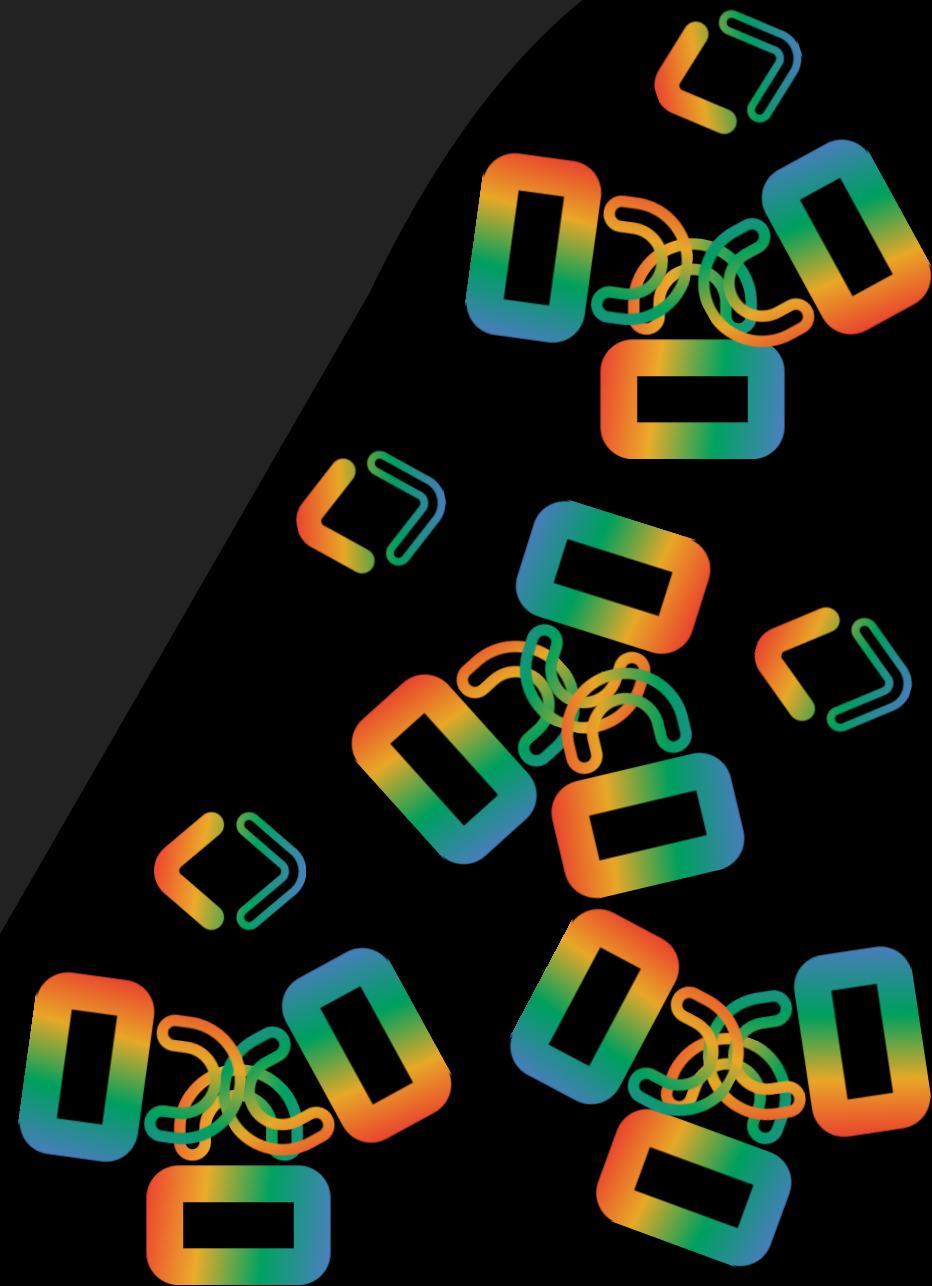




Analysing and Resolving Deadlocks

Erland Sommarskog
Data Platform MVP



Erland Sommarskog



SQL Server MVP since 2001

Independent consultant based in Stockholm

 esquel@sommarskog.se

 Slides and scripts: <https://www.sommarskog.se/present>

Agenda

- Introduction to deadlocks.
- How to get information about deadlocks?
- Understanding the deadlock XML.
- Recap on locking and isolation levels.
- Means to prevent deadlocks.
- Ways to mitigate deadlocks.

What Is a Deadlock?

[01_FirstWindow.sql](#)
[01_SecondWindow.sql](#)

- A deadlock is when two or more processes are blocking each other in such a way that neither can continue.
- SQL Server checks for this condition every five seconds.
- If a deadlock is found, SQL Server injects an error in one of the processes, the *deadlock victim*, and rolls back its transaction.
 - The deadlock victim is selected based on the amount of log records.

What Resources Can Cause Deadlock

- Message 1205 in sys.messages:
*Transaction (Process ID %d) was deadlocked on **%. *ls** resources with another process and has been chosen as the deadlock victim. Rerun the transaction.*
- According to the [Deadlocks Guide](#) in the SQL Server Docs, these are the resources that can deadlock:
 - **Locks. The only one I will talk about.**
 - Worker threads.
 - Memory.
 - Parallel query execution-related resources.
 - Multiple Active Result Sets (MARS) resources.

Are Deadlocks a Problem?

- Deadlocks occur in the best of families.
- An occasional deadlock is no cause for alarm.
- But many deadlocks per day often are.
- Users don't like errors about being “deadlock victims”.
- Too many deadlocks can reduce throughput, since it takes a few seconds until a deadlock is resolved.

Getting Information about Deadlocks

- Easiest is to query the **system_health** session which is always running.

```
SELECT CAST(event_data AS xml), timestamp_utc
FROM sys.fn_xe_file_target_read_file(
    N'system_health*.xel',
    DEFAULT, DEFAULT, DEFAULT)
WHERE object_name = 'xml_deadlock_report'
ORDER BY timestamp_utc DESC
```

[02_query_system_health.sql](#)

- Observe the asterisk – without it you get no output.
- timestamp_utc only available on SQL 2017 and later.

