RAC 12c Cache Fusion Internals

By

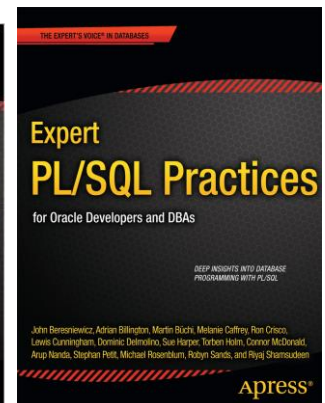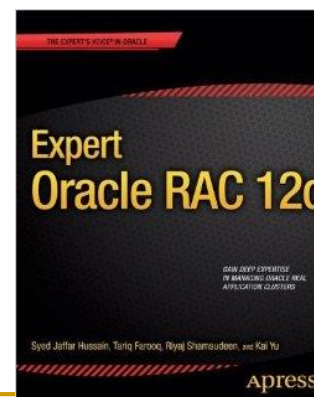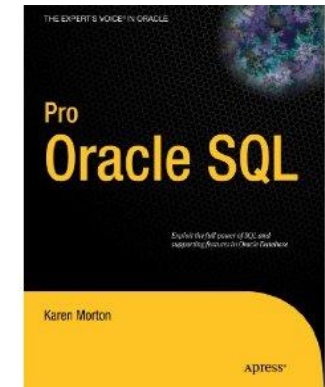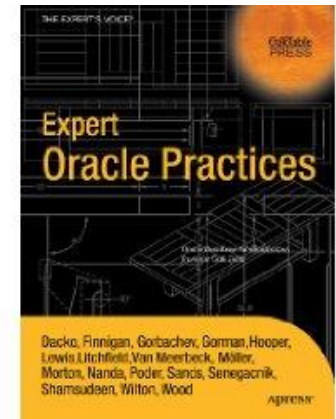Riyaj Shamsudeen

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
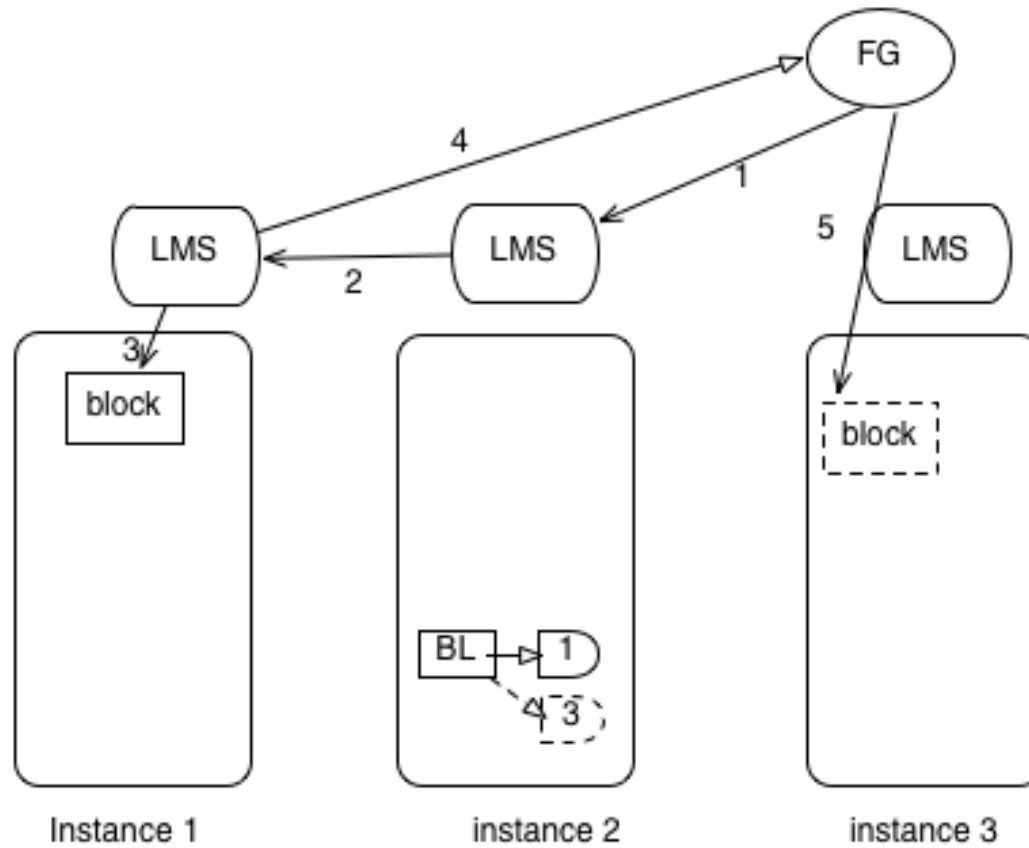
Wireshark demo
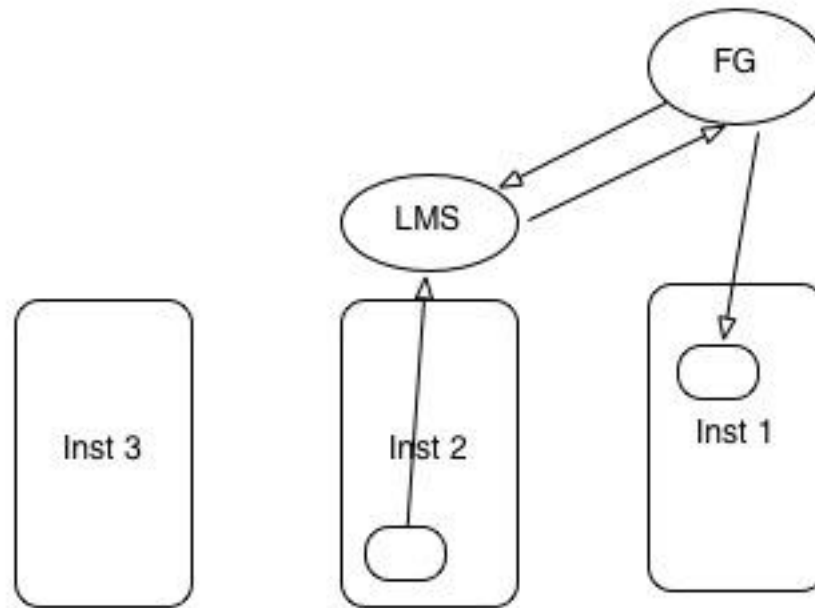
Demo: wireshark

