4. Transaction management

12 Architecture of the Transaction Management
4. Transaction management

12 Architecture of the Transaction Management

13 Locking Methods
4. Transaction management

12 Architecture of the Transaction Management

13 Locking Methods

14 Non-locking Methods
4. Transaction management

12 Architecture of the Transaction Management

13 Locking Methods

14 Non-locking Methods

15 Multiversion concurrency control
4. Transaction management

12 Architecture of the Transaction Management

13 Locking Methods

14 Non-locking Methods

15 Multiversion concurrency control

16 Transactions in SQL
Schedulers

![Scheduler Diagram]

- Transaction Manager (TM)
- Scheduler (SC)
- Storage Manager (SM)
- Recovery Manager (RM)
- Buffer Manager (BM)
- DB

Correct schedules (consisting of r, w, c, a)
Transaction Operations

- Transaction brackets
  - **BOT** (*Begin of Transaction*)
  - **EOT** (*End of Transaction*)

- Steps:
  - \( r \) (*read*),
  - \( w \) (*write*),
  - \( a \) (*abort*), and
  - \( c \) (*commit*).

\[
t = \{r|w\}^*(a|c)
\]
States of a transaction

- **initial**
- **running**
- **delayed**
- **stopped**
- **aborted**

- **BOT**
- **retry**
- **stop**
- **execute**
- **restart**
- **delay**
- **EOT**
- **committed**
Treatment of a step

- **Execute** transaction $\rightarrow$ state *running*
- **Delay** transaction $\rightarrow$ state *delayed*
- **Reject** transaction $\rightarrow$ state *aborted*
Aggressive Schedulers

- A scheduler is **aggressive**, if it allows conflicts and later attempts to identify and resolve conflicts.
- Maximizes parallelism of transactions.
- Risk: Transactions are reset at the end of an execution.
- Also known as **optimistic** methods.
Conservative Schedulers

- A scheduler is **conservative**, if it avoids conflicts and in return accepts possible delays of transactions.
- Allows only low parallelism of transactions.
- Minimizes the effort of resetting aborted transactions.
- In extreme cases, no parallelism of transactions takes place. This means that all transactions except for one will be delayed.
- Also known as **pessimistic** methods.
Lock Models

- Write and read locks in the following notation:
  - $rl(x)$: read lock on object $x$
  - $wl(x)$: write lock on object $x$

- Unlock $ru(x)$ and $wu(x)$, often summarized as unlock $u(x)$
Locking rules

- Write access $w(x)$ only possible after setting write lock $wl(x)$
- Read access $r(x)$ only valid after $rl(x)$ or $wl(x)$
- Lock only objects that have not yet been locked by another transaction
- After $rl(x)$ only $wl(x)$ valid, no lock on $x$ after that; locks of the same kind are used only once
- After $u(x)$ by $t_i$, $t_i$ may not perform another $rl(x)$ or $wl(x)$
- Before a commit, all locks must be removed
Deadlocks

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$wl(x)$</td>
<td></td>
<td>$wl(y)$</td>
</tr>
<tr>
<td>$wl(y)$</td>
<td>$wl(x)$</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>$wl(x)$</td>
<td></td>
</tr>
</tbody>
</table>

Deadlock!

Alternatives:

- Identify and remove deadlocks
- Avoid deadlocks in the first place
Identification and Removal

- Waiting Graph

- Solving by aborting a transaction, Criteria:
  - Number of resolved loops,
  - Length of a transaction,
  - Effort of resetting a transaction,
  - Importance of a transaction, ...
Livelock Problem

1. $T_1$ locks $A$
2. $T_2$ attempts to lock $A$, but has to wait
3. $T_3$ attempts to lock $A$, but has to wait as well
4. $T_1$ unlocks $A$
5. $T_3$ gets a time slice before $T_2$ locks $A$
6. $T_2$ still attempts to lock $A$, but has to wait
7. $T_4$ attempts to lock $A$, afterwards but has to wait as well
8. $T_3$ unlocks $A$
9. $T_4$ gets the next time slice before $T_2$ ...