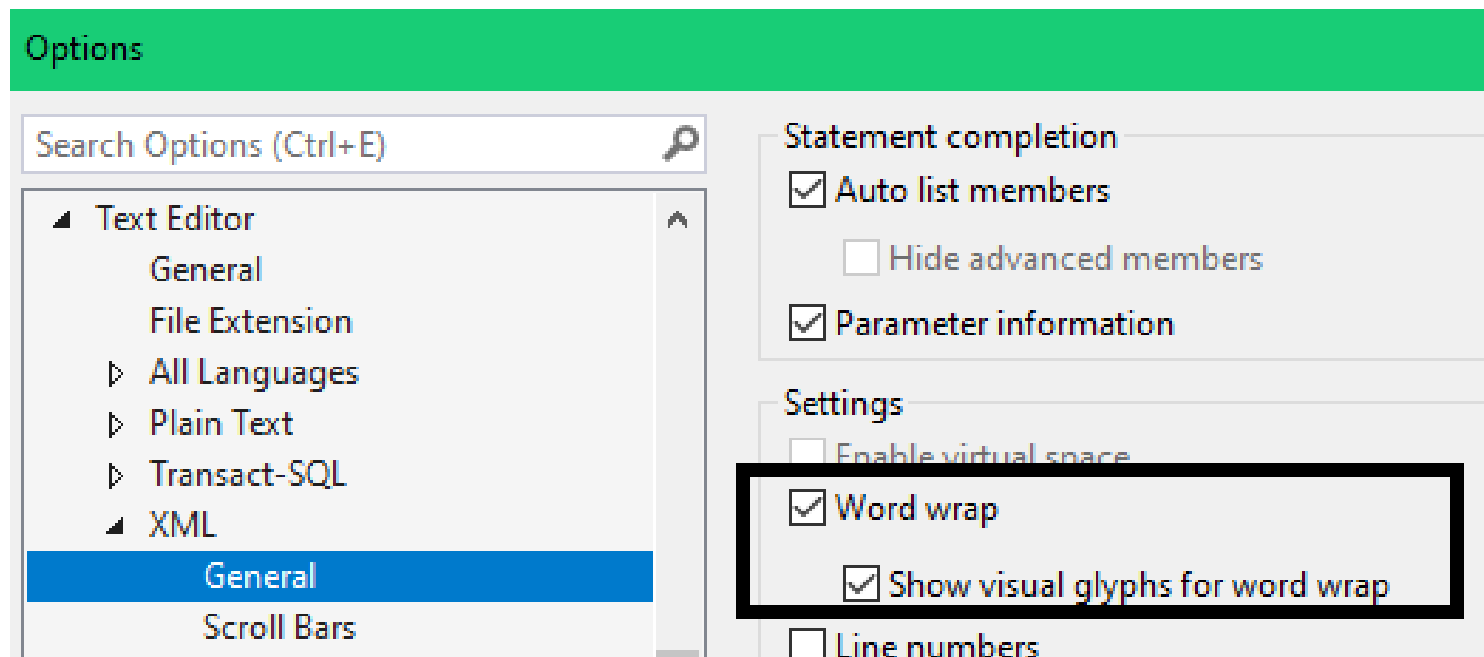
Azure SQL Database

- For Azure SQL DB, run in your **master** database:

```
; WITH CTE AS (  
    SELECT CAST(event_data AS xml) AS xml, timestamp_utc  
    FROM sys.fn_xe_telemetry_blob_target_read_file(  
        'dl', NULL, NULL, NULL))  
SELECT timestamp_utc, xml.query(  
    '/event/data[@name="xml_report"]/value/deadlock')  
FROM CTE  
WHERE xml.exist(  
    '/event/data[@name="database_name"]/value[.="DBname"]')=1
```


Tip for SSMS

- To avoid having to scroll sideways, turn on word wrap for XML:

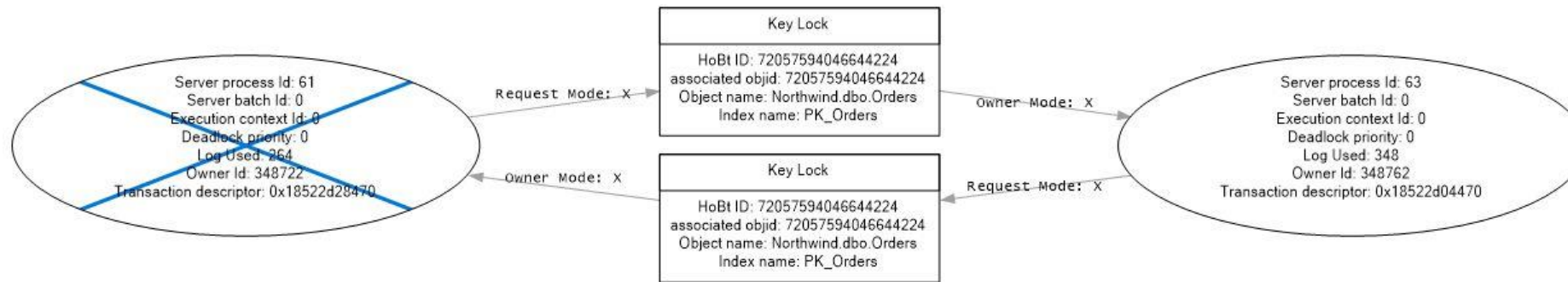


Defining Your Own XEvent Session

- Include the event **xml_deadlock_report**.
- Useful if:
 - On Azure SQL Managed Instance (write to BLOB store) where you only can access system_health through ring buffer which is unreliable.
 - You find that querying system_health session is slow.
 - You want low latency.
- Query it in the same way as the system_health session. (Don't forget the asterisk!)

Other Means to Get Information

- Turn on TF 1222 to get XML in the SQL Server errorlog.
 - Simple, but litters the errorlog, and information is difficult to read.
- Trace/Profiler. Mainly an option on SQL 2005/2008.
 - Profiler gives you a graph that looks pretty:



- But it's confusing and important information is missing. Discard!
- The TextData column has the actual XML.

The Deadlock XML

```
<deadlock>
  <victim-list>
    <victimProcess id="process18519e59088" />
  </victim-list>
  <process-list>...</process-list>
  <resource-list>...</resource-list>
</deadlock>
```

- **victim-list** – The process(es) that were rolled back.
- **process-list** – Information about the processes.
- **resource-list** – The locks involved in the deadlock.

The Process List

```
<process-list>
  <process id="process2b79fb44ca8" ... >
    <stackFrames>...</stackFrames>
    <executionStack>...</executionStack>
    <inputbuf>...</inputbuf>
  </process>
  <process id="process2b792576">...</process>
</process-list>
```

- One tag for each process in the deadlock.
- Many attributes about the processes.

Nested Elements in the Process Tag

```
<stackFrames>...</stackFrames>
```

```
<executionStack>...</executionStack>
```

```
<inputbuf>...</inputbuf>
```

- **stackFrames** – SQL Server call stack, only of interest to support engineers. (New in SQL 2022).
- **executionStack** and **inputbuf** – information about the submitted command and current statement.

The inputbuf Tag, take one

```
<inputbuf>  
UPDATE dbo.Orders  
SET     Freight = 17.23  
WHERE   OrderID = 11000  
</inputbuf>
```

- For ad-hoc commands, the `inputbuf` tag often suffices.
- For multi-statement batches, you may have use for the line numbers in the `executionStack` tag.

The inputbuf Tag, Take Two

- You may be looking at:

```
<inputbuf>  
Proc [Database Id = 2 Object Id = 757577737] </inputbuf>
```

- This is a stored procedure called through RPC.
- You need to look at the **executionStack** tag for more information.

The executionStack Tag

```
<executionStack>
  <frame procname="tempdb.dbo.inner_sp" line="10" stmtstart="502"
    stmtend="612" sqlhandle="0x03000200ee8...">
INSERT RollYourOwn(id, value) VALUES(@id, @value    </frame>
  <frame procname="tempdb.dbo.outer_sp" line="3" stmtstart="150"
    stmtend="212" sqlhandle="0x0300020027b...">
EXEC inner_sp @value, @id OUTPU    </frame>
</executionStack>
```

- You see the statement of the deadlock, and the call stack.
- In this example: the application called outer_sp.
- Last character in statement is stripped off on SQL 2014+.