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
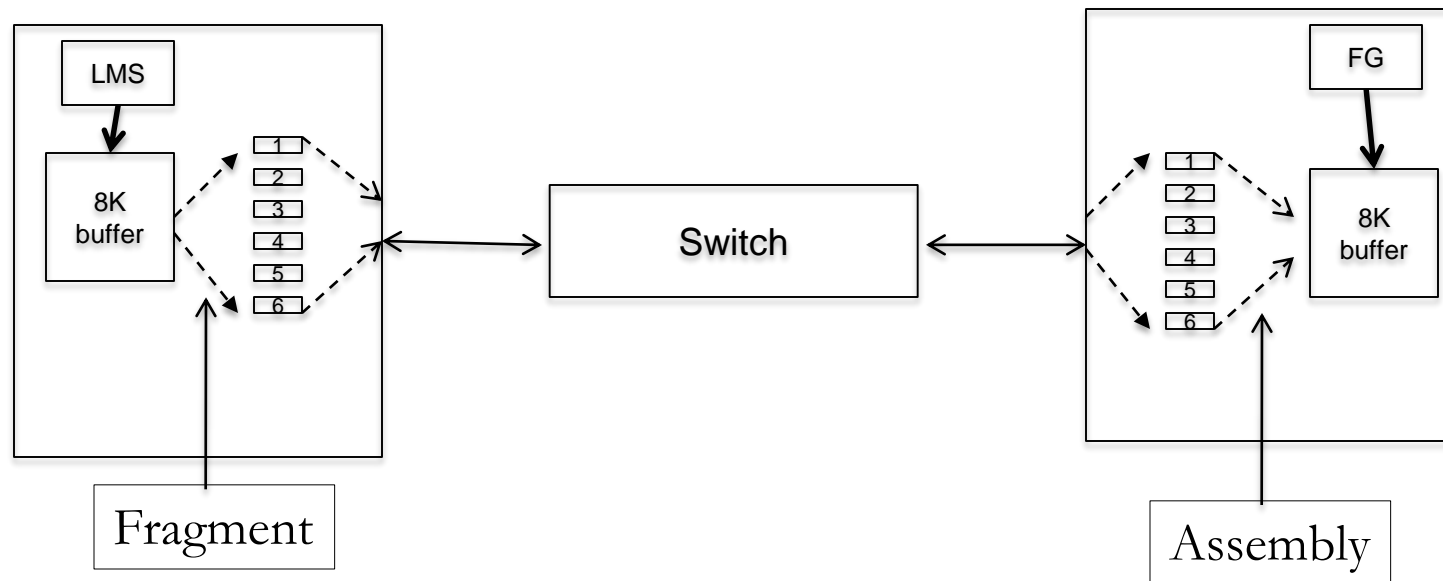
Demo using wireshark

BL resources

- BL locks protects database blocks(in RAC).

- BL resource follows a naming convention of
  [Block_id] [file_id], BL

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

Block_id = 59330.
   [0xe7c2 = 59330]

File_id = 1

Resource

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

Locks
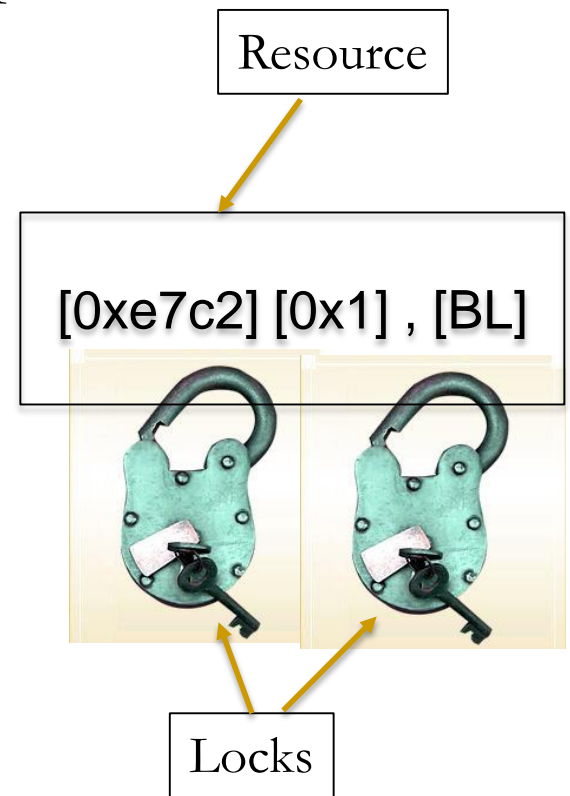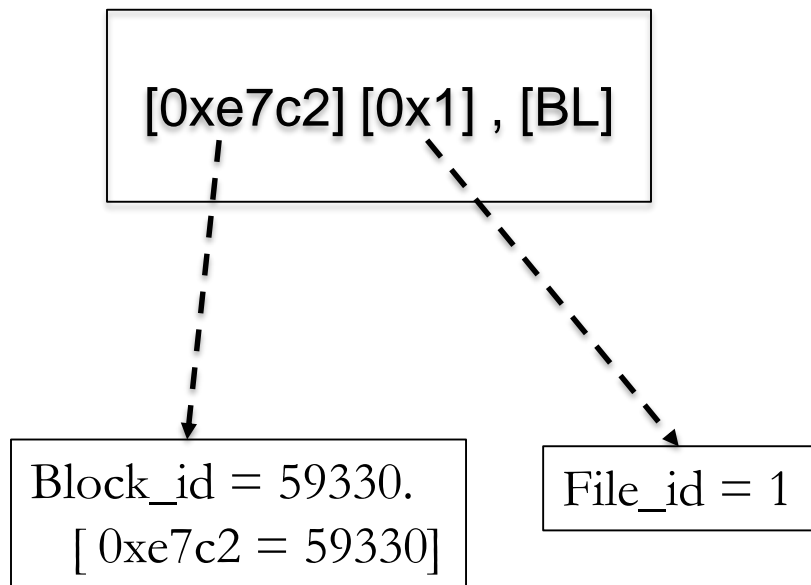
Single block read

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


```
[0x1ad3][0x7],[BL]
```

