Examples for Chapter 9 & 10:

Oracle 12c & SQL Server: B+ Trees, Transactions, and Locking

shaked19@gmail.com

http://www.ildba.co.il/author/cimid/
http://www.sqlserver.co.il/?cat=940
3rd Israeli Conference on Software Architecture
http://www.iltam.org/sw-arch2016/arch2016_page#OpenSource
http://www.youtube.com/watch?v=PgoJSgpXJ8w&t=0s&list=PLIIWwVvULn4QFsi_U962kFJi6hvDFVydv&index=1
Global Hebrew Virtual PASS Chapter:
https://www.youtube.com/watch?v=x4hGjYGBfkc
https://www.youtube.com/watch?v=eJO8G9if3EY
SqlSaturday Israel 2016:
SqlSaturday Israel 2017:
http://www.sqlsaturday.com/623/Sessions/Detail
SqlSaturday Israel 2018:
Reference and Credits

Oracle® Database Concepts
12c Release 1 (12.1)
E41396-13
https://docs.oracle.com/database/121/CNCPT/toc.htm

Oracle® Database Performance Tuning Guide
12c Release 1 (12.1)
E49058-06
https://docs.oracle.com/database/121/TGDBA/toc.htm

Oracle® Database SQL Language Reference
12c Release 1 (12.1)
E41329-20

Oracle Essentials(Oracle Database 12c), 5th; O'Reilly, 2013

Oracle OCA Oracle Database 12c Administrator Certified Associate Study Guide Exam

Pro Oracle Database 12c Administration, 2 edition ISBN 1430257288 2013

Apress Oracle Database Transactions and Locking Revealed (2014)

Oracle Learning Library

Pro SQL Server Internals 2014 Apress
SQL Server Concurrency  Locking, Blocking and Row Versioning   By Kalen Delaney
SQL 2016 Book Online

Transactional Information Systems
**SQL Server Clustered index structure: Root level**

- Table can have only one clustered index defined.
- Data pages are linked in a double-linked list where every page contains pointers to the next and previous pages in the chain. This list is called the *leaf level of the index, and it contains the actual table data*.
Clustered index structure: Leaf level

- Create a clustered index on the heap table with the data.
- First step SQL Server creates another copy of the data that is sorted based on the value of the clustered key.
- The data pages are linked in a double-linked list where every page contains pointers to the next and previous pages in the chain. This list is called the leaf level of the index, and it contains the actual table data.

```sql
create table dbo.Customers
(
    CustomerId int not null,
    Name varchar(64) not null,
    -- Other columns
)
```

```sql
create clustered index IDX_CI on dbo.Customers(CustomerId)
```
Clustered index structure: Intermediate and root levels

- When the leaf level consists of the multiple pages, SQL Server starts to build an *intermediate level of the index,***
Ordered index scan

- Data on the leaf level of the index is already sorted based on CustomerId column value
- SQL Server can scan the leaf level of the index from the first to the last page and return the rows in the order they are stored.
Ordered index scan execution plan

- An ordered scan just means that SQL Server reads the data based on the order of the index key.
- SQL Server can navigate through indexes in both directions, forward and backward.
Nonclustered index

create nonclustered index IDX_NCI on dbo.Customers(Name)

- While a clustered index specifies how data rows are sorted in a table, nonclustered indexes define a separate sorting order for a column or set of columns and persist them as a separate index structure.
- Leaf level of the nonclustered index is sorted based on the value of the index key—*Name in our case.*
- Every row on the leaf level includes the key value and row-id.
- For heap tables, row-id is the physical location of the row defined as file:page:slot address.
Transactional Information Systems:

Theory, Algorithms, and the Practice of Concurrency Control and Recovery

Gerhard Weikum and Gottfried Vossen

© 2002 Morgan Kaufmann
ISBN 1-55860-508-8

“Teamwork is essential. It allows you to blame someone else.” (Anonymous)
Part II: Concurrency Control

- 3 Concurrency Control: Notions of Correctness for the Page Model
- 4 Concurrency Control Algorithms
- 5 Multiversion Concurrency Control
- 6 Concurrency Control on Objects: Notions of Correctness
- 7 Concurrency Control Algorithms on Objects
- 8 Concurrency Control on Relational Databases
- 9 Concurrency Control on Search Structures
- 10 Implementation and Pragmatic Issues
Chapter 10: Implementation and Pragmatic Issues

• 10.2 Data Structures of a Lock Manager

• 10.3 Multi-Granularity Locking and Lock Escalation

• 10.4 Transient Versioning

• 10.5 Nested Transactions for Intra-transaction parallelism

• 10.6 Tuning Options

• 10.7 Overload Control

• 10.8 Lessons Learned

“All theory, my friend, is grey; but the precious tree of life.”

(Johann Wolfgang von Goethe)
Organization of Lock Control Blocks

Transaction Control Blocks (TCBs)

- Transaction Id
- Update Flag
- Transaction Status
- Number of Locks
- LCB Chain

Hash Table indexed by Resource Id

Resource Control Blocks (RCBs)

- Resource Id
- Hash Chain
- FirstInQueue

Lock Control Blocks (LCBs)

- Transaction Id
- Resource Id
- Lock Mode
- Lock Status
- NextInQueue
- LCB Chain
Implementation of Locking

- A **lock manager** can be implemented as a separate process to which transactions send lock and unlock requests.
- The lock manager replies to a lock request by sending a lock grant messages (or a message asking the transaction to roll back, in case of a deadlock).
- The requesting transaction waits until its request is answered.
- The lock manager maintains a data-structure called a **lock table** to record granted locks and pending requests.
- The lock table is usually implemented as an in-memory hash table indexed on the name of the data item being locked.
**Lock Table**

- Dark blue rectangles indicate granted locks; light blue indicate waiting requests
- Lock table also records the type of lock granted or requested
- New request is added to the end of the queue of requests for the data item, and granted if it is compatible with all earlier locks
- Unlock requests result in the request being deleted, and later requests are checked to see if they can now be granted
- If transaction aborts, all waiting or granted requests of the transaction are deleted
  - lock manager may keep a list of locks held by each transaction, to implement this efficiently
Bucket Count and Performance

- A hash based on the first two letters of string and can return $26 \times 26 = 676$ different hash keys.
- Assuming that the hash table can accommodate all 676 different hash buckets and you have the data shown in Figure you will need to traverse at most two rows in the chain when you run a query that looks for a specific value.

```
select * from T where Name = 'Ann'
```
Bucket Count and Performance

• Situation changes if your hash table does not have enough buckets to separate unique hash keys from each other.

• Figure illustrates the situation when a hash table has only 26 buckets and each of them stores multiple different hash keys.

• Now The same lookup of the Ann row requires you to traverse the chain of nine rows total.
Chapter 10: Implementation and Pragmatic Issues

- 10.2 Data Structures of a Lock Manager
- **10.3 Multi-Granularity Locking and Lock Escalation**
- 10.4 Transient Versioning
- 10.5 Nested Transactions for Intra-transaction parallelism
- 10.6 Tuning Options
- 10.7 Overload Control
- 10.8 Lessons Learned
Reconciling Coarse- and Fine-grained Locking

**Problem:** For reduced overhead, table scans should use coarse locks
Detect conflict of page lock with tablespace lock

**Approach:** Set “intention locks” on coarser granules

**Multi-granularity locking protocol:**
- A transaction can lock any granule in S or X mode.
- Before a granule p can be locked in S or X mode, the transaction
  needs to hold an IS or IX lock on all coarser granules that contain p.

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>X</th>
<th>IS</th>
<th>IX</th>
<th>SIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IS</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IX</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>SIX</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Typical policy:**
- use coarse locks for table scans
- use fine locks otherwise
- escalate dynamically to coarse locks when memory usage for LCBs becomes critical
Multiple-granularity

- Data items of multiple size
- Represented in form of tree
- When a node is locked all the children are automatically locked.

Problems:
1. In order to lock a lower level we need to check from root to that node if there is any locked node
2. We want to acquire a lock on DB we need to check if any of decedents should not lock by any transaction
Multiple Granularity

• Allow data items to be of various sizes and define a hierarchy of data granularities, where the small granularities are nested within larger ones.

• Can be represented graphically as a tree.

• When a transaction locks a node in the tree explicitly, it implicitly locks all the node's descendants in the same mode.

• **Granularity of locking** (level in tree where locking is done):
  - fine granularity (lower in tree): high concurrency, high locking overhead
  - coarse granularity (higher in tree): low locking overhead, low concurrency
Intention Lock Modes

- In addition to S and X lock modes, there are three additional lock modes with multiple granularity:
  - *intention-shared (IS)*: indicates explicit locking at a lower level of the tree but only with shared locks.
  - *intention-exclusive (IX)*: indicates explicit locking at a lower level with exclusive or shared locks.
  - *Shared and intention-exclusive (SIX)*: the subtree rooted by that node is locked explicitly in shared mode and explicit locking is being done at a lower level with exclusive-mode locks.

- Intention locks allow a higher level node to be locked in S or X mode without having to check all descendent nodes.
• Transaction $T_i$ can lock a node $Q$, using the following rules:
  1. The lock compatibility matrix must be observed.
  2. Root of the tree must be locked first, and may be locked in any node.
  3. A node $Q$ can be locked by $T_i$ in S or IS mode only if the parent of $Q$ is currently locked by $T_i$ in either IX or IS mode.
  4. A node $Q$ can be locked by $T_i$ in X, SIX, or IX mode only if the parent of $Q$ is currently locked by $T_i$ in either IX or SIX mode.
Multiple Granularity Cont

5. $T_i$ can lock a node only if it has not previously unlocked any node (that is, $T_i$ is two-phase).

6. $T_i$ can unlock a node $Q$ only if none of the children of $Q$ are currently locked by $T_i$.

- Explicit Lock
- Implicit Lock – Lower in the hierarchy

- **Locking**: root-to-leaf order,
- **Unlocking**: locks are released in leaf-to-root order.

- **Lock granularity escalation**: in case there are too many locks at a particular level, switch to higher granularity S or X lock.
Ex : Locks Modes

\( T_1 \rightarrow \text{read} \ a_{12} \)

\( T_2 \rightarrow \text{write} \ a_{14} \)

\( T_3 \rightarrow \text{read} \ f_1 \)

\( T_4 \rightarrow \text{read} \ d \)

\( T_1, T_2 \rightarrow \text{concurrently} \)

\( T_1, T_3, T_4 \rightarrow \text{concurrently} \)

\( T_2, T_3 \rightarrow \text{not concurrently} \)
Lock based protocol

• All the data items must be accessed in a mutually exclusive manner, i.e. when one transaction is executing then no other transaction should interrupt the same object.

• 2 types of locks.
  
  – *Shared (S) mode*: Data item can only be read. S-lock is requested using lock-S instruction
  
  – *Exclusive (X) mode*: Data item can be both read as well as written. X-lock is requested using lock-X instruction.

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>X</td>
<td>No</td>
<td>NO</td>
</tr>
</tbody>
</table>
Lock based protocol (CONT)

- A serial schedule: there is no problem to execute T1->T2 or T2->T1.

<table>
<thead>
<tr>
<th>T1: A-&gt;B (Transferring 50 rupees to B)</th>
<th>T2: Display (B+A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock-x(A)</td>
<td>Lock-s(B)</td>
</tr>
<tr>
<td>Read(A)</td>
<td>Read(B)</td>
</tr>
<tr>
<td>A=A-50</td>
<td>Unlock(B)</td>
</tr>
<tr>
<td>write(A)</td>
<td>Lock-s(A)</td>
</tr>
<tr>
<td>Unlock(A)</td>
<td>Read(A)</td>
</tr>
<tr>
<td>Lock-x(B)</td>
<td>Unlock(A)</td>
</tr>
<tr>
<td>Read(B)</td>
<td>Display(A+B)</td>
</tr>
<tr>
<td>Lock-x(B)</td>
<td>Unlock(A)</td>
</tr>
<tr>
<td>B=B+50</td>
<td></td>
</tr>
<tr>
<td>write(B)</td>
<td></td>
</tr>
<tr>
<td>Unlock(B)</td>
<td></td>
</tr>
</tbody>
</table>
Problem 1: Inconsistency in DB

- If we try to interleave the transactions then we will face the problems.
- Let's schedule $A=100$ and $B=200$ so in total we should have 300 always.
- But as we see that after till **Display (B+A)** we have $A=50$ and $B=200$ to total=250 but it should be 300 to make consistency.

<table>
<thead>
<tr>
<th>T1: A-&gt;B (Transferring 50 rupees to B)</th>
<th>T2: Display (B+A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock-x(A)</td>
<td>Lock-s(B)</td>
</tr>
<tr>
<td>Read(A)</td>
<td>Read(B)</td>
</tr>
<tr>
<td>A=A-50</td>
<td>Unlock(B)</td>
</tr>
<tr>
<td>write(A)</td>
<td>Unlock(A)</td>
</tr>
<tr>
<td>Unlock(A)</td>
<td>Display(A+B)</td>
</tr>
</tbody>
</table>

Read(B)
Lock-x(B)
B=B+50
write(B)
Unlock(B)
Problem 1: Inconsistency in DB

• Example of a transaction performing locking:

  T2: lock-S(A);
  read (A);
  unlock(A);  
  lock-S(B);
  read (B);
  unlock(B);
  display(A+B)

• Locking as above is not sufficient to guarantee serializability
  – if A and B get updated in-between the read of A and B, the displayed sum would be wrong.