Advanced: Find the Query Plan

```
<executionStack>
```

```
  <frame procname="tempdb.dbo.inner_sp" line="12" stmtstart="884"  
    stmtend="990" sqlhandle="0x0300020027b3813a ..." >
```

```
INSERT RollYourOwn(id, data) VALUES(@id, @data </frame>
```

```
  <frame ...
```

```
SELECT convert(xml, etqp.query_plan) AS query_plan  
FROM   sys.dm_exec_query_stats qs  
CROSS  APPLY sys.dm_exec_text_query_plan(qs.plan_handle,  
      qs.statement_start_offset, qs.statement_end_offset) AS etqp  
WHERE  qs.sql_handle = 0x0300020027b3813a...  
      AND qs.statement_start_offset = 884  
      AND qs.statement_end_offset = 990
```

Advanced:

Find the Query Plan - Caveats

- Execution plan may have changed since the deadlock.
- Plan may have fallen out of the cache entirely.
- The query may return multiple rows, because there are cache entries for different SET options etc.
 - Tip: look at `sys.dm_exec_query_stats.execution_count`.
- [Searching Query Store](#) may be a better option.

All the Process Attributes

```
<process id="process26eab72f848" taskpriority="0" logused="244"  
waitresource="KEY: 6:72057594047889408 (fadcdcb5e33c)"  
waittime="2728" ownerId="4260688" transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="0x270d78b8428"  
lockMode="X" schedulerid="1" kpid="9744" status="suspended"  
spid="53" sbid="0" ecid="0" priority="0" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783" lastattention="1900-  
01-01T00:00:00.783" clientapp="slask.pl" hostname="SOMMERWALD"  
hostpid="6632" loginname="PRESENT10\sommar" isolationlevel="read  
committed (2)" xactid="4260688" currentdb="6"  
currentdbname="Northgale" lockTimeout="4294967295"  
clientoption1="671088672" clientoption2="128056">
```

The Relevant Ones

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="" lockMode=""  
schedulerid="" kpid="" status="" spid="53" sbid="0" ecid="0"  
priority="" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```

Process Identification

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="" lockMode=""  
schedulerid="" kpid="" status="" spid="53" sbid="0" ecid="0"  
priority="" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```

Process Identification Details

- `id="process26eab72f848"` Mapping between `process-list`, `resource-list` and `victim-list`.
- `spid="53"` The one we know and love.
- `sbid="0"` Non-zero if MARS is in use.
- `ecid="0"` Non-zero when there is parallelism.
- `currentdbname` and `loginname`. Self-explanatory.
- `clientapp` and `hostname`. Recall that they are set in the connection string.

transactionname Attribute

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="" lockMode=""  
schedulerid="" kpid="" status="" spid="53" sbid="0" ecid="0"  
priority="" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```


Multi-Statement Transaction?

- In a multi-statement transaction, locks in the deadlock may come from previous statements in the transaction!
- If **transactionname** reads *user_transaction*, you have a multi-statement transaction.
- Same if it is *MyTran* or some other name. (Someone said **BEGIN TRANSACTION MyTran**).
- *implicit_transaction* is also a multi-statement transaction. **SET IMPLICIT_TRANSACTIONS ON** is in effect.

Multi-Statement Transaction? cont'd

- If **transactionname** reads SELECT, INSERT, UPDATE, DELETE etc, likely to be an auto-committed statement.
- But check **executionStack** – the deadlock statement could be in a trigger with multiple statements.
 - A trigger always executes in the context of a transaction defined by the statement that fired it.
- **transactionname** can read INSERT EXEC. In this case the procedure runs a multi-statement transaction defined by the INSERT statement.

Three Important Timestamps

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="" lockMode=""  
schedulerid="" kpid="" status="" spid="53" sbid="0" ecid="0"  
priority="" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```

Transaction Started by Application?

- Is **lasttranstarted** before **lastbatchstarted**?
Typically, it's a multi-batch transaction started by the application.
 - But it could be a runaway transaction without a COMMIT.
- Are **lastbatchcompleted** and **lastbatchstarted** far apart in a multi-batch transaction? Investigate!
 - Is it doing work elsewhere which increases the length of the transaction – and thus makes the window for a deadlock wider?

The lastattention Attribute

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="" lockMode=""  
schedulerid="" kpid="" status="" spid="53" sbid="0" ecid="0"  
priority="" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```

Attention Signals

- The application can send an *attention signal* to SQL Server.
- On this signal, SQL Server will:
 - Roll back the current statement.
 - Leave any user transaction open (unless XACT_ABORT is on).
- Most common reason is the error *Timeout Expired*. (Most APIs have a default timeout of 30 s.)
 - Applications must handle this error and roll back any open transactions.
- The red button in SSMS also produces an attention signal.

Unhandled Timeout Errors

- `lastattention` reflects the last time an attention signal was received.
- Is `lastattention` later than `lasttranstarted`? This is a red flag for a runaway transaction!
- Most likely the application has experienced a timeout without rolling back.
- Stop analysing the deadlock. Instead, review the error-handling code in the application.

Transaction Count

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="" lockMode=""  
schedulerid="" kpid="" status="" spid="53" sbid="0" ecid="0"  
priority="" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```


The Deceivable trancount Attribute

- For an INSERT, UPDATE, DELETE or MERGE statement, it will always be at least 2, never one or zero.
- And it's 2, regardless if it is a single-statement transaction or a multi-statement transaction.
- For a nested transaction it is 3 or higher.
- An high value could indicate a runaway transaction.
 - High value = not really matching the stack depth.
 - Missing commit or unhandled timeout.

The isolationlevel Attribute

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.783" XDES="" lockMode=""  
schedulerid="" kpid="" spid="53" sbid="0" ecid="0"  
priority="" tranq="0"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```

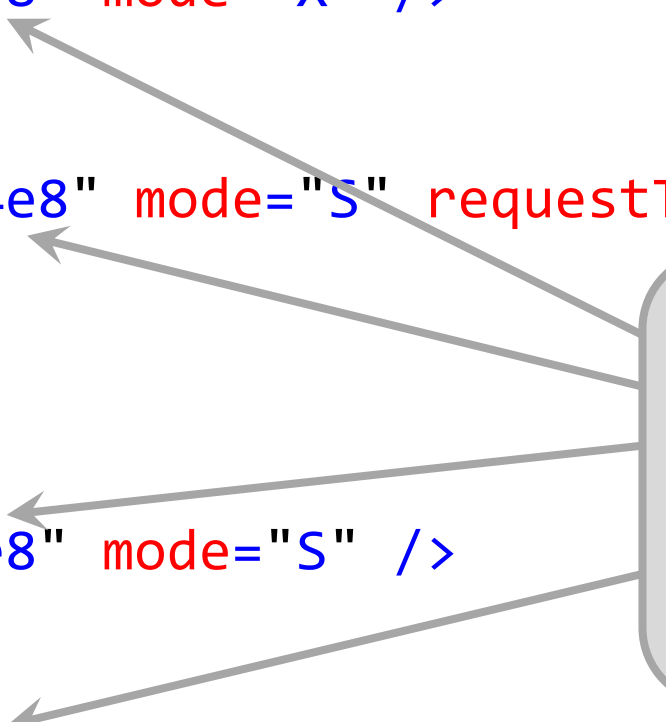
Covered later

The resource-list Element - Overview

```
<resource-list>
  <keylock attr="...">
    <owner-list>...</owner-list>
    <waiter-list>...</waiter-list>
  </keylock>
  <keylock attr="...">
    <owner-list>...</owner-list>
    <waiter-list>...</waiter-list>
  </keylock>
</resource-list>
```

- One element for each lock in the deadlock.
- Name of element depends on type of lock.
- Nested in that element:
 - **owner-list** element listing the owner(s) of the lock.
 - **waiter-list** element listing processes waiting for the lock.

```
<keylock attr="...">
  <owner-list>
    <owner id="process26eabfb68c8" mode="X" />
  </owner-list>
  <waiter-list>
    <waiter id="process26eab72e4e8" mode="S" requestType="wait" />
  </waiter-list>
</keylock>
<keylock attr="...">
  <owner-list>
    <owner id="process26eab72e4e8" mode="S" />
  </owner-list>
  <waiter-list>
    <waiter id="process26eabfb68c8" mode="X" requestType="wait" />
  </waiter-list>
</keylock>
```



A diagram with a central grey box containing the text "Mapping to **id** in the process elements." Four arrows point from this box to the `id` attributes in the XML snippets: to `process26eabfb68c8` in the first `<owner>` tag, to `process26eab72e4e8` in the first `<waiter>` tag, to `process26eab72e4e8` in the second `<owner>` tag, and to `process26eabfb68c8` in the second `<waiter>` tag.