PR mode lock

Directory node for the resource

Disk files

FG — Foreground Process
LMD — Lock Manager Daemon
GRD — Global Resource Directory
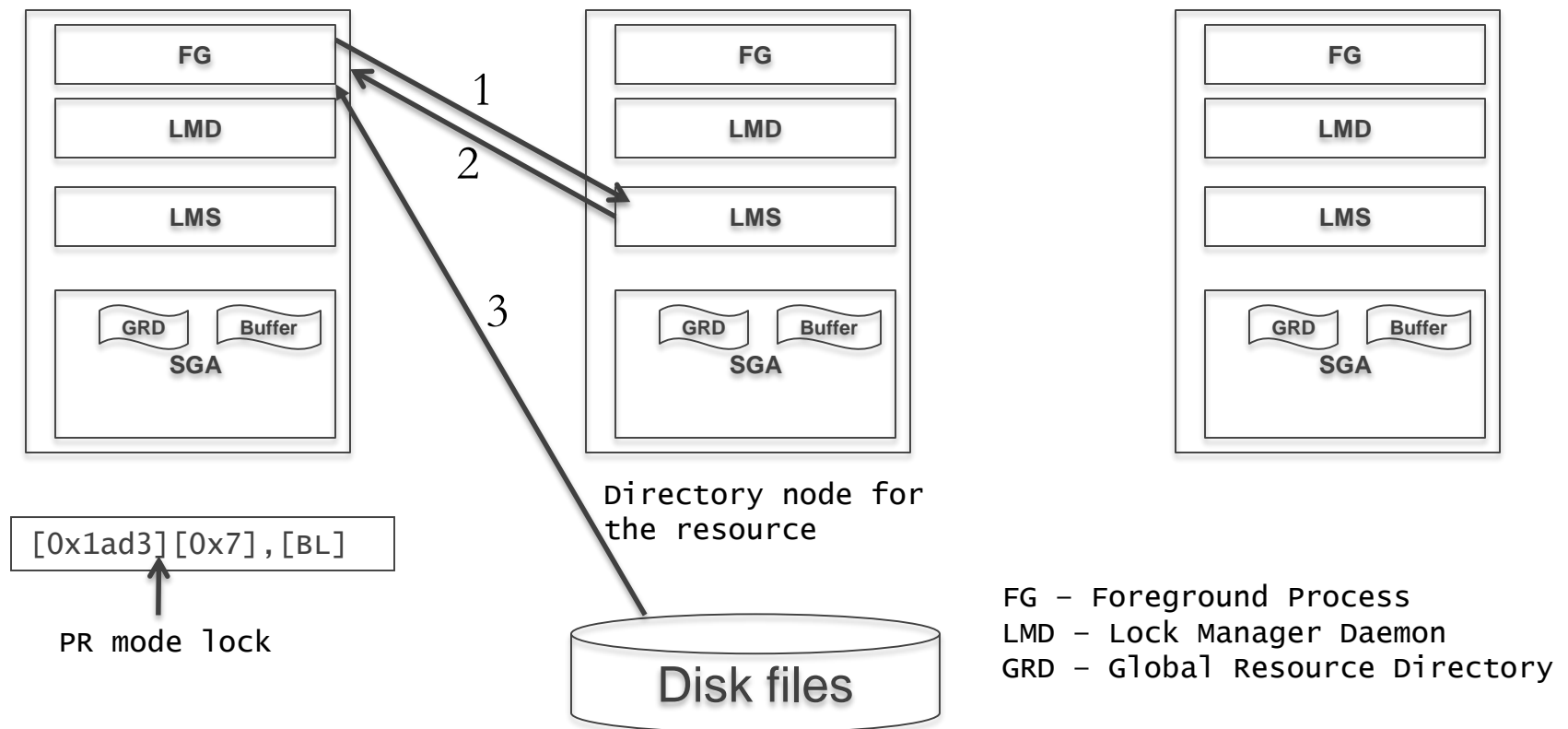
Demo: demo_01a.sql

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

```
KJBLNAME              KJBLNAME2                   KJBLGRANT  KJBLROLE KJBLREQUES
-------------------- ------------------------- ---------- -------- -----------
[0x1ad3][0x7],[BL]   6867,7,BL                 KJUSERPR          0 KJUSERNL
```

GCS structures

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

X$bh

X$le

X$kjbl

X$kjbr

BH	→ Buffer

LE
Shadow [0x1ac4][0x7],[BL]

Shadow [0x1ac4][0x7],[BL]

- - - - →

Resource [0x1ac4][0x7],[BL]

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

Instance 2 (directory instance)
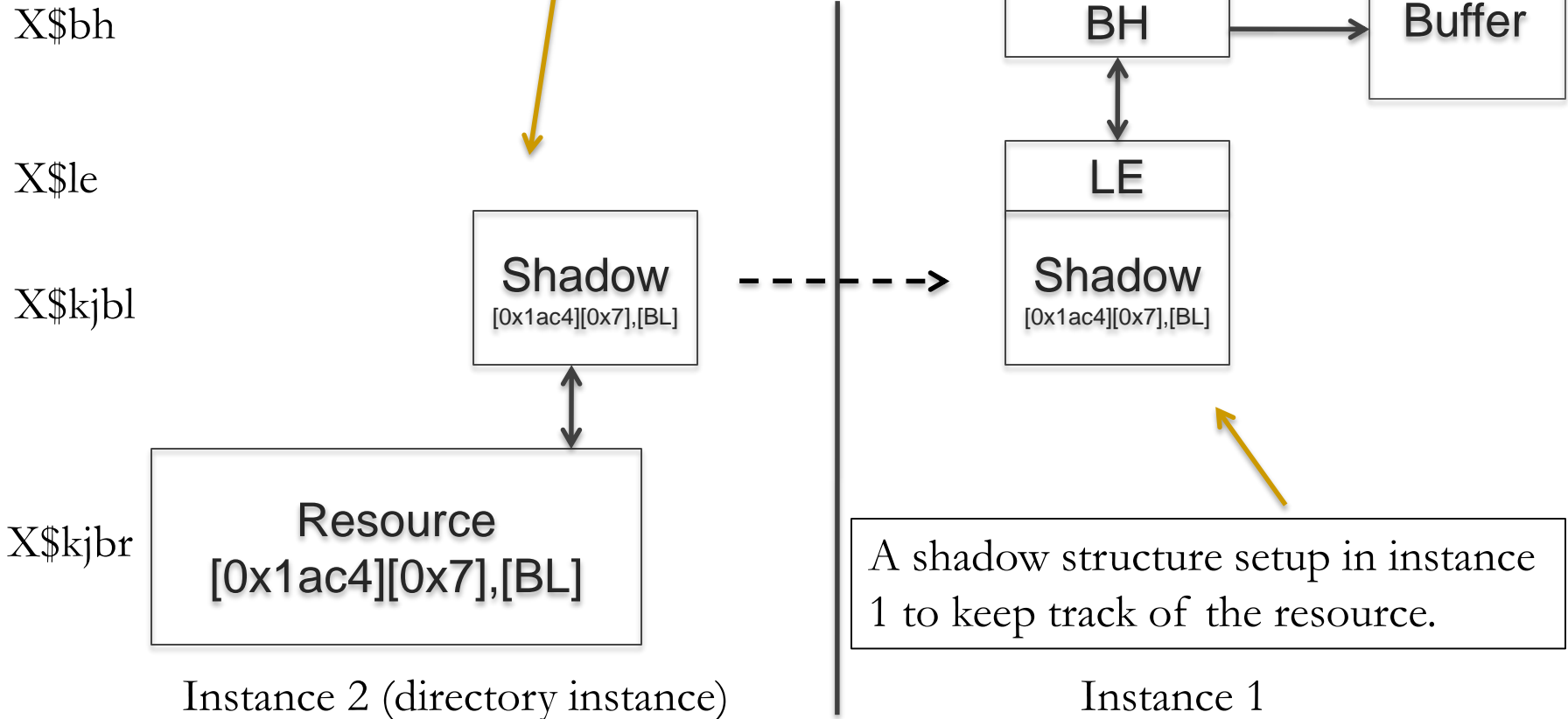
Instance 1