• A locking protocol is a set of rules followed by all transactions while requesting and
Problem 2: Deadlock

• Let’s say T2 is waiting for T1 to unlock (A) and T1 is waiting for T2 to unlock (B)

<table>
<thead>
<tr>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock-X(A)</td>
<td>Lock-x(B)</td>
</tr>
<tr>
<td>Lock-x(B)</td>
<td>Lock-x(A)</td>
</tr>
</tbody>
</table>
Two phase locking Protocol

- Ensures conflict-serializable schedules.

**Phase 1: Growing Phase**
- transaction may obtain locks
- transaction may not release locks

**Phase 2: Shrinking Phase**
- transaction may release locks
- transaction may not obtain locks

The protocol assures serializability. It can be proved that the transactions can be serialized in the order of their lock points (i.e. the point where a transaction acquired its final lock).

highlighted parts are the last locks taken. serial schedule could be: **T2 -> T3 -> T1**
2PL locking Protocol

Two-Phase Locking Protocol: Requires both Locks and Unlocks being done in two phases. (2PL)

Phases:

Growing (Expanding) Phase:
New locks on items can be acquired.

Locked Phase:
Locks are acquired.

Lock Point:
Final lock are acquired.

Shrinking phase:
Existing locks, but no new lock can be acquired.

Releasing of locks onwards.

Locked Phase.
Variation of 2PL locking Protocol

**Conservative (Static) 2PL**
- Acquire all lock before it starts.
- Release all locks after commit.
- Avoids Cascading Rollback.
- Deadlock Free.

**Strict 2PL**
- Exclusive lock can't be released until commit.
- Helps in Cascadeless Schedule.
- Deadlock May Occur.

**Rigorous 2PL**
- Shared/Exclusive can't be released until commit.
- AVOIDS Cascading Rollback.
Variation of 2PL locking Protocol

• There are some variants cascading rollback using which we can avoid cascading rollback.

• Strict 2 phase locking Protocol: All exclusive mode locks are taken by a transaction must be unlocked after commit. However we don’t bother about shared locks. Using this method schedule will be recoverable and cascade less.

• Rigorous 2 Phase Locking Protocol: All locks must be hold until the transaction commit.

• Note: Both of them can avoid cascading rollback but Deadlock is still possible.

• Conservative 2PL: Obtain the lock then start the transaction otherwise don’t. So hold and wait is gone. But Concurrency is gone. And also requirement of locks which will be taken in future- this decision is difficult.

• Note: So when we say a schedule is conservative then it will be Rigorous as well as Strict 2 PL but vice versa is not true.
**Variation of 2PL locking Protocol EX1**

<table>
<thead>
<tr>
<th>T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock-s(A)</td>
</tr>
<tr>
<td>Read(A)</td>
</tr>
<tr>
<td>Lock-x(B)</td>
</tr>
<tr>
<td>Read(B)</td>
</tr>
<tr>
<td>Unlock(A)</td>
</tr>
<tr>
<td>Write(B)</td>
</tr>
<tr>
<td>Unlock(B)</td>
</tr>
</tbody>
</table>

- **2 PL:** There is growing and shrinking phase.
- **Strict 2 PL:** There is Lock-x(B) and it is unlocked before commit so no strict 2 PL.
- **Rigorous:** If it is not strict 2 PL then it can’t be Rigorous.
- **Conservative:** If it is not strict 2 PL then it can’t be conservative.
2 PL: There is growing and shrinking phase so it is 2 PL.
Strict 2 PL: Exclusive locks are unlocked after commit. So yes it is.
Rigorous: We have unlocked all the locks after commit so it is rigorous.
Conservative: We have not taken all the locks at first then start the transaction so no conservative
Variation of 2PL locking Protocol Example
Lock Escalation

• System is decreasing the granularity of your locks
• Ex: DB turning your 100 row-level locks against a table into a single table-level lock.

• **Oracle will never escalate a lock. Never.**

• The terms lock conversion and lock promotion are synonymous.
SQL Server Locking Basics

1- Brief overview of Resource Types
2- Brief overview of Lock Modes
3- Looking at Resource Type and Lock modes on live database
4- Lock Resources    Lock Modes
   a) RID   1- S    -- Shared Lock
   b) Key   2- U    -- Update
   c) Page  3- X    -- Exclusive
   d) Extent 4- I    -- Intent (IS, IX, SIX)
   e) Table 5- Sch  -- Schema (Sch-M, Sch-S)
   f) DB    6- BU   -- Bulk Update
SQL Server Lock Resources

• Can lock user data resources at row, page, or table level.

• SQL Server will attempt to acquire row-level locks, to allow the highest degree of concurrency.

• However: lock escalation
SQL Server Locking Basics

• Unit of data locked (lock resource) – such as row, page, or table
• Type of locks acquired (lock mode) – shared, exclusive, update, and so on
• **Duration of the lock** – how long the lock is held
• **Lock ownership** – the "scope" of the lock (most locks are transaction scoped)
• **Lock metadata** – how to review current locking using the Dynamic Management
  – MSSQL : View (DMV) called sys.dm_tran_locks.
Lock resources

- When SQL Server locks a row in an index, it refers to it, and displays it, as a KEY lock, but keep in mind that SQL Server locks the entire index row, not just the key column.
- In some circumstances, SQL Server can also lock ranges of index rows.
- Locks on rows in a heap table (one without a clustered index) appear as RID (Row ID) locks in the sys.dm_tran_locks view.
- SQL Server supports two kinds of KEY locks, depending on the isolation level of the current transaction.
- If the isolation level is **READ COMMITTED** or **REPEATABLE READ**, SQL Server attempts to lock the index rows it accesses while processing the query.
- If the table has a clustered index, then the data rows are at the leaf level of the index, and so row locks for any data in a table with a clustered index will always appear as KEY locks.
- If the table is a heap, SQL Server might acquire KEY locks for the non-clustered index rows and RID locks for the data rows.
Lock duration

- **Default: Read Committed**
- SQL Server releases S locks as soon as it has read and processed the locked data.
- It holds an X lock until the end of the transaction, whether the transaction is committed or rolled back.
- It holds a U lock until the end of the transaction, unless it promoted the U lock to an X lock as with all X locks, remains for the duration of the transaction.
Lock Escalation : SQL Server

• Escalation based on SQL Server instance resource usage
• Escalation based on number of locks held by a single statement
• *Controlling Locking*, we can force SQL Server to change the default granularity of its locks with hints or index options.
Escalation based on SQL Server instance resource Usage

• In some cases, acquiring individual locks on rows may end up consuming too much of SQL Server's memory.

• Although the memory required for each lock is quite small (about 96 bytes per lock), this still adds up to a sizeable portion of the total available memory, when thousands of locks are acquired.

• When SQL Server ends up using more than 24% of its buffer pool (excluding AWE memory) to keep track of locks acquired and lock requests waiting, it will choose any session holding locks and escalate its fine-grained (row or page) locks into a table lock.
Lock Escalation : Microsoft SQL Server Lock Hierarchy

- Always have a Shared Lock (S) on DB level.

- When your query is connected to a DB (USE MyDatabase), Shared Lock prevents the dropping of the DB, or that backups are restored over that database.

- You have locks on the table, on the pages, and the records when you are performing an operation.
SQL Server Lock Escalation

• In DML: Intent Exclusive or Update Lock (IX or IU) on the table and page level, and a Exclusive or Update Lock (X or U) on the changed records.

• SQL Server always acquires locks from top to bottom to prevent Race Conditions, when multiple threads trying to acquire locks concurrently within the locking hierarchy.
SQL Server Lock Escalation

- DELETE operation on a table against 20,000 rows.

- Let’s assume that a row is 400 bytes long, means that 20 records fit onto one page of 8kb:
SQL Server Lock Escalation

- One S Lock on the database, 1 IX Lock on the table, 1,000 IX locks on the pages (20,000 records are spread across 1,000 pages), and you have finally 20,000 X locks on the records itself.

- In sum you have acquired 21,002 locks for the DELETE operation.
- Every lock needs in SQL Server 96 bytes of memory, so we look at 1.9 MB of locks just for 1 simple query.

- This will not scale indefinitely when you run multiple queries in parallel.
- For that reason SQL Server implements now the so-called Lock Escalation.
SQL Server Lock Escalation

- For more than 5,000 locks on one level in your locking hierarchy, SQL Server escalates into a simple coarse-granularity lock.
- SQL Server will by default *always* escalate directly to the table level.
- An escalation policy to the page level just doesn’t exist
- One Exclusive Lock (X) on the table level. Concurrency of your database in a very negative way
- No other session is able any more to access that table – every other query will just block
Controlling Locking: SQL Server Lock Escalation

• Since SQL Server 2008 you can also control how SQL Server performs the Lock Escalation – through the ALTER TABLE statement and the property LOCK_ESCALATION.

• 3 different options:
  • **TABLE**: Always performs the Lock Escalation to the table level
  • **AUTO**: Lock Escalation is performed to the partition level, if the table is partitioned, and otherwise to the table level.
  • **DISABLE**: Disable Lock Escalation for that specific table. Lock Manager of SQL Server can then consume a huge amount of memory. **Not Recommended !!!**

```sql
1 -- Controlling Lock Escalation
2 ALTER TABLE Person.Person
3 SET
4 (  
5     LOCK_ESCALATION = AUTO  -- or TABLE or DISABLE
6 )
7 GO
```
Escalation based on number of locks held by a single Statement

• SQL Server will also escalate locks when any individual session acquires more than 5,000 locks in a single statement.

• In this case, there is no randomness in choosing which session will get its locks escalated; it is the session that acquired the locks.
Creation of the DBlocks view to display locks in the current database: SQL Server

```sql
IF EXISTS ( SELECT 1
    FROM sys.views
    WHERE name = 'DBlocks' )
    DROP VIEW DBlocks ;
GO
CREATE VIEW DBlocks AS
SELECT request_session_id AS spid ,
    DB_NAME(resource_database_id) AS dbname ,
    CASE WHEN resource_type = 'OBJECT'
        THEN OBJECT_NAME(resource_associated_entity_id)
        WHEN resource_associated_entity_id = 0 THEN 'n/a'
        ELSE OBJECT_NAME(p.object_id)
    END AS entity_name ,
    index_id ,
    resource_type AS resource ,
    resource_description AS description ,
    request_mode AS mode ,
    request_status AS status
FROM sys.dm_tran_locks t
LEFT JOIN sys.partitions p
    ON p.partition_id = t.resource_associated_entity_id
WHERE resource_database_id = DB_ID()
AND resource_type <> 'DATABASE' ;
```
Escalation based on number of locks held by a single Statement

• First, we start a transaction and perform two UPDATE statements.
• Together, the two statements acquire more than 5,000 locks, but neither one, individually, acquires that many. Lock escalation does not occur.
• DBlocks query should reveal that the number of X locks held by the connection is 6,342.
Escalation based on number of locks held by a single Statement

- In the second transaction, we update the same 6,342 rows in a single statement. In this case, the DBlocks query reveals that the total number of locks held, right before the end of the transaction, is only one (a table lock). DBlocks query should reveal that the number of X locks held by the connection is 6,342.
Latches

• However, latches do not show up in the sys.dm_tran_locks view.

• Latches are used to protect an internal structure for brief periods while it is being read or modified, not to ensure correct transaction behavior.

• Both the data page itself and the buffer that the data is occupying are protected by latches.

• Latches protect the *physical* integrity of the data; locks protect its *logical* integrity.
Locking Examples

• Example 1: SELECT with READ COMMITTED isolation level
• Example 2: SELECT with REPEATABLE READ isolation level
• Example 3: SELECT with SERIALIZABLE isolation level
• Example 4: Update with READ COMMITTED isolation level
• Example 5: Update with SERIALIZABLE isolation level (with an index)
• Example 6: Update with SERIALIZABLE isolation level not using an index
• Example 7: Creating a table
• Example 8: RID locks
Creation of the DBlocks view to display locks in the current database: SQL Server

```sql
IF EXISTS ( SELECT 1
    FROM sys.views
    WHERE name = 'DBlocks' )
    DROP VIEW DBlocks ;
GO
CREATE VIEW DBlocks AS
SELECT request_session_id AS spid ,
    DB_NAME(resource_database_id) AS dbname ,
    CASE WHEN resource_type = 'OBJECT'
        THEN OBJECT_NAME(resource_associated_entity_id)
        WHEN resource_associated_entity_id = 0 THEN 'n/a'
        ELSE OBJECT_NAME(p.object_id)
    END AS entity_name ,
    index_id ,
    resource_type AS resource ,
    resource_description AS description ,
    request_mode AS mode ,
    request_status AS status
FROM sys.dm_tran_locks t
LEFT JOIN sys.partitions p
    ON p.partition_id = t.resource_associated_entity_id
WHERE resource_database_id = DB_ID() AND resource_type <> 'DATABASE' ;
```
<table>
<thead>
<tr>
<th>Column Name</th>
<th>Data Type</th>
<th>Allow Nulls</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProductID</td>
<td>int</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Nvarchar(50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProductNumber</td>
<td>Nvarchar(25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MakeFlag</td>
<td>Flag(bit)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FinishedGoodsFlag</td>
<td>Flag(bit)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>Nvarchar(15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SafetyStockLevel</td>
<td>Smallint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReorderPoint</td>
<td>Smallint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>StandardCost</td>
<td>Money</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ListPrice</td>
<td>Money</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>Nvarchar(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SizeUnitMeasureCode</td>
<td>Nvarchar(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WeightUnitMeasureCode</td>
<td>Nvarchar(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>Decimal(8, 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DaysToManufacture</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProductLine</td>
<td>Nvarchar(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>Nvarchar(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Style</td>
<td>Nvarchar(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProductSubcategoryID</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProductModelID</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SellStartDate</td>
<td>Datetime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SellEndDate</td>
<td>Datetime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DiscontinuedDate</td>
<td>Datetime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rowguid</td>
<td>Uniqueidentifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ModifiedDate</td>
<td>Datetime</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
[Production].[Product] table
Example 1: SELECT with READ COMMITTED isolation level

```sql
USE AdventureWorks2014;
SET TRANSACTION ISOLATION LEVEL READ COMMITTED;
BEGIN TRAN
SELECT *
FROM Production.Product
WHERE Name = 'Reflector';
SELECT *
FROM DBlocks
WHERE spid = @@spid;
COMMIT TRAN
```

- By default, SQL Server releases S locks as soon as it has finished reading the data.
- By the time we execute the SELECT from the view, SQL Server no longer holds the locks.
- As such OBJECT lock on the view (there is also a DATABASE lock on the AdventureWorks database, but the DBlocks view filtered out database locks).