The xxxlock Element

```
<keylock hobtid="72057594047758336" dbid="6"  
objectname="Northgale.dbo.Orders" indexname="PK_Orders"  
id="lock2709eb35780" mode="X"  
associatedObjectId="72057594047758336">
```

- You need to know table and index name. Ignore the rest.
- **keylock** = Row lock in index, clustered or non-clustered.
- **ridlock** = Row lock in heap when accessing data page.
- **pagelock** = What it says, lock on page level.
- **objectlock** = Lock on table level.

Index Name in Page Locks

- In **pagelock** elements, the index name is missing.
- Work from the **associatedObjectId** attribute with this query:

```
SELECT object_name(p.object_id) AS table_name,  
       i.name AS index_name  
FROM   sys.partitions p  
JOIN   sys.indexes i ON p.object_id = i.object_id  
                        AND p.index_id = i.index_id  
WHERE  p.hobt_id = @associatedObjectId
```

Advanced: Which Row is Locked?

- Look in the **process** element for the waiter:
waitresource="KEY: 6:72057594047758336
(3b86dcc184bf)"
- Copy the parenthetical part and run:
SELECT * FROM tbl WITH (INDEX = index_name)
WHERE %%LOCKRES%% = '(3b86dcc184bf)'
- Must force the index in the **keylock** element!

Deadlocks with Parallelism

- If a deadlock occurs with parallel execution, there is one **process** per thread with same **spid** and different **ecid**.
- The **resource-list** is likely to have **exchangeEvent** elements, because threads are waiting for each other.
- Try to ignore these, and focus on the locks.
- If all processes in the **process-list** have the same **spid**, you have an intra-query deadlock.
 - They are bugs in SQL Server.

Recap on Locking and Isolation Levels

- Granularity.
- The most important lock types.
- Intent locks.
- Isolation Levels.

Recap on Locking - Granularity

- For DML operations, locks can be taken on four levels:
 - Row.
 - Page.
 - (Partition).
 - Table.
- The higher the level, the bigger the risk for deadlocks.
- Normally, SQL Server takes row locks, but if it thinks that a query can touch many rows, it may start with page or table locks to preserve memory.

Lock Escalation

- If a *query* takes more than ~5000 locks on a table, SQL Server attempts to escalate to table level (Or partition, if enabled.)
- If escalation is blocked by other locks, SQL Server jogs along with the row locks and tries again later.
- Lock escalation is never to page level.
- Thus, if you see page locks in a deadlock, they were there when the query started.

Controlling Lock Escalation

- You can use ALTER TABLE SET LOCK_ESCALATION to control lock escalation per table.
 - TABLE – the default, table level.
 - AUTO – escalation to partition, if table is partitioned.
 - DISABLE – What it says.
- Trace Flags
 - 1224 – Disable lock escalation, unless there is memory pressure.
 - 1211 – Disable lock escalation, period.

The Basic Lock Modes

From strongest to weakest:

- Sch-M – Schema-modification locks.
- X – Exclusive locks.
- U – Update locks.
- S – Shared locks.
- Sch-S – Schema-stability locks.

Schema-Modification Locks (Sch-M)

- Always on table level.
- Taken by DDL statements (ALTER TABLE, CREATE INDEX etc.)
- Incompatible with all other locks.
- Held to the end of transaction.

Exclusive Locks (X)

- Can be on any level.
- Taken by INSERT, UPDATE, DELETE and MERGE.
- Incompatible with all other locks but Sch-S.
- Held to the end of transaction.

Schema-Stability Locks (Sch-S)

- Always on table level.
- Taken by SELECT and other read operations in:
 - Snapshot isolation.
 - Read-committed snapshot.
 - Read uncommitted, a.k.a. NOLOCK.
- Compatible with all locks, but Schema-modification.
 - Prevents schema from changing while query is running.
- Held to the end of the query.

Shared Locks (S)

- Can be on any level.
- Taken by SELECT and other read operations in these isolation levels:
 - Read Committed without snapshot.
 - Repeatable Read.
 - Serializable.
- Incompatible with Sch-M and Exclusive locks.
- Compatible with Sch-S, Update locks and other Shared locks.

Update Locks (U)

- Can be taken on any level.
- Taken by UPDATE, DELETE and MERGE.
- Compatible with Sch-S and Shared locks.
- Incompatible with Sch-M, Exclusive *and with other* Update locks.
- Purpose: reserve a row for update without blocking readers.
- Converted to Exclusive or released when not needed.

Intent Locks

- A.k.a “I am here” locks.
- A process wants to take an exclusive lock on table level. This should be blocked if there is a shared lock on any row.
- Check all rows if any is locked?
- That would be highly inefficient!
- Intent locks to the rescue!

Intent Locks at Work

- Consider: `SELECT col1, col2 FROM tbl WHERE keycol = 987.`
- First take an Intent Shared (IS) lock on table level.
- This will be blocked if another process holds an Exclusive (X) lock on table level.
- Once the IS lock is taken, any attempt to take an Exclusive table lock will be blocked.
- Repeat on page level. (And partition if enabled.)
- Finally, take a Shared (S) lock the row with keycol = 987.

Intent Locks and Compatibility

- There are three types of Intent Locks:
 - Intent Shared (IS).
 - Intent Update (IU).
 - Intent Exclusive (IX).
- All intent locks are compatible with each other.
- IS is incompatible with X and Sch-M.
- IU is incompatible with U, X and Sch-M.
- IX is incompatible with S, U, X and Sch-M.

Isolation Levels

- Read Uncommitted.
- Read Committed (the default, has two implementations).
- Repeatable Read.
- Serializable.
- Snapshot.

Controlling Isolation Levels

- The isolation level can be controlled with a SET command, for instance:

```
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ
```

- The specified level applies to all queries until there is a new SET TRANSACTION ISOLATION LEVEL command or the module exits.
- You can also control the isolation level for a specific table in a query with a hint, for instance:

```
SELECT * FROM tbl WITH (SERIALIZABLE)
```

Read Uncommitted

- Semantics: you accept to read data that is uncommitted and which could be inconsistent.
- Also known as “dirty reads”.
- In query hints often specified as NOLOCK.

Read Uncommitted, Semantics

- Consider this data and query:

```
SELECT SUM(Amount)
FROM   tbl
WHERE  Customer = 'Jane'
```

Customer	Account	Amount
Jane	123	100
Jane	234	200
Jane	345	300
Jane	456	400

- As query runs, a transaction books a transfer of 50€ from account 234 to 345.
- Our query may return 950, 1050 or something else depending on how the transfer is implemented.