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

X$bh — BH → Buffer

X$le — LE

X$kjbl — Client [0x1ac4][0x7],[BL]

X$kjbr — Resource [0x1ac4][0x7],[BL]

A resource structure already exists still that block is in the buffer cache.
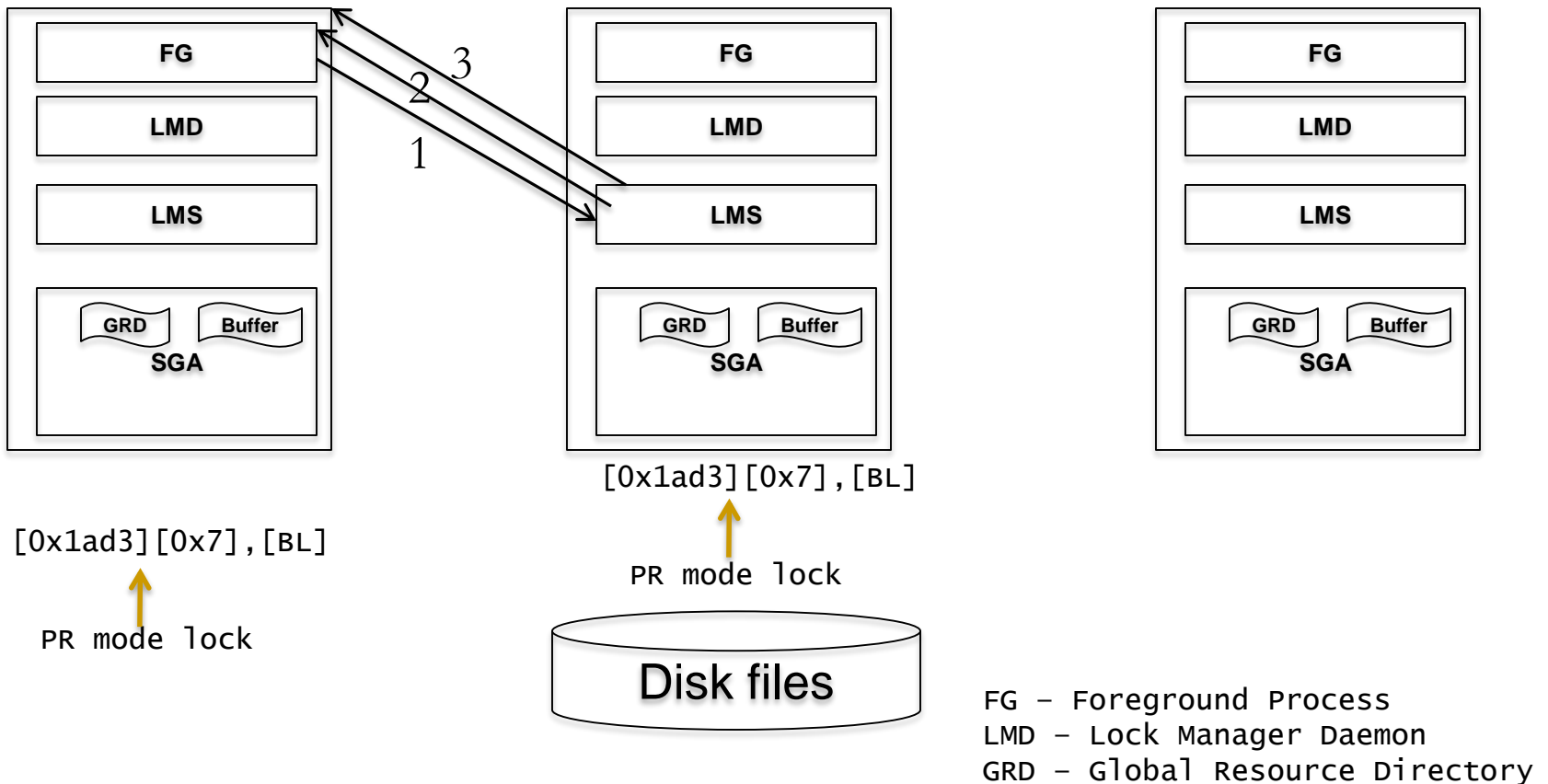
Instance 2 (Directory instance)
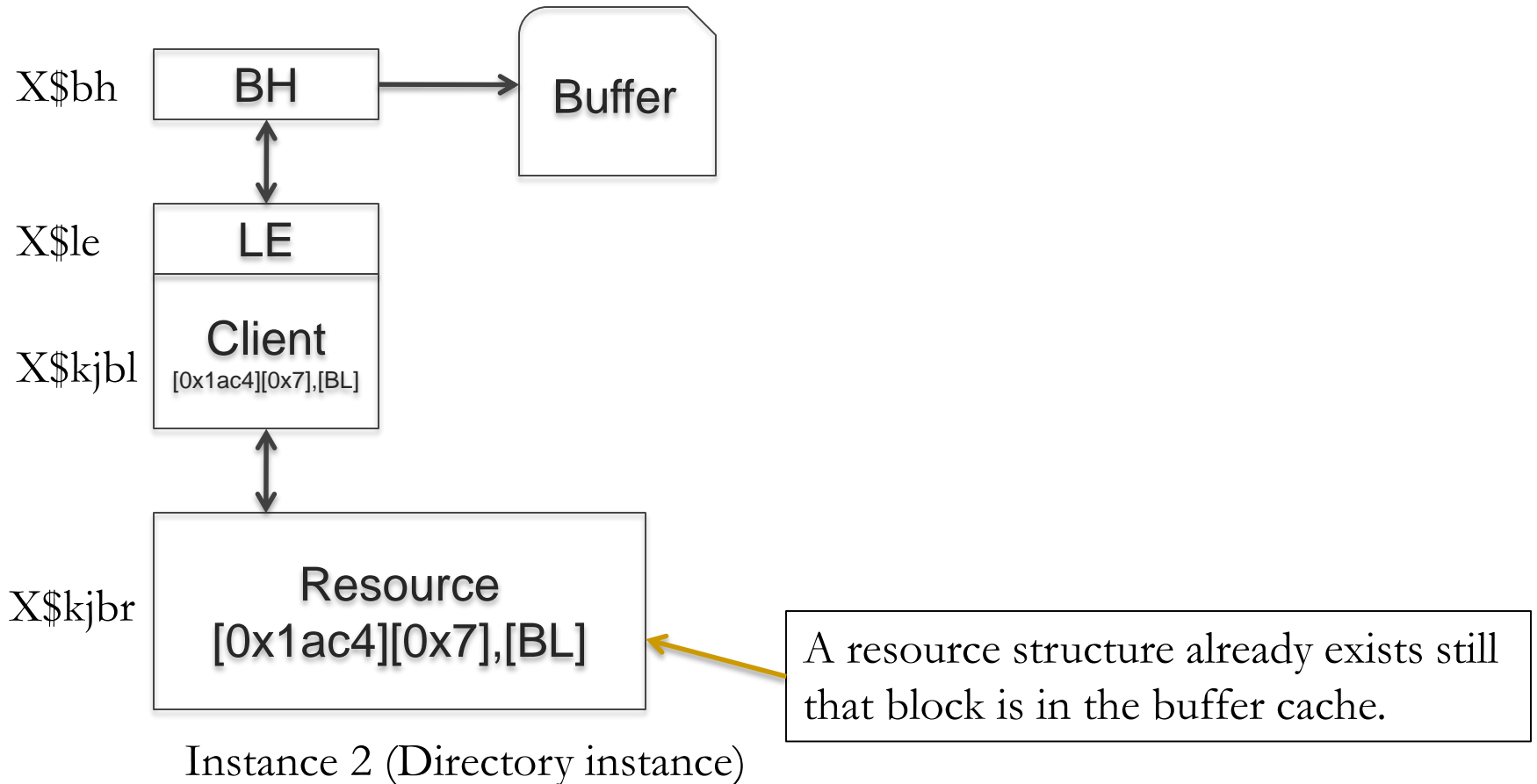
Demo: demo_01a.sql and demo_01b.sql