## Locking Protocols: Necessity

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$wl(x)$</td>
<td>$wl(x)$</td>
</tr>
<tr>
<td>$w(x)$</td>
<td>$w(x)$</td>
</tr>
<tr>
<td>$u(x)$</td>
<td>$u(x)$</td>
</tr>
<tr>
<td>$wl(y)$</td>
<td>$wl(y)$</td>
</tr>
<tr>
<td>$w(y)$</td>
<td>$w(y)$</td>
</tr>
<tr>
<td>$u(y)$</td>
<td>$u(y)$</td>
</tr>
</tbody>
</table>
Two-phase Locking Protocol

2PL guarantees conflict serializability!
Conflict caused by non-compliance with 2PL

<table>
<thead>
<tr>
<th></th>
<th>$T_1$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u(x)$</td>
<td></td>
<td>$wl(x)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$wl(y)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$wl(y)$</td>
<td></td>
<td>$u(x)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$u(y)$</td>
</tr>
</tbody>
</table>

T1

T2
Strict Two-phase Locking Protocol

- Expanding phase
- Shrinking phase

Avoids cascading aborts!
Conservative 2PL Protocol

Avoids deadlocks!
Lock Granularity

- Granularity hierarchies in databases
- Hierarchical locks
- Tree protocols for tree indexes
Granularity Hierarchies in RDBS

Granularity hierarchy

Logical

Database
  | Relation
  | Tuple
  | Attribute

Physical

Database
  | File
  | Cluster
  | Page
Hierarchical Locks

- Locks are passed from top-to-bottom (in direction to the leaves)
- Locks must not be overwritten from the top (coming from the root)
- Additionally: *Intentional locks* warn of locks further down in the hierarchy
  - *irl* (*Intentional read lock*)
  - *iw* (*Intentional write lock*)
## Compatibility Matrix

For elementary locks:

<table>
<thead>
<tr>
<th></th>
<th>$rl_i(x)$</th>
<th>$wl_i(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rl_j(x)$</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>$wl_j(x)$</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Compatibility Matrix

For hierarchical locks:

<table>
<thead>
<tr>
<th></th>
<th>$rl_i(x)$</th>
<th>$wl_i(x)$</th>
<th>$irl_i(x)$</th>
<th>$iwl_i(x)$</th>
<th>$riwl_i(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rl_j(x)$</td>
<td>√</td>
<td>—</td>
<td>√</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$wl_j(x)$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$irl_j(x)$</td>
<td>√</td>
<td>—</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>$iwl_j(x)$</td>
<td>—</td>
<td>—</td>
<td>√</td>
<td>√</td>
<td>—</td>
</tr>
<tr>
<td>$riwl_j(x)$</td>
<td>—</td>
<td>—</td>
<td>√</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Process of Hierarchical Locking

1. Locks are placed on a path from the root towards the target.
2. Data object which should be worked on is locked: Write or read lock (Compatibility matrix has to be observed!)
3. All other vertices on the path receive intentional locks.
4. Locks can be intensified, which means that $rl$ can become $wl$, $irl$ can become $rl$ and $irl$ can become $iwl$
5. Unlocking happens the other way around.
Hierarchical Locks: Example I

- **T_1**
  - OrderDB
    - irl(explicit)
  - Product
    - rl(explicit)
  - NY Espresso, ..., Arabica
    - rl(implicit)
  - Price, ..., MinQuantity
    - rl(implicit)

- **T_2**
  - OrderDB
    - irl(explicit)
  - Product
    - iwI(explicit)

- **Database**
  - (OrderDB)

- **Relation**
  - (Product)

- **Tuple**
  - (NY Espresso)

- **Attribute**
  - (Price)

- **Relation**
  - (Order)

- **Tuple**
  - (Arabica Blue)

- **Attribute**
  - (MinQuantity)
Hierarchical Locks: Example II

Database
(OrderDB)

Relation
(Product)

Tuple
(NY Espresso)

Attribute
(Price)

Tuple
(Arabica Blue)

Attribute
(MinQuantity)

OrderDB
(irl(explicit))

Product
(rl(explicit))

NY Espresso
(irl(explicit))

Price
(rl(explicit))

OrderDB
(iwl(explicit))

Product
(iwl(explicit))

Espresso
(iwl(explicit))

MinQuantity
(wl(explicit))

T1

T2

MinQuantity
(wl(explicit))

Last updated: 13.10.2017
Hierarchical Locks: \textit{riwl}

- **Transaction \( T_1 \)**
  - OrderDB
    - \textit{riwl}(explicit)
  - Product
    - \textit{riwl}(explicit)
  - NY Espresso, ..., Arabica
    - \textit{rl}(implicit)
  - NY Espresso, Arabica
    - \textit{wl}(explicit)