<table>
<thead>
<tr>
<th>spid</th>
<th>dbname</th>
<th>entity_name</th>
<th>index_id</th>
<th>resource</th>
<th>description</th>
<th>mode</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AdventureWorks</td>
<td>DBlocks</td>
<td>NULL</td>
<td>OBJECT</td>
<td>IS</td>
<td>GRANT</td>
<td></td>
</tr>
</tbody>
</table>
Example 2: SELECT with REPEATABLE READ isolation level

```
USE AdventureWorks2014;
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ;
BEGIN TRAN
SELECT *
FROM Production.Product
WHERE Name LIKE 'Racing Socks%';
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'Product';
COMMIT TRAN
```

• This time, because the transaction isolation level is REPEATABLE READ, SQL Server holds the S locks until the transaction is finished and so we can see them in our results.
• Production.Product table has a clustered index, so the rows of data are all index rows in the leaf level.
• As such, output shows that the locks on the two individual data rows returned are KEY locks.
• The table also has a non-clustered index on the Name column and we can see two KEY locks at the leaf level of this non-clustered index, used to find the relevant rows.
• We can distinguish the clustered and non-clustered indexes by the value in the Index_ID column: the data rows have an Index_id value of 1, and the non-clustered index rows have an Index_ID value of 3.
Example 3: SELECT with SERIALIZABLE isolation level

USE AdventureWorks;
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
BEGIN TRAN
SELECT *
FROM Production.Product
WHERE Name LIKE 'Racing Socks%';
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'Product';
COMMIT TRAN
Example 3: SELECT with SERIALIZABLE isolation level

- S-mode KEY locks on the rows in the clustered index, and IS locks on the parent pages and object.
- The two-part mode RangeS-S indicates a key-range lock in addition to the lock on the key itself.

1. First part (RangeS) is the lock on the range of keys between and including the key holding the lock and the previous key in the index.

2. Key-range locks prevent other transactions from inserting any new rows into the table that meet the condition of this query; that is, it's not possible to insert any new rows with a product name starting with Racing Socks.

3. Key-range locks are held on ranges in the non-clustered index on Name (Index_id = 3) because that is the index used to find the qualifying rows.

<table>
<thead>
<tr>
<th>spid</th>
<th>dbname</th>
<th>entity_name</th>
<th>index_id</th>
<th>resource</th>
<th>description</th>
<th>mode</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>NULL</td>
<td>OBJECT</td>
<td>1:9057</td>
<td>IS</td>
<td>GRANT</td>
</tr>
<tr>
<td>2</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>1</td>
<td>PAGE</td>
<td>(6b00b8eeda30)</td>
<td>IS</td>
<td>GRANT</td>
</tr>
<tr>
<td>3</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>1</td>
<td>KEY</td>
<td>(6a00dd896688)</td>
<td>S</td>
<td>GRANT</td>
</tr>
<tr>
<td>4</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>1</td>
<td>KEY</td>
<td>(9502d56a217e)</td>
<td>S</td>
<td>GRANT</td>
</tr>
<tr>
<td>5</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>3</td>
<td>PAGE</td>
<td>1:4606</td>
<td>IS</td>
<td>GRANT</td>
</tr>
<tr>
<td>6</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>3</td>
<td>KEY</td>
<td>(23027a50f6db)</td>
<td>RangeS-S</td>
<td>GRANT</td>
</tr>
<tr>
<td>7</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>3</td>
<td>KEY</td>
<td>(9602945b3a67)</td>
<td>RangeS-S</td>
<td>GRANT</td>
</tr>
<tr>
<td>8</td>
<td>AdventureWorks</td>
<td>Product</td>
<td>3</td>
<td>KEY</td>
<td></td>
<td>RangeS-S</td>
<td>GRANT</td>
</tr>
</tbody>
</table>
Example 3: SELECT with SERIALIZABLE isolation level

- The two Racing Socks rows are Racing Socks, L and Racing Socks, M.

- 3 KEY locks in the non-clustered index because SQL Server must lock three different ranges of data:
  1. Range from the key preceding the first Racing Socks row in the index (Pinch Bolt) up to the first Racing Socks row (Racing Socks, L)
  2. Range between the two rows starting with Racing Socks.
  3. The range from the second Racing Socks row (Racing Socks, M) to the next key in the index (Rear Brakes).
Example 4: Update with READ COMMITTED isolation level

```sql
USE AdventureWorks;
SET TRANSACTION ISOLATION LEVEL READ COMMITTED;
BEGIN TRAN
UPDATE Production.Product
SET ListPrice = ListPrice * 0.6
WHERE Name LIKE 'Racing Socks%';
SELECT * 
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'Product';
COMMIT TRAN
```

- 2 rows in the leaf level of the clustered index are locked with X locks.
- Page and table are then locked with IX locks.
Example 5: Update with SERIALIZABLE isolation level (with an index)

```sql
USE AdventureWorks;
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
BEGIN tran;
UPDATE Production.Product
SET ListPrice = ListPrice * 0.6
WHERE Name LIKE 'Racing Socks%';
SELECT * FROM DBlocks
WHERE spid = @@spid AND entity_name = 'Product';
COMMIT tran;
```

- 2 rows in the leaf level of the clustered index are locked with X locks.
- Page and table are then locked with IX locks.
Example 5: Update with SERIALIZABLE isolation level (with an index)

- Key-range locks are on the non-clustered index, used to find the relevant rows.
- Range interval itself needs only an S lock to prevent insertions, but the searched keys have U locks, ensuring that no other process can attempt to UPDATE them.
- The keys in the table itself (index_id = 1) obtain the X lock when the actual modification is made.
Example 6: Update with SERIALIZABLE isolation level not using an index

```
USE AdventureWorks;
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
BEGIN TRAN
UPDATE Production.Product
SET ListPrice = ListPrice * 0.6
WHERE Color = 'White' ;
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'Product' ;
COMMIT TRAN
```
Example 6: Update with SERIALIZABLE isolation level not using an index

- As there was no useful index, a clustered index scan on the entire table was required, and so all keys initially received the RangeS-U lock;
- Four rows were eventually modified, the locks on those keys escalated to the RangeX-X lock.
- Two of the RangeX-X locks, and a few of the RangeS-U locks.
- Complete output has 501 RangeS-U locks, as well as IU locks on several pages, IX locks on two pages, and an IX lock on the table.

```
SELECT * from Production.Product
WHERE Color = 'White';
```
Example 7: Creating a table

```
USE AdventureWorks;
SET TRANSACTION ISOLATION LEVEL READ COMMITTED;
BEGIN TRAN
SELECT * INTO newProducts
FROM Production.Product
WHERE ListPrice BETWEEN 1 AND 10;
SELECT * FROM DBlocks
WHERE spid = @@spid;
COMMIT TRAN
```

- When creating the new table, SQL Server acquires locks on six different system tables to record information about new table.
- In addition schema modification (Sch-M) locks
Example 8: RID locks

USE AdventureWorks;
SET TRANSACTION ISOLATION LEVEL READ COMMITTED
BEGIN TRAN
UPDATE newProducts
SET ListPrice = 5.99
WHERE Name = 'Road Bottle Cage';
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'newProducts'
COMMIT TRAN

- There are no indexes on the newProducts table
- X lock on the row (RID). For RID locks: File Number:Page number:Slot number.
- As expected, SQL Server takes IX locks on the page and the table.
### SQL Server Lock compatibility matrix.

<table>
<thead>
<tr>
<th>Requested lock mode</th>
<th>Existing granted lock mode</th>
<th>IS</th>
<th>S</th>
<th>U</th>
<th>IX</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intent shared (IS)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Shared (S)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Update (U)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Intent exclusive (IX)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Exclusive (X)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Exclusive locks

• SQL Server automatically acquires exclusive (X) locks on data in order to modify that data, during an INSERT, UPDATE, or DELETE operation.

• Only one transaction at a time can hold an X lock on a particular data resource, and X locks remain until the end of the transaction.

• The changed data is usually unavailable to any other process until the transaction holding the lock either commits or rolls back.

• READ UNCOMMITTED transaction isolation level, can read data exclusively locked by another transaction.
Intent locks

- you can have intent shared (IS) locks, intent exclusive locks (IX), and even intent update locks (IU), indicated in the request_mode column of the sys.dm_tran_locks view by IS, IX and IU, respectively.

- SQL Server can acquire locks at different levels of granularity (i.e., at the row, page, or table level), and so needs some mechanism that signals whether a component of a resource is already locked.

- Example: if one transaction attempts to lock a table, SQL Server must be able to determine whether it has already locked a row or a page of that table. Intent locks serve that purpose.
Intent locks

- Whenever a transaction acquires a lock at a lower level of granularity, it also acquires higher-level intent locks for the same object.

- For example, a transaction that holds an X lock on a row in the Customers table will also hold IX locks on both the page containing that row, and the Customers table.

- These Intent locks will prevent another transaction from locking the entire Customers table (acquiring an X lock on the table).
• Example: Escalate dynamically to coarse locks when memory usage for LCBs becomes critical
Lock conversion from S to X.

USE AdventureWorks2014;
-- Create a new table
IF OBJECTPROPERTY(OBJECT_ID('NewOrders'), 'IsUserTable') = 1
DROP TABLE NewOrders;
GO
SELECT *
INTO NewOrders
FROM Sales.SalesOrderHeader;
GO
CREATE UNIQUE INDEX NewOrder_index ON NewOrders(SalesOrderID);
GO
Lock conversion from S to X.

```sql
-- Change isolation level and start transaction
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ;
BEGIN TRAN
-- SELECT data and examine the locks
SELECT *
FROM NewOrders
WHERE SalesOrderID = 55555;
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
-- UPDATE data and examine the locks
UPDATE NewOrders
SET SalesPersonID = 277
WHERE SalesOrderID = 55555;
```

<table>
<thead>
<tr>
<th>SalesOrderID</th>
<th>RevisionNumber</th>
<th>OrderDate</th>
<th>DueDate</th>
<th>ShipDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>55555</td>
<td>8</td>
<td>2013-09-03 00:00:00.0000</td>
<td>2013-09-15 00:00:00.0000</td>
<td>2013-09-10 00:00:00.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>spid</th>
<th>dbname</th>
<th>entity_name</th>
<th>index_id</th>
<th>resource</th>
<th>description</th>
<th>mode</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>AdventureWorks2014</td>
<td>NewOrders</td>
<td>0</td>
<td>RID</td>
<td>1.24738:25</td>
<td>S</td>
<td>GRANT</td>
</tr>
<tr>
<td>79</td>
<td>AdventureWorks2014</td>
<td>NewOrders</td>
<td>2</td>
<td>KEY</td>
<td>(fc81b91b4150)</td>
<td>S</td>
<td>GRANT</td>
</tr>
<tr>
<td>79</td>
<td>AdventureWorks2014</td>
<td>NewOrders</td>
<td>0</td>
<td>PAGE</td>
<td>1.24738</td>
<td>IS</td>
<td>GRANT</td>
</tr>
<tr>
<td>79</td>
<td>AdventureWorks2014</td>
<td>NewOrders</td>
<td>2</td>
<td>PAGE</td>
<td>1.26549</td>
<td>IS</td>
<td>GRANT</td>
</tr>
<tr>
<td>79</td>
<td>AdventureWorks2014</td>
<td>NewOrders</td>
<td>NULL</td>
<td>OBJECT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lock conversion from S to X.

- Open REPEATABLE READ transaction and selects one row from the NewOrders table.
- When querying our DBlocks view we see an S lock on a RID in the database. (SalesOrderID = 55555)
- S lock is held for a key in the non-clustered index.
- Also IS locks for the page in the table that contains the selected row, and for the page in the index that contains the key for the selected row as well as an IS lock on the table itself.
Lock conversion from S to X.