Read Uncommitted, implementation

- SELECT statements do not take shared locks on row, page or table level.
- They do take Sch-S locks for all tables in the query.
- Thus, you can still be blocked by DDL, for instance index rebuild.
- INSERT, UPDATE, DELETE and MERGE will take Update/Exclusive locks for their target tables.

Read Committed

- Semantics: Guaranteed to read only committed data.
- Thus the query
`SELECT SUM(Amount) FROM tbl WHERE Customer = 'Jane'`
can only return 1000, even if a transfer is in progress.
- Say that while query is running, Jane withdraws 150€ from account 456 the ATM.
- Our query may return 850 or 1000. Both are correct under Read Committed.

Read Committed in SQL Server

- This is the default isolation level in SQL Server.
- There are two implementations: locks and snapshot.
- Locks is the default in the box product and Azure SQL Managed Instance.
- Snapshot is the default in Azure SQL Database.
- Foreign-key validations always use locks.

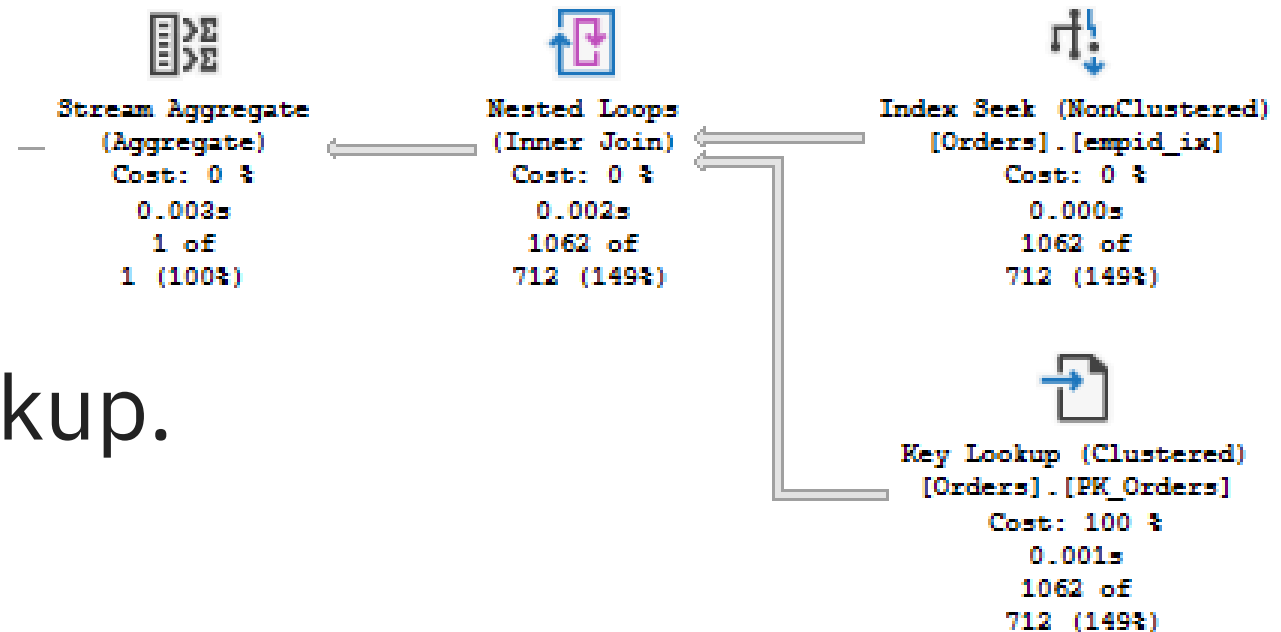
Read Committed with Locks

- SQL Server takes shared locks on row, page or table level.
- If the query
`SELECT SUM(Amount) FROM tbl WHERE Customer = 'Jane'`
runs while Jane withdraws 150€, the result will be 850, as we wait until withdrawal is committed.
- Locks are released when not needed anymore.
- Hm, what does that mean exactly?

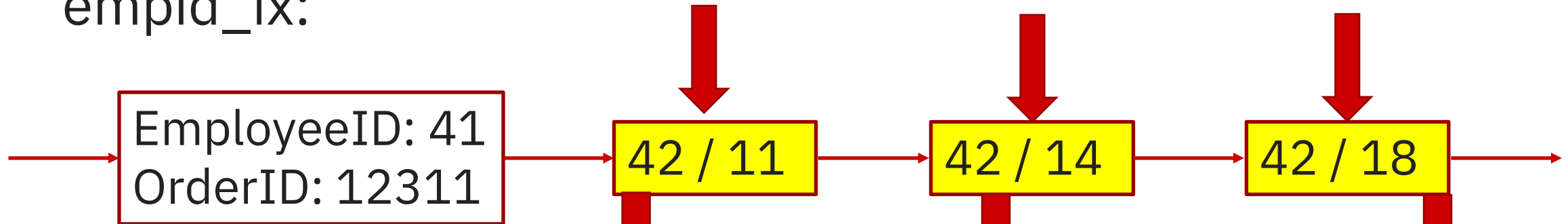
Example: When Not Needed Anymore?

`SELECT SUM(Freight) FROM Orders WHERE EmployeeID = 42`

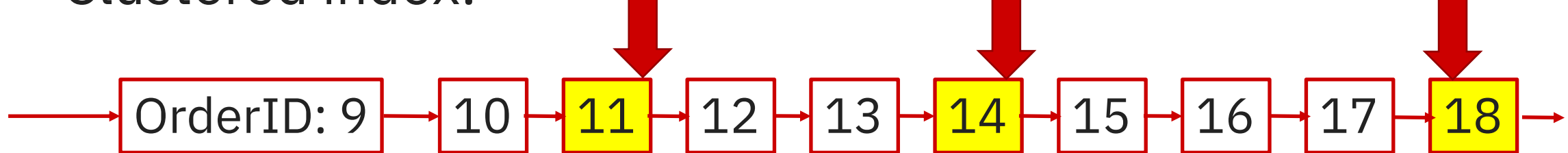
Execution plan:
Index Seek + Key Lookup.



empid_ix:



Clustered index:



Read Committed Snapshot (RCSI)

- SELECT statements do not read from the data pages, but from a version store, usually maintained in tempdb.
- If Accelerated Database Recovery is enabled, version store is in the database itself.
- You read what was committed when query started(*).
- `SELECT SUM(Amount) FROM tbl WHERE Customer = 'Jane'` runs while Jane withdraws 150€. The result will be 1000, the sum before the withdrawal.

Read Committed Snapshot, cont'd

- “when query started(*)”: If there are non-inlined user-defined functions, they have their own starting points.
- Overall, don't rely too much on “when query started”. 850 would still be correct with Read Committed.
- SELECT queries do not take shared locks with RCSI, but take Sch-S locks on all tables in queries.
- UPDATE and DELETE still take Update locks for their target tables without using the version store.
 - They use the version store for other tables.

Controlling Read Committed Snapshot

- To enable/disable Read Committed Snapshot:
`ALTER DATABASE db SET READ_COMMITTED_SNAPSHOT ON / OFF`
- Database must be in single-user mode to make the change.
- When disabled, Read Committed is always by locks.
- When enabled, Read Committed defaults to snapshot.
- You can override with the READCOMMITTEDLOCK hint, but there is no SET TRANSACTION ISOLATION LEVEL command.