GCS shadows

X$bh

X$le

X$kjbl

X$kjbr

BH → Buffer

LE

Client
[0x1ac4][0x7],[BL]

Shadow
[0x1ac4][0x7],[BL]

Resource
[0x1ac4][0x7],[BL]

Instance 2 (directory instance)

BH → Buffer

LE

Shadow
[0x1ac4][0x7],[BL]

A set of GCS shadows (locks) setup on the resource in directory instance and the requesting instance.
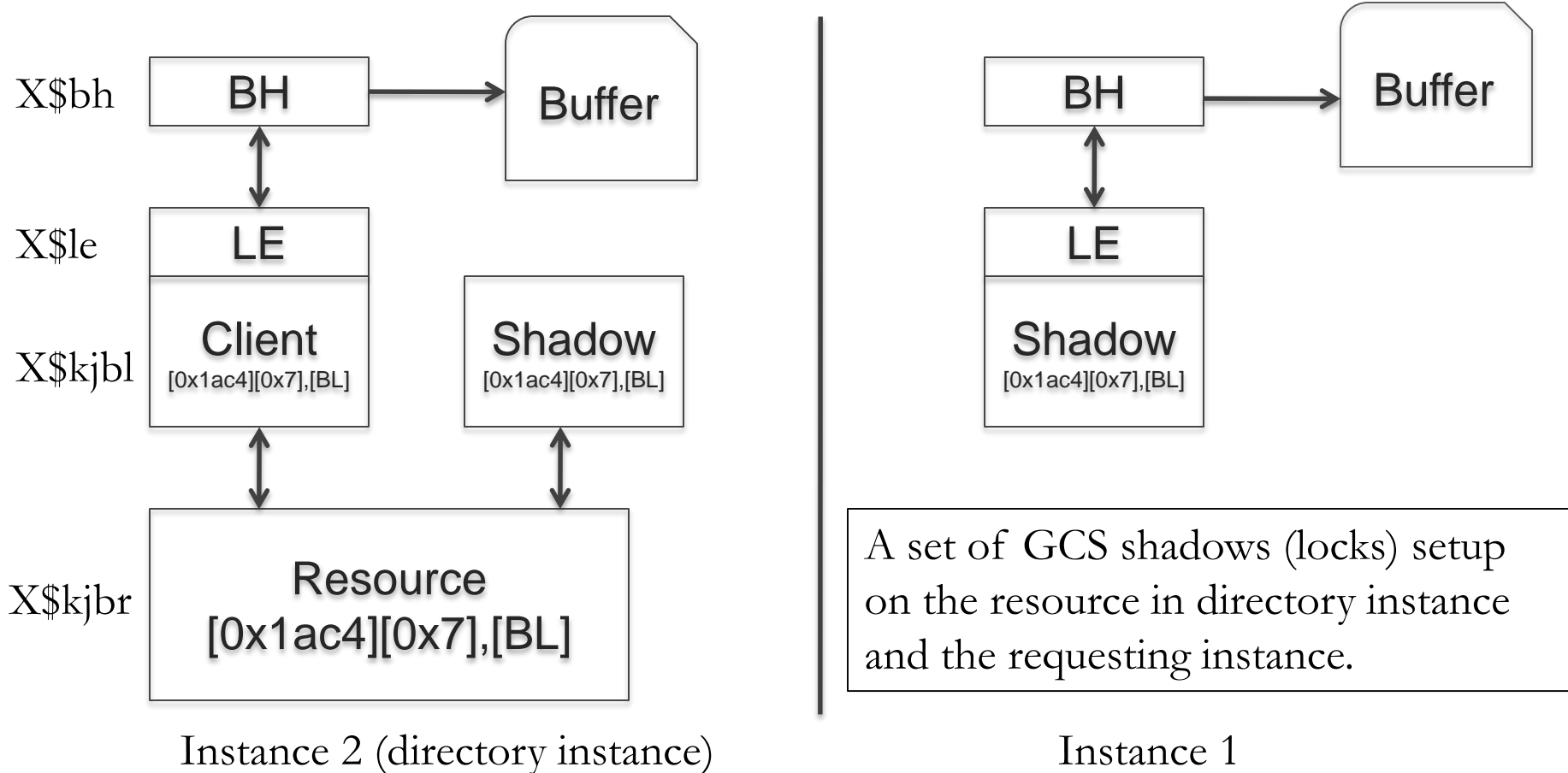
Instance 1