- After we update the row, we query the DBlocks view again, and this time there are different locks on the same resources.
- RID now has an X lock, and the page in the table, and the table itself, both have IX locks.
- Key in the index has a U lock. SQL Server acquired the U lock while searching for the row to update, and this is necessary because, until the modification to the data row happens.
- SQL Server doesn't know whether the modification will also require a change to the index.
- The page in the index containing the key has an IU lock.
Chapter 10: Implementation and Pragmatic Issues

- 10.2 Data Structures of a Lock Manager
- 10.3 Multi-Granularity Locking and Lock Escalation
- **10.4 Transient Versioning**
- 10.5 Nested Transactions for Intra-transaction parallelism
- 10.6 Tuning Options
- 10.7 Overload Control
- 10.8 Lessons Learned
Storage Organization for Transient Versioning

• update on current data moves old version to version pool
• read-only transactions follow version chains
• old versions are kept sorted by their successor timestamps
  → garbage collection simply advances begin pointer
Row Versioning: SQL Server

- With optimistic transaction isolation levels, when updates occur, SQL Server stores the old versions of the rows in a special part of tempdb called *version store*.
- The original rows in the database are also updated with 14-byte pointers that reference the old versions of the rows.
- Depending on the situation, you can have more than one version of the records stored for the row.
Row Versioning: SQL Server Cont

- Now when readers (and sometimes writers) access the row that holds exclusive (X) lock, they get the old version.
Tuning Repertoire

- Manual locking (or manual preclaiming)
- Choice of SQL isolation level(s)
- Application structuring towards short transactions
- MPL control
Definition 10.1 (Isolation Levels):

- A schedule \( s \) runs under isolation level **read uncommitted** (aka. dirty read or browse mode) if write locks are subject to S2PL.
- A schedule \( s \) runs under isolation **read committed** (aka. cursor stability) if write locks are subject to S2PL and read locks are held for the duration of an SQL operation.
- A schedule \( s \) runs under isolation level **serializability** if it can be generated by S2PL.
- A schedule \( s \) runs under isolation level **repeatable read** if all anomalies other than phantoms are prevented.

**Remark:** A scheduler can use different isolation levels for different transactions.

**Observation:** read committed is susceptible to lost updates

**Example:** \( r_1(x) \ r_2(x) \ w_2(x) \ c_2 \ w_1(x) \ c_1 \)
**SQL Server Isolation level**

- **READ UNCOMMITTED** least restrictive isolation level because it ignores locks placed by other transactions. Transactions executing under READ UNCOMMITTED can read modified data values that have not yet been committed by other transactions; these are called "dirty" reads.

- **READ COMMITTED** is the default isolation level for SQL Server. It prevents dirty reads by specifying that statements cannot read data values that have been modified but not yet committed by other transactions. Other transactions can still modify, insert, or delete data between executions of individual statements within the current transaction, resulting in non-repeatable reads, or "phantom" data.

- **REPEATABLE READ** is a more restrictive isolation level than READ COMMITTED. It encompasses READ COMMITTED and additionally specifies that no other transactions can modify or delete data that has been read by the current transaction until the current transaction commits. Concurrency is lower than for READ COMMITTED because shared locks on read data are held for the duration of the transaction instead of being released at the end of each statement.

- **SERIALIZABLE** is the most restrictive isolation level, because it locks entire ranges of keys and holds the locks until the transaction is complete. It encompasses REPEATABLE READ and adds the restriction that other transactions cannot insert new rows into ranges that have been read by the transaction until the transaction is complete.
### Which isolation levels permit which behaviors?

#### SQL Server

<table>
<thead>
<tr>
<th>Transaction Isolation Level</th>
<th>Behaviors Allowed</th>
<th>Concurrency Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dirty Read</td>
<td>Non-repeatable Read</td>
</tr>
<tr>
<td>READ UNCOMMITTED</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>READ COMMITTED</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>(default for SQL Server)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>REPEATABLE READ</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SNAPSHOT</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SERIALIZABLE</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**Concurrency Model**
- Pessimistic
- Optimistic
-- Create a database and table for testing the isolation levels
USE master
GO
IF EXISTS ( SELECT 1
FROM sys.databases
WHERE name = 'IsolationDB' )
DROP DATABASE IsolationDB ;
GO
CREATE DATABASE IsolationDB ;
GO
USE IsolationDB ;
GO
CREATE TABLE IsolationTest
(    col1 INT PRIMARY KEY ,
    col2 VARCHAR(20) ) ;
GO
INSERT INTO IsolationTest
VALUES ( 10, 'The first row' ) ;
INSERT INTO IsolationTest
VALUES ( 20, 'The second row' ) ;
INSERT INTO IsolationTest
VALUES ( 30, 'The third row' ) ;
INSERT INTO IsolationTest
VALUES ( 40, 'The fourth row' ) ;
INSERT INTO IsolationTest
VALUES ( 50, 'The fifth row' ) ;
GO
MSSQL Isolation level : Read Uncommitted
Dirty read example

```
t1
```

```
-- Step 1:
-- Start a transaction but don't commit it
USE IsolationDB ;
GO
BEGIN TRAN
UPDATE IsolationTest
SET col2 = 'New Value' ;
--<EXECUTE>
```

```
t2
```

```
-- Step 2:
-- Start a new connection and change your isolation level
USE IsolationDB ;
GO
SET TRANSACTION ISOLATION LEVEL READ UNCOMMITTED ;
SELECT * 
FROM IsolationTest ;
--<EXECUTE>
```

All the values in col2 are the string 'New Value', even though the transaction in the first connection has not yet committed.
MSSQL Isolation level : Read Uncommitted
Dirty read example

$t^3$
Back in the first connection, roll back the transaction, as shown in Step 3

$t^4$
For Step 4, rerun the SELECT statement in the second connection to see that all the values in col2 have reverted to their original values
Dirty reads problems

- If a query running under READ UNCOMMITTED isolation level performs a scan of a table it is possible that a separate transaction could update a row of data, causing that row to move to a new location.

- If the scan started before the update and read the initial version of the row, the row might move to a page not yet read, and the query could end up reading it again, later on in the same scan.
MSSQL Isolation level : Read committed (Default) Blocking Example

**t1** isolation level is the default READ COMMITTED, and then it will start a transaction that reads data from the IsolationTest table to compute an average.

```
-- Step 1:
-- Start a transaction but don't commit it
USE IsolationDB;
GO
BEGIN TRAN
UPDATE IsolationTest
SET col2 = 'New Value';
```

```
(5 row(s) affected)
```

**t2** Step 2 will UPDATE the table. Assuming that the query in Step 1 has finished processing, the UPDATE will succeed, even though the first connection is still inside a transaction. Note that the UPDATE is an auto-commit transaction and so SQL Server will commit the UPDATE and release the locks as soon as it completes.

```
-- Step 2:
-- Start a new connection and change your isolation level
USE IsolationDB;
GO
SET TRANSACTION ISOLATION LEVEL READ COMMITTED;
SELECT *
FROM IsolationTest;
```

```
-- You should notice that the process blocks, and returns
-- no data or messages!
```

A SELECT statement blocking with READ COMMITTED isolation level.
MSSQL Isolation level: Read committed (Default)

Blocking Example

-- Step 1:
-- Start a transaction but don't commit it
USE IsolationDB;
GO
BEGIN TRAN
UPDATE IsolationTest
SET col2 = 'New Value';

--<EXECUTE>

-- To finish up, perform the following two steps:
-- Step 3:
-- Return to the connection from Step 1 and issue a ROLLBACK
ROLLBACK TRANSACTION;
--<EXECUTE>

-- Step 2:
-- Start a new connection and change your isolation level
USE IsolationDB;
GO
SET TRANSACTION ISOLATION LEVEL READ COMMITTED;
SELECT * FROM IsolationTest;

--<EXECUTE>
-- You should notice that the process blocks, and returns
-- no data or messages!

-- Step 4:
-- Rerun the SELECT statement in the connection from Step 2
SELECT * FROM IsolationTest;

--<EXECUTE>
-- Verify that the data is available
MSSQL Isolation level: Read committed (Default)  
non-repeatable reads example

**t1**

Make sure the isolation level is the default READ COMMITTED, and then it will start a transaction that reads data from the Isolation Test table to compute an average.

```sql
-- Step 1:
-- Read data in the default isolation level
USE IsolationDB
SET TRANSACTION ISOLATION LEVEL READ COMMITTED
BEGIN TRAN
SELECT AVG(col1)
FROM IsolationTest;
--<EXECUTE>
```

**t2**

In the second connection, Step 2 will UPDATE the table. Assuming that the query in Step 1 has finished processing, the UPDATE will succeed, even though the first connection is still inside a transaction. Note that the UPDATE is an auto-commit transaction and so SQL Server will commit the UPDATE and release the locks as soon as it completes.

```sql
-- Step 2:
-- In a new connection, update the table:
USE IsolationDB;
UPDATE IsolationTest
SET col1 = 500
WHERE col1 = 50;
--<EXECUTE>
```
In Step 3, return to the first connection and run the same SELECT statement. The average value is now different and we have a non-repeatable read. The default READ COMMITTED isolation level prevents other connections from reading data being modified, but only prevents other connections from changing data being read, while the read operation is in progress. Once it is complete, other transactions can change the data, even if the reading transaction is still open. As a result, there is no guarantee that we'll see the same data if we rerun the SELECT within the transaction.

---

--- Step 1:
-- Read data in the default isolation level

USE IsolationDB
SET TRANSACTION ISOLATION LEVEL READ COMMITTED ;
BEGIN TRAN
SELECT AVG(col1) FROM IsolationTest ;
--<EXECUTE>

--- Step 3:
-- Go back to the first connection and run the same SELECT statement:

SELECT AVG(col1) FROM IsolationTest ;
--<EXECUTE>

--- Step 2:
-- In a new connection, update the table:

USE IsolationDB ;
UPDATE IsolationTest
SET col1 = 500
WHERE col1 = 50 ;
--<EXECUTE>

--- Step 4:
-- issue a ROLLBACK
ROLLBACK TRANSACTION ;
--<EXECUTE>
READ COMMITTED and non repeatable reads

• READ COMMITTED isolation level prevents other connections from reading data being modified, but only prevents other connections from changing data being read, while the read operation is in progress.
• Once it is complete, other transactions can change the data, even if the reading transaction is still open.
• As a result, there is no guarantee that we'll see the same data if we rerun the SELECT within the transaction.
READ_COMMITTED_SNAPSHOT

• As one would expect, its behavior is the same as the default in terms of the read phenomena, i.e. it prevents dirty reads

• Allows non-repeateable reads and phantom reads. However, this optimistic implementation of the READ COMMITTED level prevents dirty reads, *without blocking other* transactions.
**READ_COMMITTED_SNAPSHOT**

_t1_

```sql
-- Step 1:
-- First close all other connections to make sure no one is using
-- the IsolationDB database
-- Step 2:
-- Change the database option to enable "read committed snapshot"

ALTER DATABASE [IsolationDB] SET READ_COMMITTED_SNAPSHOT ON ;

--<EXECUTE>
```

_t2_

```sql
-- Step 3:
-- Start a transaction but don't commit it

USE IsolationDB ;
GO
BEGIN TRAN
UPDATE IsolationTest
SET col2 = 'New Value' ;
--<EXECUTE>
```

10-99
-- Step 6:
-- Now close all other connections to make sure no one is using the IsolationDB database:

```sql
-- Start a transaction but don't commit it
USE IsolationDB;
GO
BEGIN TRAN
UPDATE IsolationTest
SET col2 = 'New Value';
--<EXECUTE>
-- To finish up, perform the following steps:
-- Step 5:
-- Return to the connection from Step 1 and issue a ROLLBACK
ROLLBACK TRANSACTION;
--<EXECUTE>
```
READ_COMMITTED_SNAPSHOT

Step 7:
-- Change the database option to disable "read committed snapshot"

```
ALTER DATABASE [IsolationDB] SET READ_COMMITTED_SNAPSHOT OFF ;
```

--<EXECUTE>

Command(s) completed successfully.
REPEATABLE READ

• REPEATABLE READ isolation level adds to the properties of READ COMMITTED by ensuring that if a transaction re-reads data, or if a query is reissued within the same transaction, then the same data will be returned.

• Issuing the same query twice within a transaction won't pick up any changes to data values that were made by another transaction.

• A second transaction cannot modify the data that a first transaction has read, as long as that first transaction has not yet committed or rolled back.
REPEATABLE READ

Step 1:
- Read data in the Repeatable Read isolation level

```
USE IsolationDB;
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ;
BEGIN TRAN
SELECT AVG(col1)
FROM IsolationTest;
--<EXECUTE>
```

Step 2:
- In the second connection, update the table:

```
USE IsolationDB;
UPDATE IsolationTest
SET col1 = 5000
WHERE col1 = 500;
--<EXECUTE>
```
- You should notice that the UPDATE process blocks,
  - and returns no data or messages
**REPEATABLE READ**

```
-- Step 1:
-- Read data in the Repeatable Read isolation level
USE IsolationDB;
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ;
BEGIN TRAN
SELECT AVG(col1)
FROM IsolationTest;
--<EXECUTE>

-- Step 3:
-- Go back to the first connection and
-- run the same SELECT statement:
SELECT AVG(col1)
FROM IsolationTest;
--<EXECUTE>
```

**t3**

**t4**

*Issue a ROLLBACK*
REPEATABLE READ

• However, REPEATABLE READ isolation doesn't prevent all possible read phenomena.
• It protects only the data that has already been read.
REPEATABLE READ: phantoms

-- Close all connections and open two new ones
-- Step 1:
USE IsolationDB;
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ
BEGIN Tran
SELECT *
FROM IsolationTest
WHERE col1 BETWEEN 20 AND 40
--<EXECUTE>

-- Step 2:
-- In the second connection, insert new data
USE IsolationDB;
INSERT INTO IsolationTest
VALUES (25, 'New Row');
--<EXECUTE>
Upon the second execution of the same SELECT statement, the new row appears, called a phantom. The row didn't even exist the first time we ran the SELECT statement, so it wasn't locked. We can prevent phantoms with the SERIALIZABLE isolation level.

T4 Issue a ROLLBACK
SERIALIZABLE

• We pay a price to prevent phantoms. In addition to locking all the data has been read, enforcing the SERIALIZABLE isolation level, and so preventing phantoms, requires that SQL Server also lock data that doesn't exist (*Key-range Locks*).