- **Transaction \( T_2 \)**
  - OrderDB
    - \textit{irl}(explicit)
  - Product
    - \textit{irl}(explicit)
  - Arabica Blue
    - \textit{rl}(explicit)
Lock Escalation

- Choice of lock granularity
- When executing transactions: many fine-grained locks → Overhead by the lock manager

**Lock Escalation**

- Dynamic union of many fine-grained locks to one (or a few) coarse-grained locks
- Typically: When the number of locks exceed a threshold
Locks in Index Structures

- **Tree Protocol**
  1. Unlocked object \( o \) can only be locked by \( T \) if its direct predecessor has already been locked by \( T \)
  2. This rule does not apply to the first lock of a transaction
  3. Locks can be unlocked at any time
  4. No object within the same transaction \( T \) can be locked more than once
Tree Protocol: Example

\[\text{lock } A; \text{ lock } B; \text{ unlock } A; \text{ lock } D\]
## Sequence within the Tree Protocol

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lock A</td>
<td>lock B</td>
<td>lock E</td>
</tr>
<tr>
<td></td>
<td>lock B</td>
<td>lock C</td>
<td>lock F</td>
</tr>
<tr>
<td></td>
<td>lock D</td>
<td>unlock D</td>
<td>lock G</td>
</tr>
<tr>
<td></td>
<td>unlock B</td>
<td>unlock A</td>
<td>unlock E</td>
</tr>
<tr>
<td></td>
<td>lock C</td>
<td>unlock C</td>
<td>unlock F</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>unlock G</td>
</tr>
</tbody>
</table>

Transaction Management

Last updated: 13.10.2017 4–30
Properties of Tree Protocols

Tree protocol does not force 2PL!

If all transactions fulfill the tree protocol, each correct schedule is serializable.
Latches

- Lightweight (few CPU instructions) objects for short term locks
  - Transport of a record from the buffer to the storage system
  - When inserted into table: Latch for corresponding index block
- Unlike locks, only for in-memory objects
Non-locking Methods

- **Timestamp-based approaches** force a correct order of steps, by marking the transactions appropriately.

- **Serializability testers** manage a conflict graph and therefore verify the conflict serializability directly.

- **Certifiers** execute the complete transaction and later check if the conditions of serializability were fulfilled.
Timestamps

- Timestamps are unique and are allocated consecutively
- For every data object $x$ two values are managed:
  - $\text{max-r-scheduled}[x]$: Variable that contains the timestamp of the last read operation on $x$
  - $\text{max-w-scheduled}[x]$: Variable that contains the timestamp of the last write operation on $x$
Timestamp-ordering rule

An operation $p_i[x]$ is executed before $q_k[x]$ if and only if $ts(T_i) < ts(T_k)$ applies.

- $p$ and $q$: incompatible operations of transactions $T_i$ or respectively $T_k$ ($i \neq k$) on object $x$
- $ts(T_i)$ and $ts(T_k)$ are timestamps of transaction $T_i$ or respectively $T_k$
- Formally:
  
  If $p_i(x)$ and $q_j(x)$ are in conflict, the following applies:
  $p_i(x)$ is executed before $q_j(x) \iff ts(T_i) < ts(T_j)$.
Basic timestamp-ordering-algorithm

\[
\begin{align*}
\text{if } p_i[x] \text{ is fulfilled then} \\
\quad \text{if } p_i[x] \text{ is } r_i[x] \text{ then} \\
\quad \quad \text{if } ts(T_i) < \text{max-w-scheduled}[x] \text{ then} \\
\quad \quad \quad \text{reject operation;} \\
\quad \quad \text{else} \\
\quad \quad \quad \text{max-r-scheduled}[x] := \\
\quad \quad \quad \quad \max(\text{max-r-scheduled}[x], ts(T_i)); \\
\quad \quad \text{execute operation;} \\
\text{else} /* p_i[x] \text{ ist } w_i[x] */ \\
\quad \text{if } ts(T_i) < \text{max-w-scheduled}[x] \text{ or} \\
\quad \quad \text{ts}(T_i) < \text{max-r-scheduled}[x] \text{ then} \\
\quad \quad \quad \text{reject operation;} \\
\quad \text{else} \\
\quad \quad \text{max-w-scheduled}[x] := ts(T_i); \\
\quad \text{execute operation;}
\end{align*}
\]
Advantages and disadvantages of the TO-method