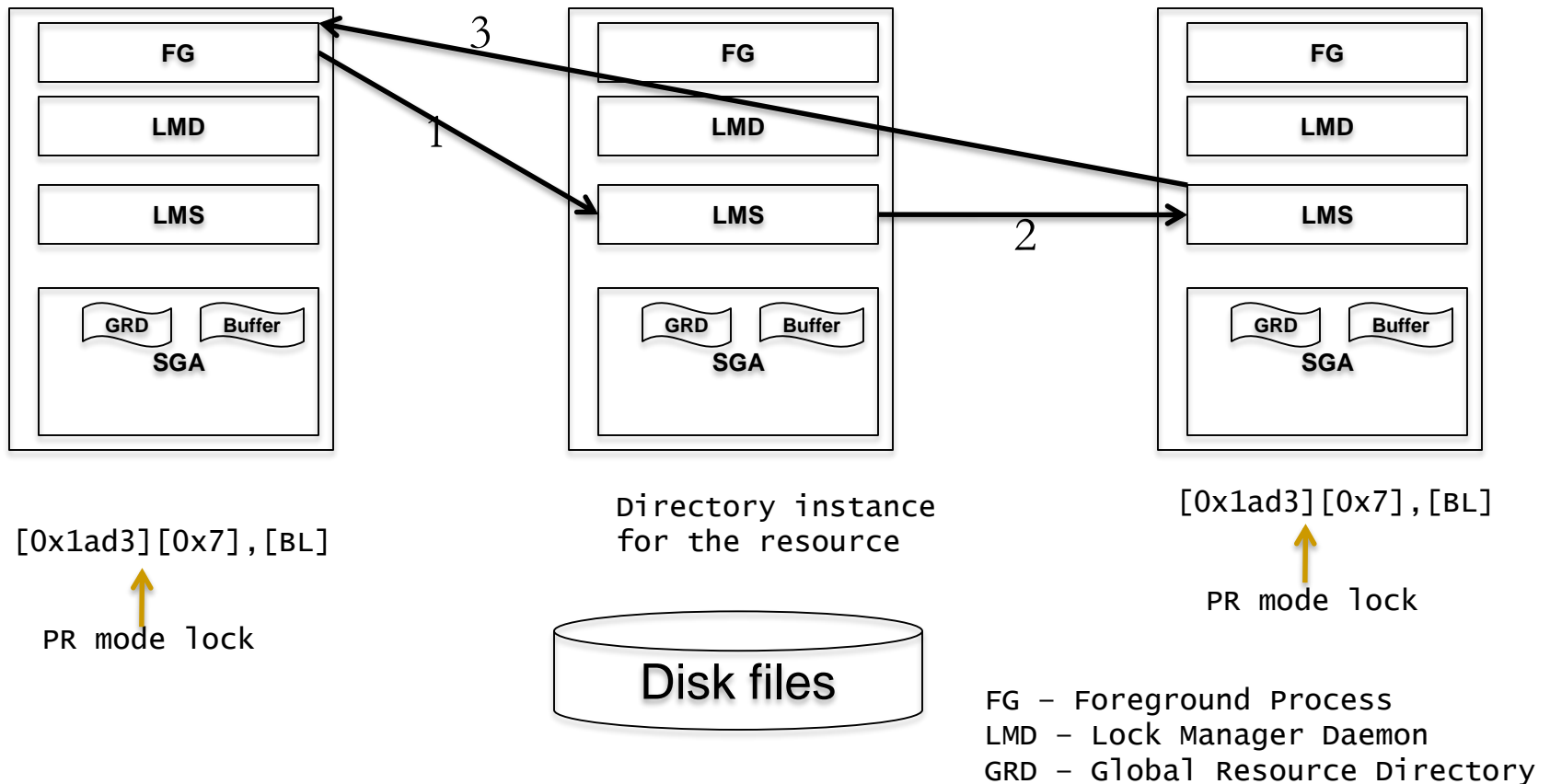
Demo: demo_01a.sql and demo_0a.sql

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.

Directory instance
for the resource

[0x1ad3][0x7],[BL]

[0x1ad3][0x7],[BL]

PR mode lock

PR mode lock

Disk files

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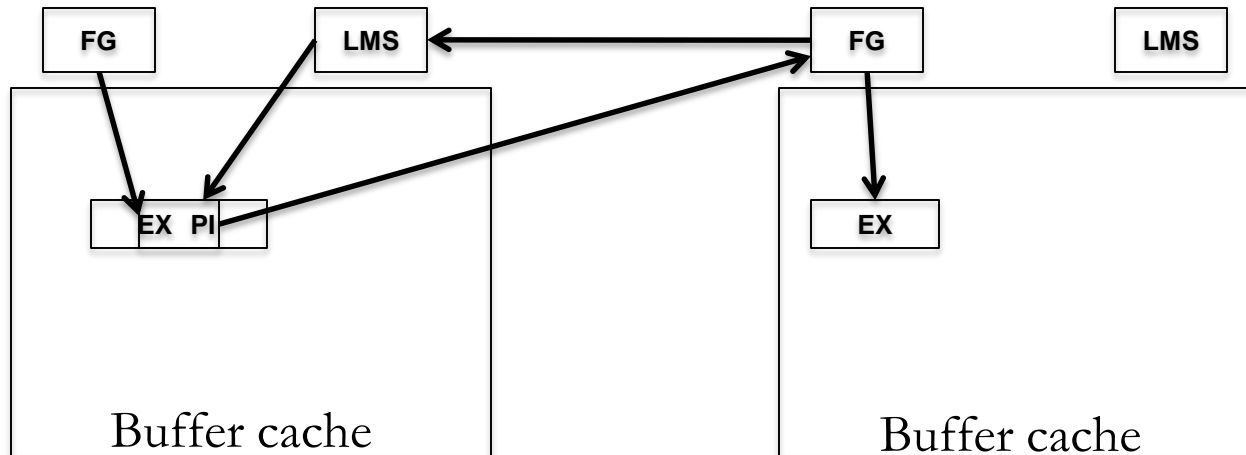LMD – Lock Manager Daemon
GRD – Global Resource Directory