Repeatable Read

- Semantics: if you read the same row twice in a transaction, you will get back the same result.
- Use case: Say that we are running
`SELECT SUM(Amount) FROM tbl WHERE Customer = 'Jane'`
We intend to use those 1000€ later in the transaction, so we want to make sure that result does not change.
- Shared locks are held to the end of the transaction.
 - Locks on rows filtered out are released directly.

Serializable

- With Repeatable Read, deposits or withdrawals to the existing accounts will be blocked.
- What if someone adds a new account, for instance a loan with a negative balance?
- With Serializable we are guaranteed that rows can not be added to a range we have read.
- Rather than plain row locks, SQL Server uses key-range locks, for instance RangeS-S, to implement Serializable.

Snapshot Isolation

- Semantics: you see the database as it was when the transaction started. You are blind to all later changes.
- Thus, our sum query *must* return 1000 when Jane is at the ATM.
- Implemented in SQL Server through the version store.
- SELECT statements take no shared locks, only Sch-S.
- UPDATE and DELETE also use the version store for their target tables when searching for rows.

Snapshot Isolation, cont'd

- By default, snapshot isolation is not available.
- You need to enable it with
`ALTER DATABASE db SET ALLOW_SNAPSHOT_ISOLATION ON`
 - Does not require single-user mode.
- The settings `ALLOW_SNAPSHOT_ISOLATION` and `READ_COMMITTED_SNAPSHOT` are independent of each other.
- To run a snapshot transaction, you need to say
`SET TRANSACTION ISOLATION LEVEL SNAPSHOT`

The isolationlevel Attribute

```
<process id="process26eab72f848" taskpriority="" logused=""  
waitresource="" waittime="" ownerId=""  
transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="" lockMode=""  
schedulerid="" kpid="" status="" spid="53" sbid="0" ecid="0"  
priority="" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783"  
lastattention="1900-01-01T00:00:00.783" clientapp="slask.pl"  
hostname="SOMMERWALD" hostpid="" loginname="PRESENT10\sommar"  
isolationlevel="read committed (2)" xactid="" currentdb=""  
currentdbname="Northgale" lockTimeout="" clientoption1=""  
clientoption2="">
```

Isolation Levels and Locks

- If you see Repeatable Read or Serializable, shared locks are held to the end of the transaction.
- Locks being held longer → higher risk for deadlock.
- Isolation level may be required to ensure correct results.
- Beware: some transaction APIs turn on Serializable by default, so it may be “by mistake”.
- Don’t forget that the isolation level may change during the transaction or be overridden by hints.

Preventing Deadlocks

- Always access resources in the same order.
- Query tuning and indexing.
- READ_COMMITTED_SNAPSHOT.
- Lock hints.
- Review application behaviour.
- Serialise with application locks.

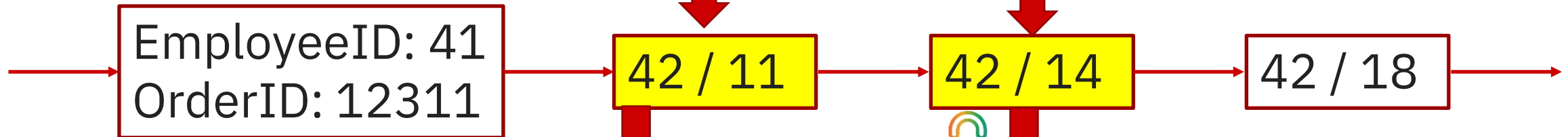
Access Resources in the Same Order

- Standard recommendation. Sounds simple and obvious.
- Can be very difficult to implement in practice.
- Business rules may mandate different access order.
- SQL Server itself does not obey to this rule when building query plans.

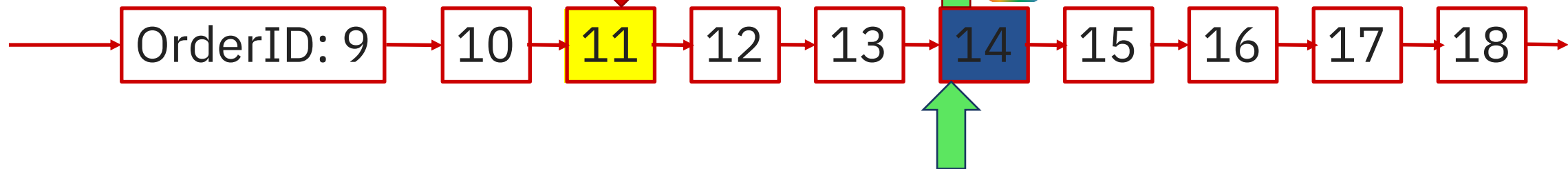
[Second_window.sql](#)
[differentaccessorder.xml](#)

```
SELECT SUM(Freight) FROM Orders WHERE EmployeeID = @empid
```

empid_ix:



Clustered index:



```
UPDATE dbo.Orders SET EmployeeID = @newempid WHERE OrderID = @orderid
```

Preventing Deadlocks

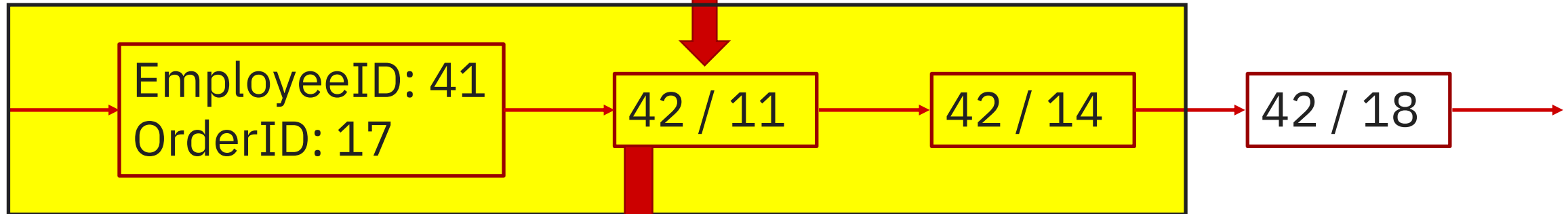
- Always access resources in the same order.
- Query tuning and indexing.
- READ_COMMITTED_SNAPSHOT.
- Lock hints.
- Review application behaviour.
- Serialise with application locks.

The Impact of Page and Table Locks

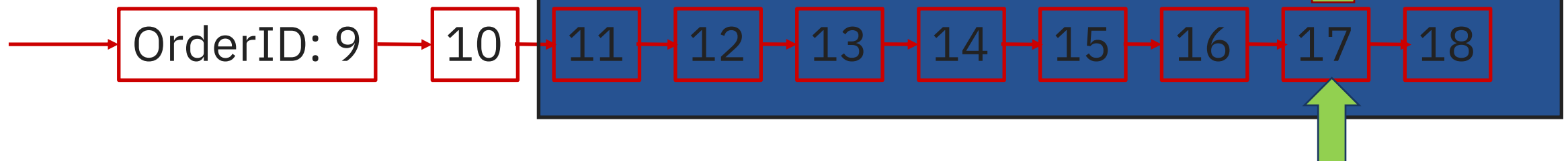
- In the deadlock we just looked at, the processes were fighting about the very same row.
- But imagine that there would only be page locks...

```
SELECT SUM(Freight) FROM Orders WHERE EmployeeID = @empid
```

empid_ix:



Clustered index:



```
UPDATE dbo.Orders SET EmployeeID = @newempid WHERE OrderID = @orderid
```

The Impact of Page and Table Locks

- Lower lock granularity increases the risk for deadlocks, as you now can get conflicts on unrelated rows.
- Thus, removing such locks helps to prevent deadlocks.
- Recall that intent locks are always on page/table. They are not a problem.
- It's shared/update/exclusive locks on higher levels you should go after.
- They are often a token of a query that needs tuning.

