  et  $c_2$ , qui sont verrouillés en lecture ;  $s$  partage deux verrous en lecture
- $T_2$  veut écrire  $s$  : conflit, donc blocage de  $T_2$ .
- $T_1$  veut écrire  $s$  : conflit, donc blocage de  $T_1$ .

C'est l'étreinte fatale ou interblocage, ou encore verrou mortel (*deadlock*)!!

Le système va rejeter une des transactions. Très regrettable, mais toujours mieux que d'introduire des anomalies (?)

Construction du **graphe d'attentes** :

- les nœuds sont les tx;
- les arcs matérialisent l'attente par la tx source d'une ressource détenue par la tx cible.

*deadlock*

**Détection de cycle** dans le graphe d'attentes !

## Quelle tx rejeter après un deadlock ?

Stratégie fondée sur l'estampillage temporel des tx.

$T_1$ ,  $T_2$  et  $T_3$  aux temps 1, 2 et 3.  $T_2$  possède un objet convoité par  $T_1$  ou par  $T_3$ .

Il existe 2 variantes :

1. *wound-wait* : la plus ancienne tx est autorisée à préempter le nuplet, la plus récente est mise en attente.
  - $T_1$  prend l'objet et  $T_2$  est annulée;
  - $T_3$  attend que  $T_2$  libère l'objet.
2. *wait-die* : la plus ancienne tx est autorisée à attendre, la plus récente est immédiatement rejetée.
  - $T_1$  attend que  $T_2$  libère l'objet;
  - $T_3$  est annulée.

 Cas de famine ? évité en rejouant les tx avec leur estampille d'origine.

- verrous de mise-à-jour; clause **for update**
- verrous d'incrément, etc.
- protocole arborescent et verrous de prédicats
- multi-granularité, escalade de verrous et verrous d'intention

## Un point sur les plans

- 2PL : protocole de verrouillage à 2 phases
- S2PL : 2PL strict
- SS2PL : 2PL rigoureux
- SI : isolation par cliché (famille MVCC - optimiste)
- REP : réparable
- PAC : prévention des annulations en chaîne
- CSER : sérialisable par conflit
- SER : sérialisable
- SERIE : sériel

 Construire un plan SER et non 2PL (à l'aide d'« écritures aveugles »).



Le **SS2PL** est le seul algorithme actuellement utilisé pour **garantir la sérialisabilité**.

Il repose intensivement sur la **pose de verrous** qui ne sont relâchés qu'à la fin de la transaction.

Il entraîne des **interblocages**, et donc le rejet de certaines transactions : difficilement explicable pour un utilisateur, pose le problème de resoumettre la transaction.