- **Disadvantages:** Problems with varying run-time of transactions
  - Transactions that take a long time, process their last steps some time after the timestamp has been created
  - These steps are very likely to "be too late"

- **Advantage:** Easy to realize in distributed systems
### Timestamp method: Schedule

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T&lt;sub&gt;1&lt;/sub&gt;</strong></td>
<td><strong>T&lt;sub&gt;2&lt;/sub&gt;</strong></td>
<td><strong>T&lt;sub&gt;3&lt;/sub&gt;</strong></td>
<td><strong>A</strong></td>
<td><strong>B</strong></td>
<td><strong>C</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>mrs</strong></td>
<td><strong>mws</strong></td>
<td><strong>mrs</strong></td>
<td><strong>mws</strong></td>
<td><strong>mrs</strong></td>
<td><strong>mws</strong></td>
</tr>
<tr>
<td><strong>ts = 200</strong></td>
<td><strong>ts = 150</strong></td>
<td><strong>ts = 175</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>read B</strong></td>
<td><strong>read A</strong></td>
<td><strong>read C</strong></td>
<td>0</td>
<td>0</td>
<td>200</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>write B</strong></td>
<td></td>
<td></td>
<td>150</td>
<td>0</td>
<td>200</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>write A</strong></td>
<td></td>
<td></td>
<td>150</td>
<td>0</td>
<td>200</td>
<td>0</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td><strong>write C</strong></td>
<td></td>
<td><strong>↓</strong></td>
<td>150</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td><strong>abort</strong></td>
<td></td>
<td></td>
<td>150</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>175</td>
<td>0</td>
</tr>
</tbody>
</table>

- TS = 200
- TS = 150
- TS = 175
Optimized timestamp-ordering method

- Optimized BTO-rule:
  - If the comparison of a write transaction with the `max-w-scheduled` stamp fails, the write is ignored, unless a conflict with the `max-w-scheduled` stamp occurs.
  - The prior value of `max-w-scheduled` is assumed and the transaction runs on.

~> “Optimization for blind writes"
## Optimized timestamp-ordering method: Schedule

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$A$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mrs</td>
<td>mws</td>
</tr>
<tr>
<td>$ts = 200$</td>
<td>$ts = 150$</td>
<td>$ts = 175$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>write $A$</td>
<td>read $A$</td>
<td>read $C$</td>
<td>150</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>150</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>150</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>write $A$ $\sqrt{\downarrow}$</td>
<td>150</td>
<td>200</td>
</tr>
</tbody>
</table>

Thomas Leich

Transaction Management

Last updated: 13.10.2017 4–40
## Livelocks and timestamp-ordering method

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T'_1$</th>
<th>$T'_2$</th>
<th>$A$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ts = 100$</td>
<td>$ts = 110$</td>
<td>$ts = 120$</td>
<td>$ts = 130$</td>
<td>$\begin{array}{cc} mrs &amp; mws \ 0 &amp; 0 \end{array}$</td>
<td>$\begin{array}{cc} mrs &amp; mws \ 0 &amp; 0 \end{array}$</td>
</tr>
<tr>
<td>write $B$</td>
<td></td>
<td>write $A$</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>read $A\downarrow$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>write $A$</td>
<td></td>
<td></td>
<td>0</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>write $B$</td>
<td></td>
<td>0</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>write $A$</td>
<td>0</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Serializability testers I

Problem: Clean up the conflict graph

\[ r_0(x) w_1(x) w_1(y_1) c_1 \ldots w_n(x) w_n(y_n) c_n w_0(z) \]

\[ T_1 \quad T_n \]

Diagram: A directed graph with nodes labeled \( t_0, t_1, \ldots, t_n \) and arrows indicating the order of transactions.
Serializability testers II

Problem: Clean up the conflict graph

Solution: Only completed transactions without incoming edges may be removed
Serializable tests III

Advantage: Accepts full classes of CSR-schedules

Disadvantage: Graph exhibits quadratic growth as the number of transactions increases
Optimistic approaches: Phases

1. **Execution phase** (**execute**): Transaction is executed (all read and write operations); write operations however written to a local copy of the data object within the **buffer**

2. **Validation phase** (**validate**)
   Any transaction that wants to perform a **commit** is checked for correctness with regard to conflict serializability

3. **Persistency phase** (**persist**):
   If no conflicts have occurred, all manipulated database objects are copied back into the persistent database