Removing Page and Table Locks

- Say that you see this query in a deadlock:

```
SELECT * FROM Orders  
WHERE dateadd(MONTH, 1, OrderDate) > getdate()
```

- Entangling column in an expression, prevents Index Seek.
- You get Index Scan which often results in page locks.
- No need to analyse the rest of the deadlock, but rewrite:

```
SELECT * FROM Orders  
WHERE OrderDate > dateadd(MONTH, -1, getdate())
```


Deadlock with Page Locks

[01_Second_window.sql](#)
[pagelock-deadlock.xml](#)

```
-- Reader.  
SELECT SUM(Freight) FROM TallOrders WHERE OrderDate = @date  
-- Writer.  
UPDATE TallOrders SET ShipVia = @newshipid WHERE OrderID = @orderid
```

- The index in the deadlock is on (ShipVia, OrderDate). There is no index on OrderDate.
- Reader scans this index, taking page locks.
- Parallel plan, thus multiple pages locked simultaneously.

Deadlock with Page Locks, cont'd

- Writer needs to modify *two* pages in index on ShipVia, one for the old value, one for the new value.
- Thus, writer needs intent locks (IX) on both pages.
- Locks the first page, tries to lock the next but is blocked by a reader thread.
- Another reader thread wants to read rows from first page, but is blocked by writer.
- Possible resolution: add an index on OrderDate.
 - And we could tell this from the start....

Are Row Locks Permitted?

- If you have table or page locks you are not able to understand, run this query:

```
SELECT * FROM sys.indexes
WHERE  object_name(object_id) = 'YourTable'
      AND (allow_row_locks = 0 OR allow_page_locks = 0)
```

- If it returns any rows, enable page and row locks with:

```
CREATE INDEX ix ON tbl(col) WITH
  (ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON,
   DROP_EXISTING = ON)
```

Only Row Locks, All Good?

- Even if there are only row locks, query tuning can still be the solution.
- Recall this deadlock:

```
-- Reader
SELECT SUM(Freight) FROM Orders WHERE EmployeeID = @empid
-- Writer
UPDATE Orders SET EmployeeID=@newempid WHERE OrderID=@orderid
```

- Possible resolution: Add Freight as an included column in index on EmployeeID.

One More Deadlock Case

Consider this:

1. Process P starts a transaction.
2. P updates row(s) in table T.
3. P runs query to populate a temp table for two minutes.
4. Process Q starts SELECT joining S and T, starting from S.
5. The rows in T are updated by P, so Q is blocked.
6. P attempts to update row in S that Q has locked.
7. DEADLOCK!

The Window for a Deadlock

- What if we can tune P's INSERT query so that it runs 200 ms?
- That does not entirely *prevent* the deadlock.
- But the *window* where the deadlock can occur is drastically reduced.
- Shorter transactions → Less risk for deadlocks.
- ...but that does not mean you should split up a transaction that needs to be atomic in multiple.

Too Many Indexes?

- The more indexes there are on a table, INSERT/DELETE/ UPDATE statements takes longer. → The deadlock window gets wider.
- Two DML statements that update indexes in different order may clash.

Too Many Indexes, cont'd.

- Check `sys.dm_db_index_usage_stats` for unused indexes.
 - Keep in mind that DMV is cleared on server restarts.
 - You can also trawl query plans in Query Store.
- Check for overlapping indexes. For instance:
 - Indexes on both (a) and (a, b). Drop first index.
 - Note: indexes on (b) and (a, b) are *not* overlapping!
 - Indexes on (a, b, c) INCLUDE (d) and (a, b, c) INCLUDE (e). Consolidate into (a, b, c) INCLUDE (d, e).

Preventing Deadlocks

- Always access resources in the same order.
- Query tuning and indexing.
- **READ_COMMITTED_SNAPSHOT.**
- Lock hints.
- Review application behaviour.
- Serialise with application locks.

READ_COMMITTED_SNAPSHOT

- This is a great way to eliminate all deadlocks between readers and writers.
- Readers don't block writers. Writers don't block readers.
- (Except for readers that use Repeatable Read or Serializable.)

RCSI, Technical Caveats

- There is a price: Incurs an overhead to update and delete operations to maintain the version store.
- Generally, increases the load on tempdb (unless you have Accelerated Database Recovery turned on).
- Adds a 14-byte pointer to updated rows, leading to page splits.
 - These pointers are removed on index rebuild!
 - Rebuild indexes with 90% fill factor?

RCSI, Application Caveats

- You are reading stale data. Does that matter?
- Example: Cannot have an inactivated securable in a portfolio. Cannot inactivate a securable if it's in at least one portfolio. This is checked by triggers.
 - Transaction A: adds a securable to a portfolio.
 - Transaction B: inactivates the same securable.
 - Triggers read from snapshot and finds that all good.
 - Business-rule violation!
- READCOMMITTEDLOCK hint can prevent this.

If RCSI makes you nervous

- Rather than enabling READ_COMMITTED_SNAPSHOT, enable the database for true SNAPSHOT isolation.
- Then use SNAPSHOT isolation in places where you know it is safe.

Preventing Deadlocks

- Always access resources in the same order.
- Query tuning and indexing.
- READ_COMMITTED_SNAPSHOT.
- Lock hints.
- Review application behaviour.
- Serialise with application locks.

The UPDLOCK hint

- This hint is great to prevent conversion deadlocks.

```
SELECT @val = col FROM tbl WITH (REPEATABLEREAD)
WHERE id = @id
-- Code that computes a new value for @val
UPDATE tbl SET col = @val WHERE id = @id
```

- Say two processes run this code for the same @id.
- Both read the row and keep a shared lock on it.
- When they try to update, they block each other.
- Change REPEATABLEREAD to UPDLOCK to prevent this.
 - Now process that comes in second will have to wait until first commits.

The ROWLOCK hint

- If you get deadlocks with page or table locks, you can use the ROWLOCK hint to force row locks.
- SQL Server chooses page or table locks to preserve memory. Forcing row locks may threaten server stability.
- Better to review indexes and/or tune queries.
- Can be a temporary measure until you get a better solution in place.
- Cannot use ROWLOCK to prevent lock escalation.

NOLOCK?

- Sure, NOLOCK can remove plenty of deadlocks.
- ...and your queries start to produce spurious incorrect results that are difficult to reproduce.
- Keep in mind that with NOLOCK you may:
 - Read uncommitted data.
 - Fail to read committed data that was moved due to page splits.
 - Read the same data twice.
- NOLOCK is not a solution to deadlocks, period!
- Use READ_COMMITTED_SNAPSHOT instead.

Preventing Deadlocks

- Always access resources in the same order.
- Query tuning and indexing.
- READ_COMMITTED_SNAPSHOT.
- Lock hints.
- Review application behaviour.
- Serialise with application locks.