• Running multiple SERIALIZABLE transactions at the same time is the equivalent of running them one at a time – that is, serially.
-- Open two new connections
-- Step 1:
-- In the first connection, start a transaction

USE IsolationDB;
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
BEGIN TRAN
SELECT *
FROM IsolationTest
WHERE col1 BETWEEN 20 AND 40;
--<EXECUTE>

-- Step 2:
-- In the second connection, insert new data
USE IsolationDB
INSERT INTO IsolationTest
VALUES (35, 'Another New Row');
-- Notice that the INSERT will block
--<EXECUTE>
-- In the first connection, start a transaction

USE IsolationDB;
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
BEGIN TRAN

SELECT *
FROM IsolationTest
WHERE col1 BETWEEN 20 AND 40;

-- <EXECUTE>

-- Step 3:
-- Go back to the first connection and rerun the SELECT

SELECT *
FROM IsolationTest
WHERE col1 BETWEEN 20 AND 40;

-- <EXECUTE>

-- Notice no new rows
SERIALIZABLE

**T4**

Rollback

```sql
USE IsolationDB;
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
BEGIN TRAN
SELECT * FROM IsolationTest WHERE col1 BETWEEN 20 AND 40;
--<EXECUTE>

-- Step 3:
-- Go back to the first connection and rerun the SELECT
SELECT * FROM IsolationTest WHERE col1 BETWEEN 20 AND 40;
--<EXECUTE>
-- Notice no new rows

-- Step 4:
-- issue a ROLLBACK
ROLLBACK TRANSACTION;
--<EXECUTE>
```

Command(s) completed successfully.
Definition 10.2 (Multiversion Read Committed and Snapshot Isolation Levels):

- A transaction runs under isolation level **multiversion read committed** if it reads the most recent committed versions as of the transaction‘s begin and uses S2PL for writes.
- A transaction runs under **snapshot isolation** if it reads the most recent versions as of the transaction‘s begin and its write set is disjoint with the write sets of all concurrent transactions.

**Observation: snapshot isolation does not guarantee MVSR**

**Example:**

\[
\begin{align*}
& r_1(x_0) \ r_1(y_0) \ r_2(x_0) \ r_2(y_0) \ w_1(x_1) \ c_1 \ w_2(y_2) \ c_2 \\
\end{align*}
\]

Possible interpretation:
- constraint \( x + y \geq 0, x_0 = y_0 = 5, \)
- \( t_1 \) subtracts 10 from \( x \), \( t_2 \) subtracts 10 from \( y \)
Multiversion Two-Phase Locking

- Multiversion schemes keep old versions of data item to increase concurrency.
  - Multiversion Timestamp Ordering
  - Multiversion Two-Phase Locking
- Each successful write results in the creation of a new version of the data item written.
- Use timestamps to label versions.
- When a read(Q) operation is issued, select an appropriate version of Q based on the timestamp of the transaction, and return the value of the selected version.
- Reads never have to wait as an appropriate version is returned immediately.
Multiversion Two-Phase Locking

- Differentiates between read-only transactions and update transactions.
- Update transactions acquire read and write locks, and hold all locks up to the end of the transaction. That is, update transactions follow rigorous two-phase locking.
  - Each successful write results in the creation of a new version of the data item written.
  - Each version of a data item has a single timestamp whose value is obtained from a counter `ts-counter` that is incremented during commit processing.
- Read-only transactions are assigned a timestamp by reading the current value of `ts-counter` before they start execution; they follow the multiversion timestamp-ordering protocol for performing reads.
Multiversion Two-Phase Locking

- When an update transaction wants to read a data item, it obtains a shared lock on it, and reads the latest version.
- When it wants to write an item, it obtains X lock on; it then creates a new version of the item and sets this version's timestamp to $\infty$.
- When update transaction Ti completes, commit processing occurs:
  - Ti sets timestamp on the versions it has created to $ts$-counter $+1$
  - Ti increments $ts$-counter by 1
- Read-only transactions that start after Ti increments $ts$-counter will see the values updated by Ti.
- Read-only transactions that start before Ti increments the $ts$-counter will see the value before the updates by Ti.
- Only serializable schedules are produced.
Advanced Locking Concepts

• Lock Compatibility
• Lock Mode Conversion
• Special Intent Locks
• Shared intent exclusive (SIX)
• Update intent exclusive (UIX)
• Shared intent update (SIU)
Advanced Locking Concepts

- **lock compatibility** – which lock types are compatible, and so can exist simultaneously on the same resource, and which are incompatible, and so will lead to blocking

- **lock mode conversion** – how SQL Server converts lock modes in response to the operations being performed by a given transaction, in order to ensure enforcement of the ACID transaction properties

- **special intent locks** – acquired when a non-intent lock is requested on a resource on which either an IX or an IU lock is already held

- **key-range locks** – a type of lock acquired in SERIALIZABLE isolation level when scanning or modifying a range of data;

- **We take a closer look at the four most common modes of key-range lock.**
### SQL Server Lock compatibility matrix.

<table>
<thead>
<tr>
<th>Requested lock mode</th>
<th>IS</th>
<th>S</th>
<th>U</th>
<th>IX</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intent shared (IS)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Shared (S)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Update (U)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Intent exclusive (IX)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Exclusive (X)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Three more lock modes

• **Sch-S: schema stability lock**
  1. SQL Server acquires a Sch-S lock whenever it is compiling and optimizing a query.
  2. Most compatible of all the lock modes, and do not block on any transactional locks, including X locks.
  3. It's perfectly fine for one session to be modifying data in a table while SQL Server is optimizing a query for another session that is accessing that table.

• **Sch-M: schema modification lock**
  1. SQL Server acquires a Sch-M lock when performing certain DDL operations that change a table's definition (its schema).
  2. These operations include adding and dropping columns from the table, or changing a column's data type.
  3. Sch-M locks are the least compatible lock mode and a request for every other mode will block on a Sch-M lock, and vice versa; a session cannot get a Sch-M lock if any other session has any other lock on the table.
  4. when a session is making a schema change to a table, no other sessions can do anything with the table.
Three more lock modes

**BU: bulk update lock**

1. SQL Server acquires a BU lock on a table only if a session explicitly requests one during a bulk insert operation into the table.

2. With the BULK INSERT command we can specify a BU lock, using the TABLOCK hint, and with the bcp utility we can use the –h "TABLOCK" option.

3. We can specify that SQL Server take BU locks, by default, for a particular table, during bulk updates.

4. BU locks allow multiple threads to load data into the same table concurrently, and they are only compatible with other BU locks and with Sch-S locks.
Creation of the DBlocks view to display locks in the current database.

```
IF EXISTS ( SELECT 1
    FROM sys.views
    WHERE name = 'DBlocks' )
    DROP VIEW DBlocks ;
GO
CREATE VIEW DBlocks AS
SELECT request_session_id AS spid ,
    DB_NAME(resource_database_id) AS dbname ,
    CASE WHEN resource_type = 'OBJECT'
    THEN OBJECT_NAME(resource_associated_entity_id)
    WHEN resource_associated_entity_id = 0 THEN 'n/a'
    ELSE OBJECT_NAME(p.object_id)
    END AS entity_name ,
    index_id ,
    resource_type AS resource ,
    resource_description AS description ,
    request_mode AS mode ,
    request_status AS status
FROM sys.dm_tran_locks t
LEFT JOIN sys.partitions p
    ON p.partition_id = t.resource_associated_entity_id
WHERE resource_database_id = DB_ID()
AND resource_type <> 'DATABASE' ;
```
USE AdventureWorks;
-- Create a new table
IF OBJECTPROPERTY(OBJECT_ID('NewOrders'), 'IsUserTable') = 1
    DROP TABLE NewOrders;
GO
SELECT * INTO NewOrders
FROM Sales.SalesOrderHeader;
GO
CREATE UNIQUE INDEX NewOrder_index ON NewOrders(SalesOrderID);
GO

-- Change isolation level and start transaction
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ;
BEGIN TRAN

-- SELECT data and examine the locks
SELECT *
FROM NewOrders
WHERE SalesOrderID = 55555;

SELECT *
FROM DBlocks
WHERE spid = @@spid
    AND entity_name = 'NewOrders';

-- UPDATE data and examine the locks
UPDATE NewOrders
SET SalesPersonID = 277
WHERE SalesOrderID = 55555;

SELECT *
FROM DBlocks
WHERE spid = @@spid
    AND entity_name = 'NewOrders';

ROLLBACK Tran
Creation of the DBlocks view to display locks in the current database.

The code in Listing will drop the NewOrders table if it already exists, then re-create it and build a non-clustered index on the SalesOrderID column. It then sets the isolation level to REPEATABLE READ in order that SQL Server holds S locks until the end of the transaction rather than just the end of the current statement, as is the case in the default READ COMMITTED level.
Lock conversion from S to X.

CREATE UNIQUE INDEX NewOrder_index ON NewOrders(SalesOrderID);
GO
-- Change isolation level and start transaction
SET TRANSACTION ISOLATION LEVEL REPEATABLE READ;
BEGIN TRAN
-- SELECT data and examine the locks
SELECT *
FROM NewOrders
WHERE SalesOrderID = 55555;
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
-- UPDATE data and examine the locks
UPDATE NewOrders
SET SalesPersonID = 277
WHERE SalesOrderID = 55555;
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
ROLLBACK TRAN

The code then opens a REPEATABLE READ transaction and selects one row from the NewOrders table. When querying our DBlocks view, we see an S lock on a RID in the database. This is the RID for the row that was selected, with SalesOrderID = 55555. Also, note that an S lock is held for a key in the non-clustered index. There are also IS locks for the page in the table that contains the selected row, and for the page in the index that contains the key for the selected row as well as an IS lock on the table itself.
Lock conversion from S to X.

After we update the row, we query the DBlocks view again, and this time there are different locks on the same resources. The same RID now has an X lock, and the page in the table, and the table itself, both have IX locks. The key in the index has a U lock. SQL Server acquired the U lock while searching for the row to update, and this is necessary because, until the modification to the data row happens, SQL Server doesn't know whether the modification will also require a change to the index. The page in the index containing the key has an IU lock.
Lock conversion from S to X.

Locks acquired by a SELECT and then UPDATE in the same transaction.
Special Intent locks

- Shared intent exclusive (SIX).
- Update intent exclusive (UIX)
- Shared intent update (SIU)
Special Intent Locks

• SQL Server holds an S lock on a row, then it also holds IS locks on the page and the table containing that row.

• SQL Server acquires an IU lock on an index page, when the component (a key) of that index had a U lock.

• In addition to the IS, IX, and IU there are three more types of intent locks that can be considered conversion locks.

• SQL Server will acquire these types of lock when a non-intent lock is requested on a resource on which either an IX or an IU lock is already held.
SIX Mode

• Compromise
• Both SI and X. Safer.
• Something inside is written
• When SQL Server has one or more rows locked with X locks, the pages and the table that contains the rows will acquire IX locks. When the same transaction performs an operation that requires an S lock, SQL Server will acquire a SIX lock on the table.
USE AdventureWorks2014;
--Step 1: Create a new table and set the isolation level
IF OBJECTPROPERTY(OBJECT_ID('NewOrders'), 'IsUserTable') = 1
DROP TABLE NewOrders;
GO
SELECT *
INTO NewOrders
FROM Sales.SalesOrderHeader;
GO
CREATE UNIQUE INDEX NewOrder_index ON NewOrders(SalesOrderID);
GO
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
GO
SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
GO

-- Step 2: Generate an SIX lock
BEGIN TRAN
UPDATE dbo.NewOrders
SET ShipDate = ShipDate + 1
WHERE SalesOrderID = 55555;
GO

SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
GO

SELECT *
FROM dbo.NewOrders WITH ( TABLOCK, REPEATABLEREAD )
WHERE SalesOrderID = 55555;
GO

SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
GO
ROLLBACK TRAN
GO
Special Intent Locks : SIX MODE

<table>
<thead>
<tr>
<th>spid</th>
<th>dbname</th>
<th>entity_name</th>
<th>index_id</th>
<th>resource</th>
<th>description</th>
<th>mode</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adventure2014</td>
<td>NewOrders</td>
<td>NULL</td>
<td>OBJECT</td>
<td></td>
<td>IX</td>
<td>GRANT</td>
</tr>
<tr>
<td>2</td>
<td>Adventure2014</td>
<td>NewOrders</td>
<td>2</td>
<td>KEY</td>
<td>(fc81b91b4150)</td>
<td>U</td>
<td>GRANT</td>
</tr>
<tr>
<td>3</td>
<td>Adventure2014</td>
<td>NewOrders</td>
<td>2</td>
<td>PAGE</td>
<td>1:24877</td>
<td>IU</td>
<td>GRANT</td>
</tr>
<tr>
<td>4</td>
<td>Adventure2014</td>
<td>NewOrders</td>
<td>0</td>
<td>RID</td>
<td>1:26952.6</td>
<td>X</td>
<td>GRANT</td>
</tr>
<tr>
<td>5</td>
<td>Adventure2014</td>
<td>NewOrders</td>
<td>0</td>
<td>PAGE</td>
<td>1:26952</td>
<td>IX</td>
<td>GRANT</td>
</tr>
</tbody>
</table>

- starts with the UPDATE statement, which will acquire an X lock on updated row, and IX locks on the page and the table containing the row.
- Subsequent query against the NewOrders table, in the same transaction, will obtain an S lock on the table and hold it to the end of the transaction, thanks to the REPEATABLE READ hint.
- Query against the DBlocks view reveals that SQL Server has acquired an SIX lock on the table.
Key-Range Locks

- when considering lock resources and SERIALIZABLE isolation level.