Step 2 and 3 often performed together

**valpersist**: Joint atomic execution of **validate** and **persist**
Validation criterion

- **Transaction counter TC**
  \[ \forall T_i, T_j : n(T_i) < n(T_j) \text{ with} \]
  \[ n(T_i) = \text{Value of TC after validation of } T_i : \]
  - \( T_i \) finishes its **valpersist**-phase, before \( T_j \) starts.
  - \( x \in ws(T_i) \cap rs(T_j) \Rightarrow T_i \text{ finishes its } \text{valpersist} \text{-phase, before } T_j \text{ reads } x. \)
  - \( x \in rs(T_i) \cap ws(T_j) \Rightarrow T_i \text{ reads } x \text{ before } T_j \text{ begins its } \text{valpersist} \text{-phase} \)

- **ws**\( (T_i) \) : write-set of \( T_i \) (all DB-objects written by \( T_i \))
- **rs**\( (T_i) \) : read-set of \( T_i \) (read DB-objects)
Optimistic scheduler I

T_1
- execute
- validate
- persist

T_2
- execute
- validate
- persist

T_3
- execute
- validate
- persist

backward oriented validation

T_1
- execute
- validate
- persist

T_2
- execute

T_3
- execute

forward oriented validation
Optimistic scheduler II

- **Backward oriented validation:**
  - Test $rs(T_i)$ against $ws(T_j)$ for already completed $T_j$
  - Excluded are $T_j$, that have already performed a commit before the begin of $T_i$
  - In case of conflict: Reset $T_i$
  - Example: Test $rs(T_1)$ against $ws(T_2)$ and $ws(T_3)$

- **Forward oriented validation:**
  - Test $ws(T_i)$ against $rs(T_j)$ for transactions $T_j$, that are active (in execution phase)
  - In case of conflict: Reset $T_j$
  - Example: $ws(T_1)$ against $rs(T_2)$ and $rs(T_3)$
Multiversion concurrency control (MVCC)

- Realization in Oracle, InterBase, PostgreSQL, ...
- Given schedule $s$:

$$s = r_1(x)w_1(x)r_2(x)w_2(y)r_1(y)w_1(z)c_1c_2$$

- $s \notin CSR$
- Tolerable if $r_1(y)$ is able to read prior version of $y$ so that $r_1(y)$ is consistent to $r_1(x)$
- Then: $s$ equivalent to $s' = t_1t_2$
MVCC: Principle

- Each write operation $w$ creates a new version of the current data object
- Read operations can choose a version
- Versions are transparent for applications
- Advantages:
  - Decoupling of read operation and write operation $\leadsto$ Read transaction has the same view on the database as at BOT
  - No synchronization of read transactions as well as against read transactions necessary $\leadsto$ Reduction of conflict probability
  - Synchronization of manipulating operations by other approaches (locking, timestamps)
**MVCC: Example**

<table>
<thead>
<tr>
<th></th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w(x_0 \rightarrow x_1)$</td>
<td>$w(y_0 \rightarrow y_1)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>commit</td>
<td></td>
<td>$r(y_0)$</td>
</tr>
<tr>
<td></td>
<td>$w(x_1 \rightarrow x_2)$</td>
<td></td>
<td>$r(x_0)$</td>
</tr>
<tr>
<td></td>
<td>commit</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MVCC: Version Management I

- Task: Determination of versions that are to be read
  - Global transaction counter \( TNC \) (transaction number count)
  - Commit timestamp \( cts \) for manipulating transactions and \( BOT \) timestamps \( bts \) for read transactions
  - Write timestamp \( wts(x) \) for object \( x \)
MVCC: Version Management II

- Commit of an update transaction $T_U$:
  - Current $TNC$-value as a commit timestamp $cts$ as well as incrementation of $TNC$
  - Modified Object: Write timestamp $wts(x) = cts(T_U)$

- Read Transaction $T_R$:
  - Current $TNC$-value as BOT-timestamp: $bts(T_R) = TNC$
  - Has access to most recent version of $x$ with:
    
    $$wts(x) < bts(T_R)$$

    i.e, Read from $x_i$ with
    
    $$i = \max\{j \mid wts(x_j) < bts(T_R)\}$$
MVCC: Version Management: III

- Task: Release of no longer required versions (*garbage collection*)
  - BOT timestamp of oldest read transaction: $bts_{\text{min}}$
  - Version $x_i$ of object $x$ may be deleted if more recent version $x_j$ exists, so that the following requirement holds:
    \[
    wts(x_i) < wts(x_j) < bts_{\text{min}}
    \]

