RAC 12c Cache Fusion Internals

By
Riyaj Shamsudeen
Me

- 23+ years using Oracle products
- OakTable member
- Oracle ACE Director
- Specializes in RAC, performance tuning and Internals.
- Slowly in to BigData

- rshamsud@orainternals.com
- orainternals.wordpress.com
- Web: www.orainternals.com
Process architecture
Wireshark demo

Demo: wireshark
MTU

- MTU defines Maximum Transmission Unit of a packet. Limits the size of a packet, default is ~1500 bytes.
- Transfer of an 8K UDP => Transfer of 6 IP packets

Demo using wireshark
BL resources

- BL locks protect database blocks (in RAC).

- BL resource follows a naming convention of
  
  \[\text{Block_id} \] \[\text{file_id}\], BL

  
  \[0xe7c2\] \[0x1\], [BL]

  
  Block_id = 59330.
  File_id = 1
Single block read

- Block is not in any buffer cache. LMS grants a PR mode lock on the resource and asks FG to read from the disk.

Demo: demo_01a.sql
Trace lines

WAIT #18446741324875049632: nam='gc cr grant 2-way' ela= 499 p1=7 p2=6867 p3=1 obj#=76484 tim=4597940025

WAIT #18446741324875049632: nam='db file sequential read' ela= 758 file#=7 block#=6867 blocks=1 obj#=76484 tim=4597941129

PR mode

<table>
<thead>
<tr>
<th>KJBLNAME</th>
<th>KJBLNAME2</th>
<th>KJBLGRANT</th>
<th>KJBLROLE</th>
<th>KJBLREQUEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0x1ad3][0x7], [BL]</td>
<td>6867, 7, BL</td>
<td>KJUSERPR</td>
<td></td>
<td>0 KJUSERNL</td>
</tr>
</tbody>
</table>
**GCS structures**

A resource structure created in the directory instance, a lock created in instance 2

- $X$bh
- $X$le
- $X$kjbl
- $X$kjbr

A shadow structure setup in instance 1 to keep track of the resource.

**Demo:** tc_one_row.sql, gcs_locks.sql, gcs_resources.sql
**Single block transfer - 2 way**

- Block is in the directory instance in a compatible mode. Both block transfer and grant performed by the LMS process running in instance 2.

---

FG – Foreground Process
LMD – Lock Manager Daemon
GRD – Global Resource Directory
GCS Directory instance

Demo: demo_01a.sql and demo_01b.sql

A resource structure already exists still that block is in the buffer cache.
A set of GCS shadows (locks) setup on the resource in directory instance and the requesting instance.

Demo: demo_01a.sql and demo_0a.sql
Single block transfer - 3 way

- Block is in the buffer cache of instance 3. Instance 2 is the directory instance of the resource. LMS process transfers the blocks from instance 3 over the interconnect.

**Diagram:**

1. Instance 2 transfers block to instance 3.
2. LMS transfers block from instance 3 to instance 2.
3. LMD transfers block from instance 3 to instance 2.

**Annotations:**

- Directory instance for the resource
- PR mode lock
- Disk files
- FG – Foreground Process
- LMD – Lock Manager Daemon
- GRD – Global Resource Directory

©OraInternals Riyaj Shamsudeen
CUR mode

- Concurrent changes to the same block.
- Row level lock vs BL lock.
- EX grants must be acquired to change blocks.

Demo: upd_one_row_100.sql, upd_one_row_101.sql
CUR mode

- Two pending transactions in the same block.
Buffer changes

- Changes under EX mode.

- Downgrade by other instances.

<table>
<thead>
<tr>
<th>KJBLNAME</th>
<th>KJBLNAME2</th>
<th>KJBLGRANT</th>
<th>KJBLROLE</th>
<th>KJBLREQUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0x1ac4][0x7],[BL][ext 0x0,0x0 6852,7,BL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KJUSEREX</td>
<td>0</td>
<td>KJUSERNL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enter value for block: 6852

<table>
<thead>
<tr>
<th>STATE</th>
<th>MODE_HELD</th>
<th>LE_ADDR</th>
<th>DBARFIL</th>
<th>DBABLK</th>
<th>CR_SCN_BAS</th>
<th>CR_SCN_WRP</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>000000006D3E3AB0</td>
<td>7</td>
<td>6852</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
CR buffers

- Instance 1 acquired EX mode lock.
- Instance 2 requests the block, and LMS in instance 1 ships CR copy.

Demo, upd_100.sql, buffer_state, tc_one_row
Busy

- `gc cr block busy, gc current block busy`

- LMS constructed buffer applying undo records.

- Excessive *busy events = No application affinity.

- Application affinity will reduce *busy events as the buffers will be modified in the same instance.
Congested

- Congested wait events also imply concurrency, but at an higher level.

- If LMS process can not get to a request in 1ms time, then the response for the request will be marked with ‘congestion’ wait event.

- Review RT priority and LMS session/process metrics.
**gc cr grants 2-way**

Open request File #5, blk #8

Send time On ksxp

Wire

Read from the disk

Charge time to ‘gc cr grant 2-way’
Why DRM?

Excessive amount of GC activity for One object!

Demo: remaster demo. Refer sqldeveloper

©OraInternals  Riyaj Shamsudeen
In 11g, all resources are frozen during the reconfiguration.

DRM (11g)

Request queue

LCK

LMD

LMON

LMS0 LMS1 LMSn

Resources

Inst 1 Inst 2 Inst 3
In 12c, only set of resources in a window is frozen.

Resources

Inst 1

Inst 2

Inst 3
Resource names

- From 12 onwards, resource names are coded with con_id.

```sql
select resource_name from gv$ges_resource
where resource_name like '[0x15f29][0x0],[TM]%' /
```

<table>
<thead>
<tr>
<th>RESOURCE_NAME</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0x15f29][0x0],[TM][ext 0x3,0x0]</td>
<td>hrdb1 PDB – GSTEST table – object_id=89897</td>
</tr>
<tr>
<td>[0x15f29][0x0],[TM][ext 0x4,0x0]</td>
<td>hrdb2 PDB – GSTEST table – object_id=89897</td>
</tr>
</tbody>
</table>

- Comes handy if you are debugging RAC trace files, to identify the PDB generating the errors.
THANK YOU

- **Email**: rshamsud@orainternals.com
- **Blog**: orainternals.wordpress.com
- **Web**: www.orainternals.com