- If the isolation level is SERIALIZABLE and a query scans a range of data within a transaction, SQL Server needs to lock enough of the table to ensure that another transaction cannot insert a new value into the range currently being scanned, because if we reissued the same query that value would then appear as a phantom.

- A key-range lock is associated with a specific index key, but includes the range of possible values less than or equal to the key with which the lock is associated, and greater than the previous key in the index leaf level.

- Another way to say it would be that a key-range lock spans the range between two keys, and includes the key at the end, but not the key at the beginning.

- For example, if an index leaf level included the sequential values "James" and "Jones," a key-range lock on "Jones" would lock out all key values greater than "James" and less than or equal to "Jones."
Key-Range Locks

USE AdventureWorks;

--Step 1: Create a new table and set the isolation level

IF OBJECTPROPERTY(OBJECT_ID('NewOrders'), 'IsUserTable') = 1
    DROP TABLE NewOrders;
GO

SELECT *
INTO NewOrders
FROM Sales.SalesOrderHeader;
GO
USE AdventureWorks;

--Step 1: Create a new table and set the isolation level
IF OBJECTPROPERTY(OBJECT_ID('NewOrders'), 'IsUserTable') = 1
    DROP TABLE NewOrders;
GO

SELECT *
INTO NewOrders
FROM Sales.SalesOrderHeader;
GO

CREATE UNIQUE INDEX NewOrder_index ON NewOrders(SalesOrderID);
GO

SET TRANSACTION ISOLATION LEVEL SERIALIZABLE;
GO
Generate RangeS-S locks

-- Step 2: Generate RangeS-S locks
BEGIN TRAN
SELECT * 
FROM dbo.NewOrders 
WHERE SalesOrderID BETWEEN 55555 AND 55557;
GO
SELECT * 
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
GO
ROLLBACK TRAN
GO
Generate RangeS-S locks

- Running as a SERIALIZABLE transaction, which requests a range of orders from the NewOrders table, based on SalesOrderID

- Results from the DBlocks view shows that SQL Server acquired four RangeS-S locks on KEY resources and that the SELECT statement returns three rows. It is normal to see one more key-range lock than the number of rows affected, because the ranges are open at the lower-valued end.

```
-- Step 2: Generate RangeS-S locks
BEGIN TRAN
SELECT * 
FROM dbo.NewOrders
WHERE SalesOrderID BETWEEN 55555 AND 55557;
GO
SELECT * 
FROM DBlocks
WHERE spid = @@spid 
AND entity_name = 'NewOrders';
GO
ROLLBACK TRAN
GO
```
Generate RangeS-S locks

- The four key-range locks, indicated by the four RangeS-S locks in DBlocks view, cover:

1. Range starting just after the key 55554 up to and including the key 55555
2. Range starting just after the key 55555 up to and including the key 55556
3. Range starting just after the key 55556 up to and including the key 55557
4. Range starting just after the key 55557 up to and including the key 55558.
Generate RangeS-S locks

- The four key-range locks, indicated by the four RangeS-S locks in DBlocks view, cover:

1. Range starting just after the key 55554 up to and including the key 55555

2. Range starting just after the key 55555 up to and including the key 55556

3. Range starting just after the key 55556 up to and including the key 55557

4. Range starting just after the key 55557 up to and including the key 55558.
Generate RangeS-S locks

- SQL Server will try to store any newly-inserted values, and remember that a range lock prevents SQL Server from inserting a new row into the locked range.

- Since the SalesOrderID column has an index, the rows will be stored in order of the SalesOrderID.

- Key-range locks will have to include the range from the key just prior to the first one selected up to and including the first key, so that the first key itself cannot be modified.

- Key-range locks will also have to include a range starting just after the highest-valued key selected up to the next key in the index, so that no values equal to the highest key selected can be inserted.
Generate RangeS-U locks

BEGIN TRAN
UPDATE dbo.NewOrders
SET ShipDate = ShipDate + 1
WHERE SalesOrderID BETWEEN 55555 AND 55557;
GO
SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
GO
ROLLBACK TRAN
Generate RangeS-U locks

- If a non-clustered index is used to locate and update rows in a heap, while in SERIALIZABLE isolation level, and if the column being updated is *not the indexed* column used for access, the SQL Server will acquire a lock of Type RangeS-U.
- This means that there is an S lock on the range between the index keys, but the index key itself has a U lock.
- The rows in the heap will have the expected X lock on the RID.
Generate RangeS-U locks

• If a non-clustered index is used to locate and update rows in a heap, while in SERIALIZABLE isolation level, and if the column being updated is *not the indexed* column used for access, the SQL Server will acquire a lock of Type RangeS-U.

• This means that there is an S lock on the range between the index keys, but the index key itself has a U lock.

• The rows in the heap will have the expected X lock on the RID.

```
-- Step 3: Generate RangeS-U locks
BEGIN TRAN
UPDATE dbo.NewOrders
SET ShipDate = ShipDate + 1
WHERE SalesOrderID BETWEEN 55555 AND 55557;
GO
SELECT * FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
GO
ROLLBACK TRAN
```
Generate Range X-X locks

-- STEP 4: Generate RangeX-X locks
-- We need a clustered index to see these locks

CREATE UNIQUE CLUSTERED INDEX NewOrder_index ON NewOrders(SalesOrderID)
WITH DROP_EXISTING;
GO

BEGIN TRAN
UPDATE dbo.NewOrders
SET ShipDate = ShipDate + 1
WHERE SalesOrderID BETWEEN 55555 AND 55557;
GO

SELECT *
FROM DBlocks
WHERE spid = @@spid
AND entity_name = 'NewOrders';
GO

ROLLBACK TRAN
Generate Range X-X locks

If updating rows in an index while in SERIALIZABLE isolation level, the session will acquire exclusive key-range locks.

Starts by converting the non-clustered index on SalesOrderID to a clustered index, and then updates the same range of rows as previous.

In order to observe RangeX-X locks, the updated rows must be index keys, which is true when the table has a clustered index, and would also occur when updating one of the key columns of a non-clustered index.
Generate Range X-X locks

- If updating rows in an index while in SERIALIZABLE isolation level, the session will acquire exclusive key-range locks.
- Starts by converting the non-clustered index on SalesOrderID to a clustered index, and then updates the same range of rows as previous.
- In order to observe RangeX-X locks, the updated rows must be index keys, which is true when the table has a clustered index, and would also occur when updating one of the key columns of a non-clustered index.

```
-- STEP 4: Generate RangeX-X locks
-- We need a clustered index to see these locks
CREATE UNIQUE CLUSTERED INDEX NewOrder_index ON NewOrders(SalesOrderID)
WITH DROP_EXISTING;
GO
BEGIN TRAN
UPDATE dbo.NewOrders
SET ShipDate = ShipDate + 1
WHERE SalesOrderID BETWEEN 55555 AND 55557;
GO
SELECT *
FROM DBlocks
WHERE spid = @spid
    AND entity_name = 'NewOrders';
GO
ROLLBACK TRAN
```
### Conversion key-range locks

<table>
<thead>
<tr>
<th>Lock 1</th>
<th>Lock 2</th>
<th>Conversion Lock</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Rangel-N</td>
<td>Rangel-S</td>
</tr>
<tr>
<td>U</td>
<td>Rangel-N</td>
<td>Rangel-U</td>
</tr>
<tr>
<td>X</td>
<td>Rangel-N</td>
<td>Rangel-X</td>
</tr>
<tr>
<td>Rangel-N</td>
<td>RangeS-S</td>
<td>RangeX-S</td>
</tr>
<tr>
<td>Rangel-N</td>
<td>RangeS-U</td>
<td>RangeX-U</td>
</tr>
</tbody>
</table>
Controlling Locking

• Changing the transaction isolation level (the most common method).

• Changing the lock timeout period so that a transaction either skips past the locked rows, or rolls back.

• Using lock hints in SQL statements to control lock granularity, or specify custom behavior on encountering locked rows.

• Using bound connections to allow multiple connections to share the same locks.

• Using application locks to extend the resources that can be locked.
Controlling Concurrency and Locking Via the Isolation Level

- **READ UNCOMMITTED** – A transaction operating in READ UNCOMMITTED isolation level takes no locks while performing SELECT operations so it cannot block on locks held by other transactions.

- **READ COMMITTED** – The default isolation level, in which SQL Server holds shared locks only until the data has been read, and holds exclusive locks until the end of the transaction.

- **REPEATABLE READ** – A transaction operating in REPEATABLE READ isolation level keeps shared locks and exclusive locks until the end of the transaction.
Controlling Concurrency and Locking Via the Isolation Level

• **SERIALIZABLE** – The most restrictive isolation level, SERIALIZABLE adopts a special locking mechanism, using key-range locks, and holds all locks until the end of the transaction, so that users can't insert new rows into those ranges.

• **SNAPSHOT** – Has the outward appearance of SERIALIZABLE, but operates under a completely different concurrency model, optimistic concurrency,
Application-level “Optimistic Locking”

Idea: strive for short transactions or short lock duration

Approach:
- aim at two-phase structure of transactions: read phase + short write phase
- run queries under relaxed isolation level (typically read committed)
- rewrite program to test for concurrent writes during write phase

Example:

```
Select Balance, Counter Into :b, :c
From Accounts Where AccountNo = :x
...
compute interests and fees, set b, ...
...
Update Accounts
Set Balance = :b, Counter = Counter + 1
Where AccountNo = :x And Counter = :c
```

avoids lost updates, but cannot guarantee consistency
Pessimistic Locking

• A row lock would be placed as soon as the user indicates his intention to perform an update on a specific row that he has selected

Example:

> select empno, ename, sal from emp where deptno = 10;

<table>
<thead>
<tr>
<th>EMPNO</th>
<th>ENAME</th>
<th>SAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>7782</td>
<td>CLARK</td>
<td>2450</td>
</tr>
<tr>
<td>7839</td>
<td>KING</td>
<td>5000</td>
</tr>
<tr>
<td>7934</td>
<td>MILLER</td>
<td>1300</td>
</tr>
</tbody>
</table>
Pessimistic Locking- Cont

- Bind the values the user selected so we can query the database and make sure the data hasn’t been changed yet.

```sql
> variable empno number
> variable ename varchar2(20)
> variable sal number
> exec :empno := 7934; :ename := 'MILLER'; :sal := 1300;
PL/SQL procedure successfully completed.
```

- we are going to lock the row using FOR UPDATE NOWAIT

```sql
> select empno, ename, sal from emp
where empno = :empno
and decode( ename, :ename, 1 ) = 1
and decode( sal, :sal, 1 ) = 1
for update nowait
/
EMPNO ENAME SAL
---------- ---------- ----------
7934 MILLER 1300
```

decode( column, :bind_variable, 1 ) = 1 shorthand way of expressing
where (column = :bind_variable OR (column is NULL and :bind_variable is NULL))“.
since NULL = NULL is never true (nor false!)
Pessimistic Locking- Cont

• If the underlying data has not changed, we will get our MILLER row back, and this row will be locked from updates (but not reads) by others.
• If another user is in the process of modifying that row, we will get an ORA-00054 resource busy error. We must wait for the other user to finish with it.
• If, in the time between selecting the data and indicating our intention to update, someone has already changed the row, then we will get zero rows back.

SCOTT@ORA12CR1> update emp
set ename = :ename, sal = :sal
where empno = :empno;
row updated.
> commit;
Commit complete.
Optimistic Locking

• Defers all locking up to the point right before the update is performed.
• One popular implementation of optimistic locking is to keep the old and new values in the application, and upon updating the data, use an update like

\[
\text{Update table}
\]
\[
\text{Set column1 = :new_column1, column2 = :new_column2, ...}
\]
\[
\text{Where primary_key = :primary_key}
\]
\[
\text{And decode( column1, :old_column1, 1 ) = 1}
\]
\[
\text{And decode( column2, :old_column2, 1 ) = 1}
\]

Other Options:
• Optimistic Locking Using a Version Column (systimestamp column)
• Optimistic Locking Using a Checksum
Optimistic Locking Using a Version Column

- Insert a copy of the DEPT data into this table:

```
EODA@ORA12CR1> create table dept
2   ( deptno  number(2),
3    dname   varchar2(14),
4    loc     varchar2(13),
5    last_mod timestamp with time zone
6    default systimestamp
7    not null,
8    constraint dept_pk primary key(deptno)
9  )
10 /
Table created.
```

LAST_MOD column NOT NULL. It must be populated with default value: current system time.

TIMESTAMP data type has the highest precision available in Oracle (microsecond)

```
EODA@ORA12CR1> insert into dept( deptno, dname, loc )
2   select deptno, dname, loc
3    from scott.dept;
4 rows created.
```

```
EODA@ORA12CR1> commit;
Commit complete.
```
Optimistic Locking Using a Version Column

EODA@ORA12CR1> variable deptno number
EODA@ORA12CR1> variable dname varchar2(14)
EODA@ORA12CR1> variable loc varchar2(13)
EODA@ORA12CR1> variable last_mod varchar2(50)
EODA@ORA12CR1>
EODA@ORA12CR1> begin
   2   :deptno := 10;
   3   select dname, loc, to_char( last_mod, 'DD-MON-YYYY HH.MI.SSXFF AM TZR' )
   4       into :dname,:loc,:last_mod
   5       from dept
   6       where deptno = :deptno;
   7   end;
   8   /
PL/SQL procedure successfully completed.