- Implementation as a ring buffer
Transactions in SQL-DBS

Weakening of ACID in SQL-92: Isolation levels

```sql
set transaction
[ { read only | read write }, ]
[isolation level
  { read uncommitted |
    read committed |
    repeatable read |
    serializable }, ]
[ diagnostics size ...]
```

Standard setting:

```sql
set transaction read write,
  isolation level serializable
```
Isolation levels I

- **read uncommitted**
  - Weakest level: Access to uncommitted data, only for read only transactions
  - Statistical and similar transactions (rough overview, no correct values)
  - No locks → can be executed efficiently, no other transactions are impaired

- **read committed**
  - Reading of committed data only, but non-repeatable read possible

- **repeatable read**
  - No non-repeatable read, but phantom problem may occur

- **serializable**
  - Guaranteed serializability
# Isolation levels II

<table>
<thead>
<tr>
<th>Isolation level</th>
<th>Dirty Read</th>
<th>Nonrepeatable Read</th>
<th>Phantom Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Uncommitted</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Read Committed</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Repeatable Read</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Serializable</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
**Isolation levels: read committed**

<table>
<thead>
<tr>
<th></th>
<th>$T_1$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>set transaction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>isolation level</td>
<td></td>
</tr>
<tr>
<td></td>
<td>read committed</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>select LastName from CUSTOMER where CNr = 4711</td>
<td>update CUSTOMER</td>
</tr>
<tr>
<td></td>
<td>→ Jagellowsk</td>
<td>set LastName = 'Heuer'</td>
</tr>
<tr>
<td>3</td>
<td>select LastName from CUSTOMER where CNr = 4711</td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ Jagellowsk</td>
<td>update CUSTOMER</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>set LastName = 'Heuer'</td>
</tr>
<tr>
<td>5</td>
<td>select LastName from CUSTOMER where CNr = 4711</td>
<td>commit</td>
</tr>
<tr>
<td></td>
<td>→ Heuer</td>
<td></td>
</tr>
</tbody>
</table>
### Isolation levels: read committed

<table>
<thead>
<tr>
<th></th>
<th>$T_1$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><code>set transaction isolation level read committed</code></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><code>select LastName from CUSTOMER where CNr = 4711</code></td>
<td><code>update CUSTOMER set LastName = 'Heuer' where CNr = 4711</code></td>
</tr>
<tr>
<td>3</td>
<td><code>update CUSTOMER set LastName = 'Saake' where CNr = 4711</code></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ <strong>blocked</strong></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><code>commit</code></td>
<td><code>commit</code></td>
</tr>
<tr>
<td>5</td>
<td><code>commit</code></td>
<td></td>
</tr>
</tbody>
</table>
## Isolation levels: Serializable

<table>
<thead>
<tr>
<th></th>
<th>$T_1$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>set transaction</td>
<td>update CUSTOMER</td>
</tr>
<tr>
<td></td>
<td>isolation level</td>
<td>set LastName = 'Heuer'</td>
</tr>
<tr>
<td></td>
<td>serializable</td>
<td>where CNr = 4711</td>
</tr>
<tr>
<td></td>
<td></td>
<td>update CUSTOMER</td>
</tr>
<tr>
<td>2</td>
<td>select LastName from CUSTOMER where CNr = 4711</td>
<td>set LastName = 'Heuer'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>where CNr = 4711</td>
</tr>
<tr>
<td></td>
<td></td>
<td>commit</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>update CUSTOMER</td>
<td></td>
</tr>
<tr>
<td></td>
<td>set LastName = 'Saake'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>where CNr = 4711</td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ error</td>
<td></td>
</tr>
</tbody>
</table>
Oracle

- Support of isolation level *Read Committed* and *Serializable*
- Moreover: *Read-Only*-mode (not a component of SQL-92)

```
set transaction isolation level read committed;
set transaction isolation level serializable;
set transaction isolation level read only;
```
Oracle /2

- Isolation levels for a set of transactions
  
  ```sql
  alter session
  set isolation_level isolation level;
  ```

- Explicit commands to put locks
  
  ```sql
  lock table Table in row share mode;
  lock table Table in share mode;
  lock table Table in row exclusive mode;
  lock table Table in share row exclusive mode;
  lock table Table in exclusive mode;
  ```