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.

```
KJBLNAME                        KJBLNAME2            KJBLGRANT  KJBLROLE KJBLREQUES
----------------------------- -------------------- ---------- -------- ----------
[0x1ac4][0x7],[BL][ext 0x0,0x0 6852,7,BL            KJUSEREX          0 KJUSERNL


Enter value for block: 6852
     STATE   MODE_HELD LE_ADDR              DBARFIL     DBABLK CR_SCN_BAS CR_SCN_WRP     CLASS
---------- ---------- ---------------- ---------- ---------- ---------- ---------- ----------
         1          0 000000006D3E3AB0          7       6852          0          0          1
```

CR buffers

- Instance 1 acquired EX mode lock.

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

| FG | LGWR | LMS | | FG | | LMS |

EX

Apply undo

CR

CR

Buffer cache

CR

CR

Buffer cache

`[0x1ad3][0x7],[BL]`

EX mode lock
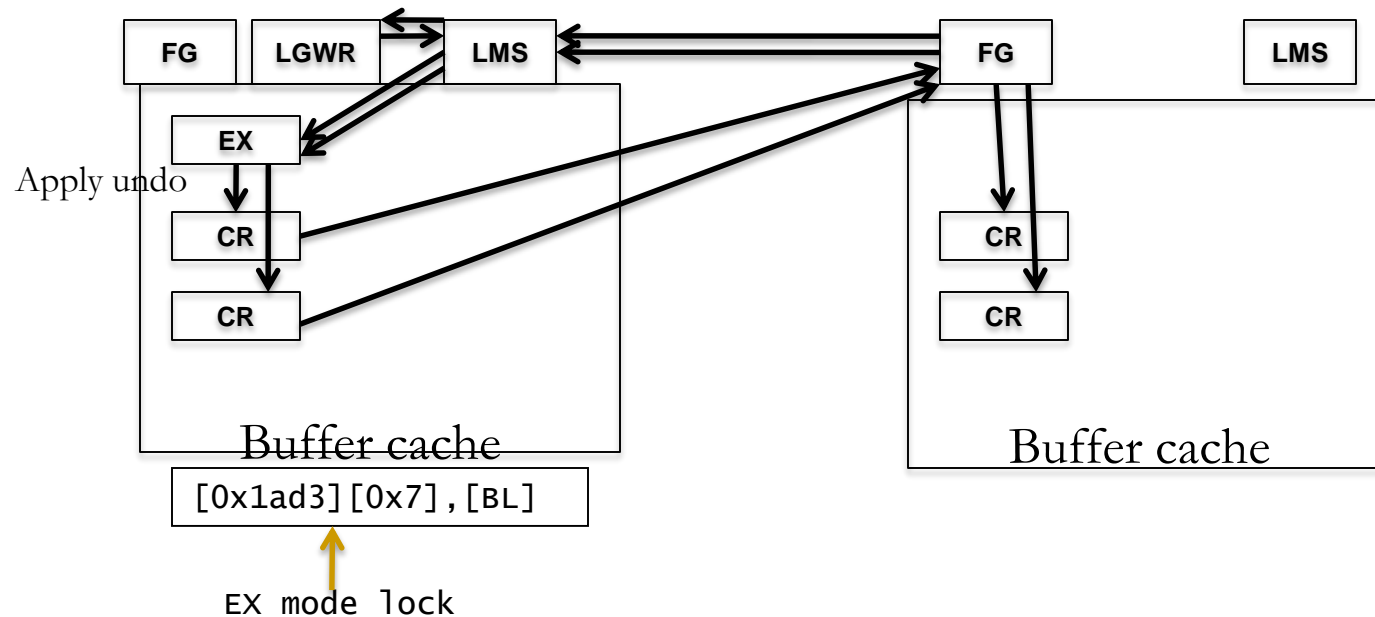
Demo, upd_100.sql , buffer_state, tc_one_row

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.