Think of What You Can Do Differently

- Is application running a multi-batch transaction with a lot work between the batches?
- Rewrite one-by-one processing into set-based.
- If two operations clash, can one be rescheduled to a different time?
- Disable user updates when you need to run a big import during business hours.
- Split up transactions into several in a *controlled* way.
(Not a step to take lightly!)

Preventing Deadlocks

- Always access resources in the same order.
- Query tuning and indexing.
- READ_COMMITTED_SNAPSHOT.
- Lock hints.
- Review application behaviour.
- Serialise with application locks.

Application Locks

- You may have an operation where parallel execution isn't really meaningful.
- You can serialise with application locks.

```
EXEC sp_getapplock @Resource = 'MyLock',  
                  @LockMode = 'Exclusive'
```

- *MyLock* = name of your lock. Up to 32 characters. Scope of name is the database.
- Blocks until lock is available.
- Lock is held until end of the transaction.

Application Locks, cont'd

- Use application locks when parallel processing should not occur.
- But don't use application locks to prevent occasional deadlocks – can hamper concurrency in the system.
- Application locks can also be taken on session level. Great if you want to prevent parallel execution of a stored procedure entirely.
- See further [Books Online](#).

Mitigating Deadlocks

- SET DEADLOCK_PRIORITY.
- Retry on deadlock.
- SET LOCK_TIMEOUT.

Deadlock Priority

- If a background process clashes with end users, you may prefer the background process to be the deadlock victim.
- Use this statement: `SET DEADLOCK_PRIORITY LOW`
- Rather than LOW, you can give HIGH, NORMAL or a number from -10 to 10. LOW is the same as -5.
- I've never had use for anything but LOW.

Deadlock Priority in the XML

```
<process id="process26eab72f848" taskpriority="5" logused="244"  
waitresource="KEY: 6:72057594047889408 (fadcdcb5e33c)"  
waittime="2728" ownerId="4260688" transactionname="UPDATE"  
lasttranstarted="2022-01-30T14:57:47.787" XDES="0x270d78b8428"  
lockMode="X" schedulerid="1" kpid="9744" status="suspended"  
spid="53" sbid="0" ecid="0" priority="-5" trancount="2"  
lastbatchstarted="2022-01-30T14:57:47.783"  
lastbatchcompleted="2022-01-30T14:57:47.783" lastattention="1900-  
01-01T00:00:00.783" clientapp="slask.pl" hostname="SOMMERWALD"  
hostpid="6632" loginname="PRESENT10\sommar" isolationlevel="read  
committed (2)" xactid="4260688" currentdb="6"  
currentdbname="Northgale" lockTimeout="4294967295"  
clientoption1="671088672" clientoption2="128056">
```

Retry on Deadlock

- Rather than just displaying an error and quit, we can re-attempt the operation that caused the deadlock.
- Some people say: Always retry on deadlock.
- I say: *Only* implement retry in situations where deadlocks are a real problem.
- Incorrectly implemented deadlock retry can lead to data errors that are difficult to explain.
- Testing is difficult – you need a deadlock to test.

Two Key Rules for Deadlock Retries

- Never redo only part of a transaction.
 - => **Never do deadlock retry when called inside a transaction.**
 - On deadlock, the *entire* transaction must be rolled back – you cannot only roll back your own part.
- Never redo a committed transaction. That could lead to data being doubled.
 - E.g.: after the transaction there is a SELECT that deadlocks.
- These sort of errors occur only very occasionally and can be very difficult to troubleshoot.

Demo on Deadlock Retry

- This demo also shows conversion deadlocks and how to prevent them with the UPDLOCK hint.
- Additionally, it sports application locks and deadlock priority. [03_deadlockdemo_main.sql](#)
[03_deadlockdemo2.sql](#)
[03_deadlockdemo3.sql](#)
- Notice how the deadlock retry clutters the code.
- This makes it more difficult to discern the business logic, and makes it more difficult to spot bugs.

Generic Retry in Data Layer?

- Avoids the code clutter – but opens trapdoors.
- You must still check for open transactions.
- How do you know that all stored procedures run at most one transaction? What if there is code after COMMIT that can deadlock?
- Strict coding guidelines?
- Retry of read-only operations is OK, if you can discern them.

Retry in the Business Layer?

- If that is where transaction starts, you have no choice.
- You may be back to the code clutter.
- Then again, if you have a background task that runs once a minute – there is an automatic retry.
- And for that matter, telling the user to try again is also a retry...

Why Lock Timeouts?

- Background processes can lower their deadlock priority and implement good retry.
- This saves more important processes from being deadlock victims.
- Still, they can be held up for five seconds – that may be bad enough. (Reduced throughput, irritated users etc.)
- What if the background process could step out of the clash at an early stage?

SET LOCK_TIMEOUT

- `SET LOCK_TIMEOUT 100`
- If blocked by a lock, wait 100 ms to get the lock.
- When timeout expires, error 1222 is raised.

Msg 1222, Level 16, State 51, Line 3
Lock request time out period exceeded.

- On this error, rollback and retry (after a short wait).
- `SET LOCK_TIMEOUT 0` = don't wait. `-1` = wait forever.
- There is a `lockTimeout` attribute in deadlock XML.

Advanced: The NOWAIT hint

- Frivolous use of SET LOCK_TIMEOUT can lead to the background process never getting work done.
- What if there is a table in a specific query that is particularly prone to deadlocks?
- You can use the NOWAIT hint to have a lock timeout of zero for that table only:

```
SELECT ...  
FROM   TableA A WITH (NOWAIT)  
JOIN   TableB B ON A.col = B.col
```

Summary: Key Points of this Session

- An occasional deadlock is no cause for alarm.
- Get the deadlock XML from the system_health session.
- Is any process in a multi-statement transaction?
The transaction name will tell you.
- Is any process running a multi-batch transaction?
Is **lasttranstarted** before **lastbatchstarted**?
- Remember: The faster your operations are, the smaller the window where a deadlock can occur.

Summary: Don't Shy the Shortcuts!

- Understanding a deadlock in full detail can be hard.
- The good news: you only need to understand as much to make a decision on prevention or mitigation.
 - Is **lastattention** after **lasttranstarted**? Troubleshoot the unhandled timeout.
 - When you see page and table locks, apply query and index tuning before anything else.
 - READ_COMMITTED_SNAPSHOT to isolate readers and writers.
 - Use deadlock priority + deadlock retry for background processes.
 - When things should not run in parallel, serialise or separate.

Concurrency Conflicts in Databases

- Deadlocks is just one way they can manifest.
 - Update errors with snapshot isolation or In-Memory OLTP.
 - Or the PK error in the demo.
- Bad throughput due to too strict isolation level or serialisation in the wrong place.
- Worst of them all: Spurious data errors.
 - Too liberal isolation level (which could be RCSI).
 - Incorrect transaction scope.
 - Incorrect deadlock retry.

Summary: Choose Wisely

- There is often more than one way to prevent or mitigate a deadlock.
- You need to choose wisely.
- The cure can be worse than the disease.
 - Spurious data errors.
 - Serialisation reducing throughput of the system.
 - Etc.
- Take a holistic view of the situation.

Session evaluation

Your feedback is important to us



Evaluate this session at:

www.PASSDataCommunitySummit.com/evaluation



Thank you

Analysing and Resolving Deadlocks

Erland Sommarskog

esquel@sommarskog.se

Slides and scripts:

<https://www.sommarskog.se/present>