• Currently

EODA@ORA12CR1> select :deptno dno, :dname dname, :loc loc, :last_mod lm
   2   from dual;

<table>
<thead>
<tr>
<th>DNO</th>
<th>DNAME</th>
<th>LOC</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>ACCOUNTING</td>
<td>NEW YORK</td>
<td>15-APR-2014 07.04.01.147094 PM -06:00</td>
</tr>
</tbody>
</table>
Optimistic Locking Using a Version Column

• **Update row by primary key (DEPTNO).** Verify LAST_MOD column had not been modified by any other session between the time we read it first and the time we did the update.

```sql
EODA@ORA12CR1> update dept
2    set dname = initcap(:dname),
3        last_mod = systimestamp
4    where deptno = :deptno
5    and last_mod = to_timestamp_tz(:last_mod, 'DD-MON-YYYY HH.MI.SSXFF AM TZR');
1 row updated.
```

line 3 of the UPDATE statement updates the LAST_MOD column to be the current time if the row is found to be updated.

• **Run same update again.** Notice how 0 rows updated because the predicate on LAST_MOD was not satisfied.

```sql
EODA@ORA12CR1> update dept
2    set dname = upper(:dname),
3        last_mod = systimestamp
4    where deptno = :deptno
5    and last_mod = to_timestamp_tz(:last_mod, 'DD-MON-YYYY HH.MI.SSXFF AM TZR');
0 rows updated.
```
Optimistic or Pessimistic Locking?

- Pessimistic locking works very well in Oracle
  - Has many advantages over optimistic locking
  - Requires a stateful connection to DB, like a client/server connection => unrealistic in many cases today

- Optimistic concurrency recommend for most applications with version column approach with a timestamp
  - less computationally expensive than a hash or checksum on large columns like LONG, LONG RAW, CLOB, BLOB.
• Simply locks the table, not the rows in the table.

• If you start modifying the rows, they will be locked as normal. So, this is not a method to save on resources (as it might be in other RDBMSs).

• **Locking a Table: Example for a large batch update**
  
  – locks the employees table in exclusive mode but does not wait if another user already has locked the table:

  ```sql
  LOCK TABLE employees IN EXCLUSIVE MODE NOWAIT;
  ```
DBMS_LOCK package

- Use this package to serialize access to some resource external to Oracle.
  - Using the UTL_FILE routine that allows you to write to a file on the server’s file system.
- File is external, Oracle won’t coordinate the many users trying to modify it simultaneously.
- Before you open, write, and close the file, you will request a lock named after the file in exclusive mode, and after you close the file, you will manually release the lock.
- Only one person at a time will be able to write a message to this file. Everyone else will queue up.
- **DBMS_LOCK package allows you to manually release a lock when you are done with it, or to give it up automatically when you commit, or even to keep it as long as you are logged in.**
DBMS_LOCK package

- **ALLOCATE_UNIQUE** Procedure: Allocates a unique lock identifier.

  **Syntax**
  
  ```sql
  DBMS_LOCK.ALLOCATE_UNIQUE (lockname IN VARCHAR2, lockhandle OUT VARCHAR2, expiration_secs IN INTEGER DEFAULT 864000);
  ```

- **CONVERT** Function: Converts a lock from one mode to another.

  **Syntax**
  
  ```sql
  DBMS_LOCK.CONVERT( id IN INTEGER || lockhandle IN VARCHAR2, lockmode IN INTEGER, timeout IN NUMBER DEFAULT MAXWAIT) RETURN INTEGER;
  ```

- **RELEASE** Function: Explicitly releases a lock previously acquired using the REQUEST function.

  **Syntax**
  
  ```sql
  DBMS_LOCK.RELEASE (id IN INTEGER) RETURN INTEGER;
  DBMS_LOCK.RELEASE (lockhandle IN VARCHAR2) RETURN INTEGER;
  ```
Unrestricted **multiprogramming level (MPL)** can lead to performance disaster known as **data-contention thrashing**:  
- additional transactions cause superlinear increase of lock waits  
- throughput drops sharply  
- response time approaches infinity
Benefit of MPL Limitation

system admin sets **MPL limit**: during load bursts excessive transactions wait in **transaction admission queue**

![Graph showing the benefit of MPL limitation](image)

avoids thrashing, but poses a tricky tuning problem:
- overly low MPL limit causes long waits in admission queue
- overly high MPL limit opens up the danger of thrashing

problem is even more difficult for highly heterogeneous workloads
Chapter 10: Implementation and Pragmatic Issues

• 10.2 Data Structures of a Lock Manager
• 10.3 Multi-Granularity Locking and Lock Escalation
• 10.4 Transient Versioning
• 10.5 Nested Transactions for Intra-transaction parallelism
• 10.6 Tuning Options
• 10.7 Overload Control
• 10.8 Lessons Learned
Conflict-ratio-driven Overload Control

\[
\text{conflict ratio} = \frac{\text{# locks held by all trans.}}{\text{# locks held by running trans.}}
\]

Critical conflict ratio \(\approx 1.3\)

\[
\text{conflict ratio} = \frac{\text{# locks held by all trans.}}{\text{# locks held by running trans.}}
\]
Conflict-ratio-driven Overload Control Algorithm

upon begin request of transaction t:
    if conflict ratio < critical conflict ratio
    then admit t else put t in admission queue fi

upon lock wait of transaction t:
    update conflict ratio
    while not (conflict ratio < critical conflict ratio)
        among trans. that are blocked and block other trans.
        choose trans. v with smallest product
        #locks held * #previous restarts
        abort v and put v in admission queue od

upon termination of transaction t:
    if conflict ratio < critical conflict ratio then
        for each transaction q in admission queue do
            if (q will be started the first time) or
                (q has been a rollback/cancellation victim and
                all trans. that q was waiting for are terminated)
            then admit q fi od fi
The problem

The OS has no ability to characterize Oracle session traffic

Mixed workloads represent real-world Oracle traffic

How can we prevent certain sessions from consuming too many resources on the server?
SQL Server 2016 Resource Governor

• Resource pools. A resource pool represents the server’s physical resources

• Workload groups. A workload group serves as a container for session requests that have similar classification criteria

• Classification. The classification process assigns incoming sessions to a workload group based on the characteristics of the session
Resource Governor limitations

- Resource management is limited to the SQL Server Database Engine. Resource Governor cannot be used for Analysis Services, Integration Services, and Reporting Services.
- No workload monitoring or management exists between SQL Server instances.
- Resource Governor can manage OLTP workloads, but these types of queries—typically very short in duration—are not always on the CPU long enough to apply bandwidth controls. This may skew the statistics returned for CPU usage percent.
- The ability to govern physical I/O only applies to user operations and not system tasks. System tasks include write operations to the transaction log and lazy writer I/O operations. The Resource Governor applies primarily to user read operations because most write operations are typically performed by system tasks.
- You cannot set I/O thresholds on the internal resource pool.
Enabling Resource Governor
Creating a resource pool

- **Resource pools** are sections of resources (CPU and memory) that can be used by one or more workload groups (groups of applications.)

- SQL Server has two resource pools by default: **internal for the server itself**, and **default for all other unassigned workloads**.

```sql
-- Create a new resource pool explicitly stating all values
-- As all values being given are defaults, this could alternatively be done
-- as CREATE RESOURCE POOL dbapool;
CREATE RESOURCE POOL dbapool
WITH
(  
  MIN_CPU_PERCENT = 0,  -- how much must be assigned to this pool
  MAX_CPU_PERCENT = 100,  -- how much would be assigned if possible (note, can be exceeded if no conter
  CAP_CPU_PERCENT = 100,  -- cannot-be-exceeded maximum, useful for predictable billing
  AFFINITY SCHEDULER = AUTO,
  MIN_MEMORY_PERCENT = 0,  -- memory allocated to this pool that cannot be shared
  MAX_MEMORY_PERCENT = 100,  -- percentage total server memory which is allowed to be used by this pool
  MIN_IOPS_PER_VOLUME = 0,  -- minimum number of I/O operations per second per disk volume to reserve
  MAX_IOPS_PER_VOLUME = 2147483647  -- maximum. Note, this is the max allowed value for this property.
);
```
Creating a workload group

• Sessions are allocated into workload groups by the classifier function.

```sql
-- Create a workload group using the resource pool we just created
CREATE WORKLOAD GROUP dbagroup
WITH (
    IMPORTANCE = MEDIUM, -- relative importance compared to other workgroups
    REQUEST_MAX_MEMORY_GRANT_PERCENT = 50, -- how much memory a single process can request from the pool
    REQUEST_MAX_CPU_TIME_SEC = 0, -- how long a single request can take without generating a CPU Threshold
    MAX_DOP = 0, -- max degree of parallelism allowed
    GROUP_MAX_REQUESTS = 0 -- num simultaneous events allowed, 0 means unlimited
)
USING dbapool
;
GO

-- update resource governor to changes
ALTER RESOURCE GOVERNOR RECONFIGURE;
GO
```
Creating a classifier function

- The **classifier function** allocates incoming queries into workgroups.
- You will allocate all queries run by a DBA into the newly created DBA workgroup.

```sql
-- make a function returning the workgroup name for the session to be allocated to
-- allocates new sessions to dbapool if dba in login name
USE Master
GO
CREATE FUNCTION fx_DBAClassifier()
RETURNS sysname
WITH SCHEMABINDING
AS
BEGIN

DECLARE @wg sysname

IF
    SUSER_NAME() LIKE '%dba%'
    SET @wg = 'dbagroup'
ELSE SET @wg = 'default'
RETURN @wg
END;
GO

-- tell the resource governor to use the function
ALTER RESOURCE GOVERNOR with (CLASSIFIER_FUNCTION = dbo.fx_DBAClassifier);
-- restart the resource governor
ALTER RESOURCE GOVERNOR RECONFIGURE;
```
Setting up a DBA session for testing Resource Governor behavior
Setting up a DBA session for testing Resource Governor behavior
Setting up a DBA session for testing Resource Governor behavior
Setting up a DBA session for testing Resource Governor behavior
Setting up a DBA session for testing Resource Governor behavior
Setting up a DBA session for testing Resource Governor behavior

```
USE MASTER
GO
select g.name from sys.dm_exec_sessions s INNER JOIN sys.dm_resource_governor_workload_groups g
  ON s.group_id=g.group_id where s.session_id = @@SPID;
```
Testing

- Testing performance impact of workload group assignment
- *Currently, the dbapool has default parameters, so performance will respond as if the session were in the default group*
Testing

- In the table at the bottom, clear the Show checkbox for the %Processor Time counter
Testing

5. In the **Performance Monitor** command bar, click .
6. Scroll to **SQLServer:Resource Pool Stats**, and expand it by clicking the down arrow.
### Setting up a DBA session for testing Resource Governor behavior

<table>
<thead>
<tr>
<th>Available counters</th>
<th>Added counters</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQLServer:Replication Merge</td>
<td>Counter</td>
</tr>
<tr>
<td>SQLServer:Replication Snapshot</td>
<td></td>
</tr>
<tr>
<td>SQLServer:Resource Pool Stats</td>
<td></td>
</tr>
<tr>
<td>SQLServer:SQL Errors</td>
<td></td>
</tr>
<tr>
<td>SQLServer:SQL Statistics</td>
<td></td>
</tr>
<tr>
<td>SQLServer:SSIS Pipeline 13.0</td>
<td></td>
</tr>
<tr>
<td>SQLServer:SSIS Service 13.0</td>
<td></td>
</tr>
<tr>
<td>SQL Server:Transactions</td>
<td></td>
</tr>
</tbody>
</table>

1. In the Performance Monitor command bar, click `...`.
2. Scroll to SQLServer:Resource Pool Stats, and expand it by clicking the down arrow.
3. Select Disk Read IO/sec, and then click `Add>>`. 

---

**Example:**

- Select `...` from the Performance Monitor command bar.
- Navigate to `SQLServer:Resource Pool Stats` and expand it.
- Choose `Disk Read IO/sec` and click `Add>>`.

---

**Note:** Ensure that the correct counters are added for monitoring resource usage effectively.
Setting up a DBA session for testing Resource Governor behavior

- 7. Select Disk Read IO/sec, and then click Add>>.
- 8. Select Disk Write IO/sec, and then click Add>>.
- 9. Click OK.
- Do not close the Performance Monitor
Setting up a DBA session for testing Resource Governor behavior

- 7. Select Disk Read IO/sec, and then click Add>>.
- 8. Select Disk Write IO/sec, and then click Add>>.
- 9. Click OK.
- Do not close the Performance Monitor
Setting up a DBA session for testing Resource Governor behavior

- 7. Select Disk Read IO/sec, and then click Add>>.
- 8. Select Disk Write IO/sec, and then click Add>>.
- 9. Click OK.
- Do not close the Performance Monitor
Setting up a DBA session for testing Resource Governor behavior

- 7. Select Disk Read IO/sec, and then click Add>>.
- 8. Select Disk Write IO/sec, and then click Add>>.
- 9. Click OK.
- Do not close the Performance Monitor
Executing test query

```sql
-- NOTE: reporting service performance is not affected by resource governance
-- so we must show improvements here by a query
-- clear the buffers
CHECKPOINT
GO

DBCC DROPCLEANBUFFERS
DBCC FREEPROCCACHE
GO

USE [AdventureworksDW2016CTP3]
GO

-- make a temporary table to do IO-intensive write operations on
CREATE TABLE Temp (ID int);
GO

-- do the IO-intensive write operation
INSERT INTO Temp
    select top 10000000 ProductKey from FactResellerSalesXL_CCI;
GO

-- clean up
DROP TABLE Temp
```
Executing test query

- Switch to the Performance Monitor window again by clicking in the taskbar.
- See how the disk write-and-read counters spiked when you ran the query.
Executing test query

- Switch to the **Performance Monitor** window again by clicking in the taskbar.
- *See how the disk write-and-read counters spiked when you ran the query.*
Alerting resource pools

- Altering Resource Governance resource pool `dbapool` to limit `MAX_IOPS_PER_VOLUME`
Alerting resource pools

• Altering Resource Governance resource pool dbapool to limit MAX_IOPS_PER_VOLUME

```sql
USE [AdventureworksDW2016CTP3]
GO
ALTER RESOURCE POOL dbapool
    WITH (MAX_IOPS_PER_VOLUME = 50 -- This is a very low IOPS, so should have a noticeable effect);
GO
ALTER RESOURCE GOVERNOR RECONFIGURE;
```
Checking effect of altered rules

```
-- NOTE: reporting service performance is not affected by resource governance
-- so we must show improvements here by a query
-- clear the buffers
CHECKPOINT
GO

DBCC DROPCLEANBUFFERS
DBCC FREEPROCCACHE
GO

USE [AdventureworksDW2016CTP3]
GO

-- make a temporary table to do IO-intensive write operations on
CREATE TABLE Temp (ID int);
GO

-- do the IO-intensive write operation
INSERT INTO Temp
    select top 10000000 ProductKey from FactResellerSalesXL_CCI;
GO

-- clean up
DROP TABLE Temp
```
Executing test query

- Note that the spike is wider and shorter than when you previously ran the query. Notice that the maximum value for the read-and-write I/O is limited to approximately 50 as well.
Executing test query

- Note that the spike is wider and shorter than when you previously ran the query. Notice that the maximum value for the read-and-write I/O is limited to approximately 50 as well.
Example:

- Oracle Resource manager
Introducing Database Resource Manager

- DBRM is a quality of service (QoS) tool where we can schedule and guarantee sessions a certain level of service

- What is a “session”?
  - Oracle user, application, computer
  - Sessions are organized into consumer groups

- What resources can be scheduled?
  - CPU utilization
  - Parallel execution
  - Concurrent active sessions
  - Session time limits
  - Etc…
Database Resource Manager (DBRM)

DBRM allows us to guarantee service levels for sessions

DBRM also enables us to cap sessions such that they don’t overreach

DBRM consists of several related components
Resource Manager Components

- Resource Consumer Group
- Plan Directive
- Resource Plan
Configuring Database Resource Manager with EM
Configuring Database Resource Manager with EM
Configuring Database Resource Manager with EM

The Database Resource Manager helps you manage how resources such as CPU and Parallel Queuing are allocated among user sessions.

**TIP** Administration Privileges Specify which users or roles have the system privilege "Administer Resource Manager", which is required for configuring Resource Manager.

- **Consumer Groups** Define Consumer Groups, which are user sessions grouped together based on resource processing requirements.
- **Consumer Group Mappings** Define consumer group mapping rules, which are used to map user sessions to consumer groups.
- **Plans**
  - Define Resource Plans for a single database, which contain directives that specify how resources are allocated to Consumer Groups.
- **Performance Statistics** Monitor the statistics for the currently enabled Resource Plan.
Configuring Database Resource Manager with EM

### Consumer Groups: `psight.localdomain`

<table>
<thead>
<tr>
<th>Consumer Group</th>
<th>Enabled</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATCH_GROUP</td>
<td>NO</td>
<td>Consumer group for batch operations</td>
</tr>
<tr>
<td>DEFAULT_CONSUMER_GROUP</td>
<td>YES</td>
<td>Consumer group for users not assigned to any consumer group</td>
</tr>
<tr>
<td>DSS_CRITICAL_GROUP</td>
<td>NO</td>
<td>Consumer group for critical DSS queries</td>
</tr>
<tr>
<td>DSS_GROUP</td>
<td>NO</td>
<td>Consumer group for DSS queries</td>
</tr>
<tr>
<td>ETL_GROUP</td>
<td>NO</td>
<td>Consumer group for ETL</td>
</tr>
<tr>
<td>INTERACTIVE_GROUP</td>
<td>NO</td>
<td>Consumer group for interactive, OLTP operations</td>
</tr>
<tr>
<td>LOW_GROUP</td>
<td>NO</td>
<td>Consumer group for low-priority sessions</td>
</tr>
<tr>
<td>ORA$APPQOS_0</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$APPQOS_1</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$APPQOS_2</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$APPQOS_3</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$APPQOS_4</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$APPQOS_5</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$APPQOS_6</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$APPQOS_7</td>
<td>YES</td>
<td>Consumer group for Application QOS</td>
</tr>
<tr>
<td>ORA$AUTOTASK</td>
<td>YES</td>
<td>Consumer group for autotask operations</td>
</tr>
<tr>
<td>SYS_GROUP</td>
<td>YES</td>
<td>Consumer group for system administrators</td>
</tr>
</tbody>
</table>
Configuring Database Resource Manager with EM
Configuring Database Resource Manager with EM
A Simple Resource Plan

A Simple Resource Plan Plan

Resource Plan "DAYTIME"

Directive 1
75% of CPU

Directive 2
15% of CPU

Directive 3
10% of CPU

Consumer Group "OLTP"

Consumer Group "REPORTING"

Consumer Group "OTHER_GROUPS"
Resource Plan Key points

- Only one plan can be active on a server at a time
  - In RAC environments, keep plans and schedules the same
  - Use the Scheduler and maintenance windows to switch plans
  - You can use subplans

- A complex resource plan can extend up to 8 priority levels

- We create plans by using DBMS_RESOURSE_PLAN or (as always) Cloud Control 12c
# Built in Resource Plan

<table>
<thead>
<tr>
<th>Consumer Group Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFAULT_PLAN</td>
<td>A general purpose plan that prioritizes SYS to allow the DBA to debug hangs and unexpected heavy loads.</td>
</tr>
<tr>
<td>DEFAULT_MAINTENANCE_PLAN</td>
<td>Plan for the maintenance windows. Provides some CPU resources for automated maintenance tasks.</td>
</tr>
<tr>
<td>MIXED_WORKLOAD_PLAN</td>
<td>Plan for managing mixed workloads (interactive and batch).</td>
</tr>
<tr>
<td>DSS_PLAN</td>
<td>Plan for managing DSS workloads (DSS, ETL, and batch).</td>
</tr>
<tr>
<td>ETL_CRITICAL_PLAN</td>
<td>Similar to DSS_PLAN but prioritizes ETL over DSS.</td>
</tr>
</tbody>
</table>

# Built in Consumer Groups

<table>
<thead>
<tr>
<th>Consumer Group Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERACTIVE_GROUP</td>
<td>Consumer group for interactive, OLTP operations.</td>
</tr>
<tr>
<td>BATCH_GROUP</td>
<td>Consumer group for batch operations.</td>
</tr>
<tr>
<td>DSS_GROUP</td>
<td>Consumer group for decision support system (DSS) queries.</td>
</tr>
<tr>
<td>DSS_CRITICAL_GROUP</td>
<td>Consumer group for critical DSS queries.</td>
</tr>
<tr>
<td>ETL_GROUP</td>
<td>Consumer group for data load, or ETL jobs.</td>
</tr>
<tr>
<td>SYS_GROUP</td>
<td>Consumer group for system administrators.</td>
</tr>
<tr>
<td>LOW_GROUP</td>
<td>Consumer group for low-priority sessions.</td>
</tr>
<tr>
<td>OTHER_GROUPS</td>
<td>Default consumer group for all sessions.</td>
</tr>
</tbody>
</table>

# Consumer Groups Mapping Rules

<table>
<thead>
<tr>
<th>Session Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>service_module_action</td>
<td>A combination of service name, module name, and action name, in this format: service_name.module_name.action_name</td>
</tr>
<tr>
<td>service_module</td>
<td>A combination of service and module names in this format: service_name.module_name</td>
</tr>
<tr>
<td>module_name_action</td>
<td>A combination of module and action names in this format: module_name.action_name</td>
</tr>
<tr>
<td>module_name</td>
<td>The module name in the currently running application.</td>
</tr>
<tr>
<td>service_name</td>
<td>The service name used by the client to establish a connection.</td>
</tr>
<tr>
<td>oracle_user</td>
<td>The Oracle database user name.</td>
</tr>
<tr>
<td>client_program</td>
<td>The name of the client program used to log onto the server.</td>
</tr>
<tr>
<td>client_os_user</td>
<td>The operating system user name of the client.</td>
</tr>
<tr>
<td>client_machine</td>
<td>The name of the computer from which the client is making the connection.</td>
</tr>
<tr>
<td>client_id</td>
<td>The client identifier.</td>
</tr>
</tbody>
</table>

Managing Resource Consumers Groups

- Their initial group is used during initial connection – this happens dynamically depending upon your mapping rules
- You can then either manually or programmatically switch consumer group membership for sessions

```
BEGIN
DBMS_RESOURCE_MANAGER.Switch_Consumer_Group_for_Sess ('17', '12345', 'HIGH_PRIORITY');
END;
```

```
BEGIN
DBMS_RESOURCE_MANAGER.Create_Plan_Directive (plan => 'DAYTIME',
GROUP_OR_SUBPLAN => 'OLTP',
COMMENT => 'OLTP group',
MGMT_P1 => 75,
SWITCH_GROUP => 'LOW_GROUP',
SWITCH_TIME => 5);
END;
```
Resource Manager in Multitenant Environment

CDB resource plan in root

- **Directive**: share = 3
  - PDB salespdb

- **Directive**: share = 3
  - PDB servicespdb

- **Directive**: share = 1
  - PDB hrpdb

Chapter 10: Implementation and Pragmatic Issues

- 10.2 Data Structures of a Lock Manager
- 10.3 Multi-Granularity Locking and Lock Escalation
- 10.4 Transient Versioning
- 10.5 Nested Transactions for Intra-transaction parallelism
- 10.6 Tuning Options
- 10.7 Overload Control
- 10.8 Lessons Learned
Lessons Learned

• Locking can be efficiently implemented, with flexible handling of memory overhead by means of multi-granularity locks
• Tuning options include
  • choice of isolation levels
  • application-level tricks
  • MPL limitation
• Tuning requires extreme caution to guarantee correctness: if in doubt, don‘t do it!
• Concurrency control is susceptible to data-contention thrashing and needs overload control
Lessons Learned

• we looked some more advanced locking topics, including lock mode conversion, when SQL Server acquires additional locks on data that is already locked.

• We covered the special lock mode called key-range locks that can be held on ranges of index keys when running queries under SERIALIZABLE isolation level.

• We looked at when, and how, SQL Server will escalate locks on smaller resources into table or partition locks.

• Finally, we explored latches and compile locks
Types of deadlock

• Cycle deadlock
• Conversion deadlock
A cycle deadlock

- Transaction 1 starts, acquires an exclusive table lock on the Supplier table, and requests an exclusive table lock on the Part table.
- Simultaneously, Transaction 2 starts, acquires an exclusive lock on the Part table, and requests an exclusive lock on the Supplier table.
- The two transactions become deadlocked – caught in a "deadly embrace." Each transaction holds a resource needed by the other process. Neither can proceed and, without intervention, both would be stuck in deadlock forever.
Generating a cycle deadlock

USE AdventureWorks
-- On one connection, start Transaction 1:
BEGIN TRAN
UPDATE Purchasing.PurchaseOrderDetail
SET OrderQty = OrderQty + 200
WHERE ProductID = 922
    AND PurchaseOrderID = 499;
GO

-- Open a second connection, and start Transaction 2:
BEGIN TRAN
UPDATE Production.Product
SET ListPrice = ListPrice * 0.9
WHERE ProductID = 922;
GO

-- Go back to the first connection, and execute this update statement:
UPDATE Production.Product
SET ListPrice = ListPrice * 1.1
WHERE ProductID = 922;
GO

-- At this point, this first connection should block.
-- It is not deadlocked yet, however. It is waiting for a lock
-- on the Production.Product table, and there is no reason
-- to suspect that it won't eventually get that lock.
Generating a cycle deadlock

-- Now go back to the second connection,
-- and execute this update statement:

```
UPDATE Purchasing.PurchaseOrderDetail
SET OrderQty = OrderQty - 200
WHERE ProductID = 922
    AND PurchaseOrderID = 499;
GO
-- At this point a deadlock occurs.
```

One of the processes will receive the following error message. (Of course, the actual process ID reported will probably be different.)

```
Msg 1205, Level 13, State 51, Line 1
Transaction (Process ID 57) was deadlocked on lock resources with another process and has been chosen as the deadlock victim. Rerun the transaction.
```
A conversion deadlock.

Problems:
Two transactions were both searching for the same row to modify (for example, the same customer row in the Customers table), using different access paths, and they could both reach the desired resource at the same time.

Each transaction could acquire an S lock on that row, but then each transaction would attempt to convert this lock to an X lock in order to perform the modification, but the S lock held by one transaction prevents the other from doing so.
**SQL server : U (Update) Lock**

A conversion deadlock.

Problems:

- Two transactions were both searching for the same row to modify (for example, the same customer row in the Customers table), using different access paths, and they could both reach the desired resource at the same time.

- Each transaction could acquire an S lock on that row, but then each transaction would attempt to convert this lock to an X lock in order to perform the modification, but the S lock held by one transaction prevents the other from doing so.

---

Transactional Information System 10-222