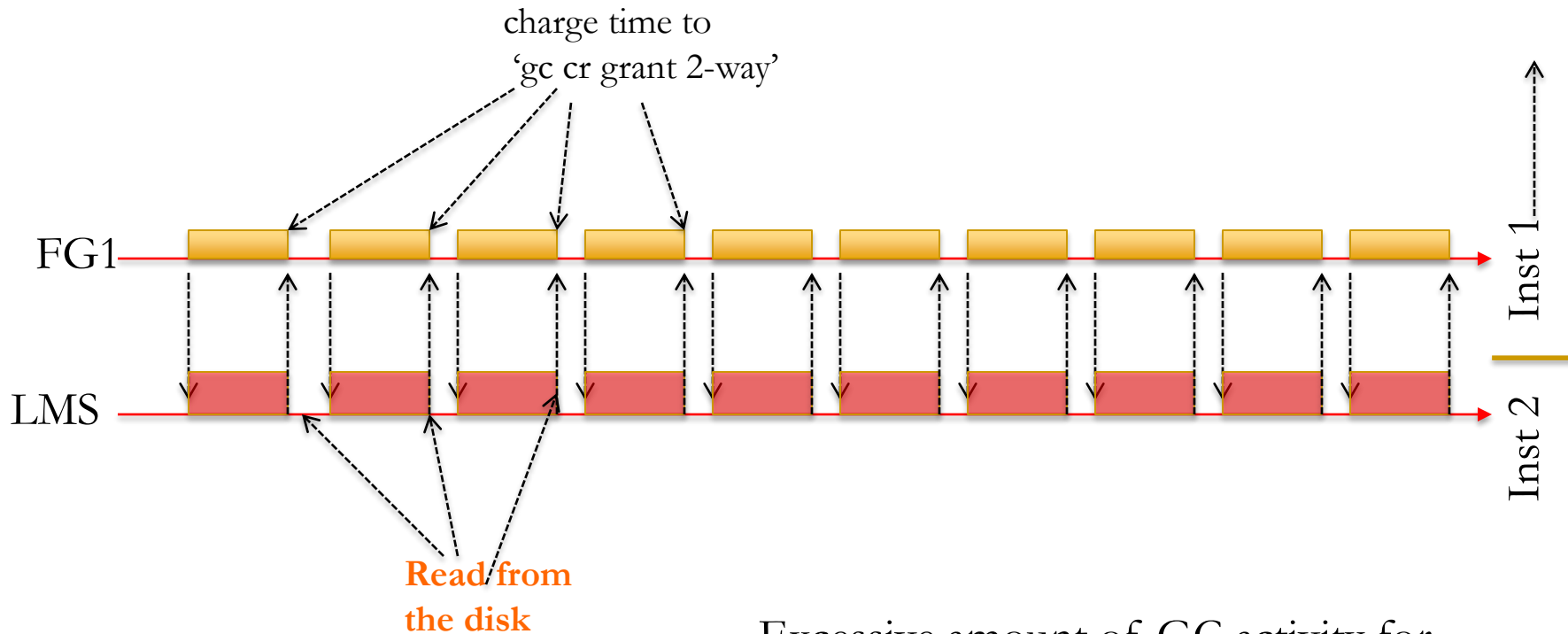# Congested

- 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.
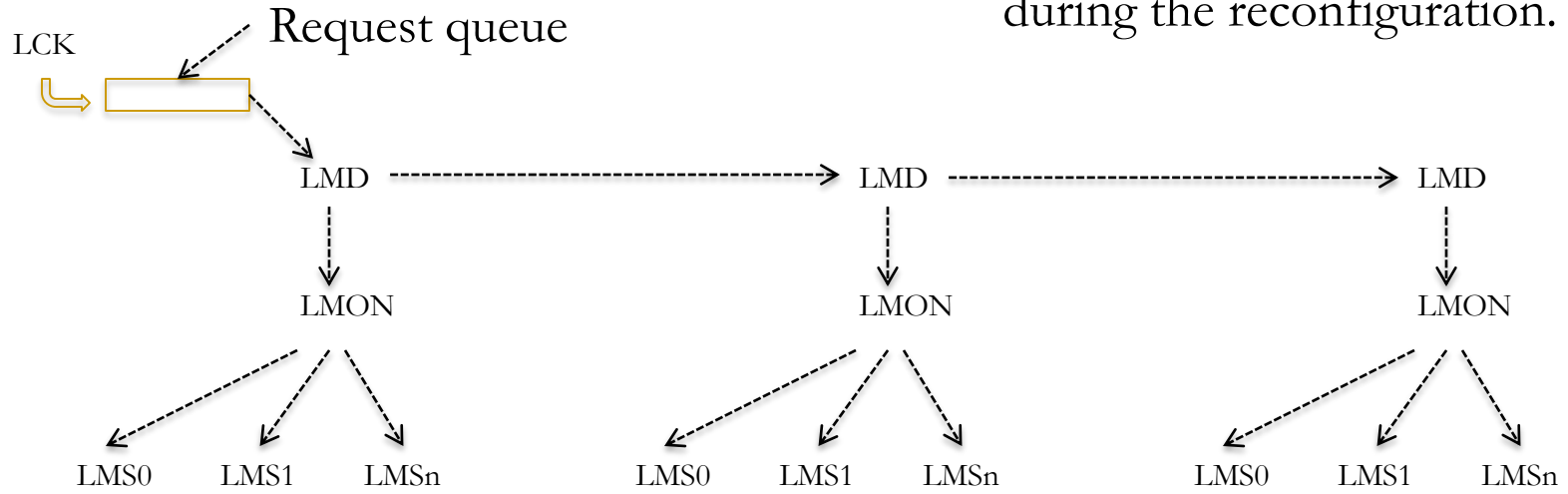
gc cr grants 2-way

Why DRM?

charge time to
'gc cr grant 2-way'

FG1

Inst 1

LMS

Inst 2

Read from
the disk
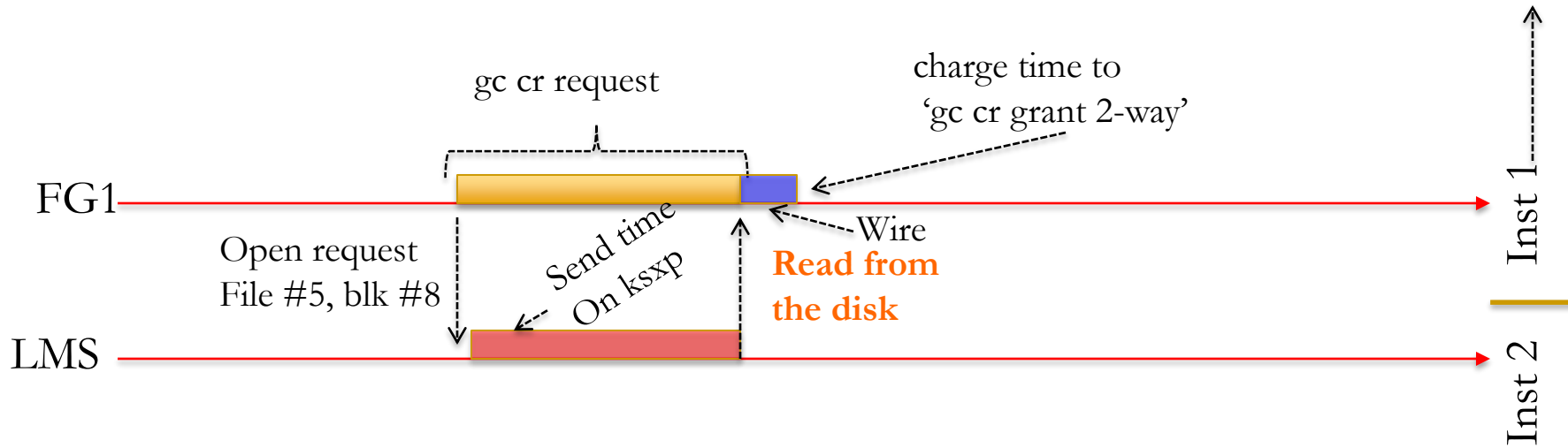
Excessive amount of GC activity for
One object!
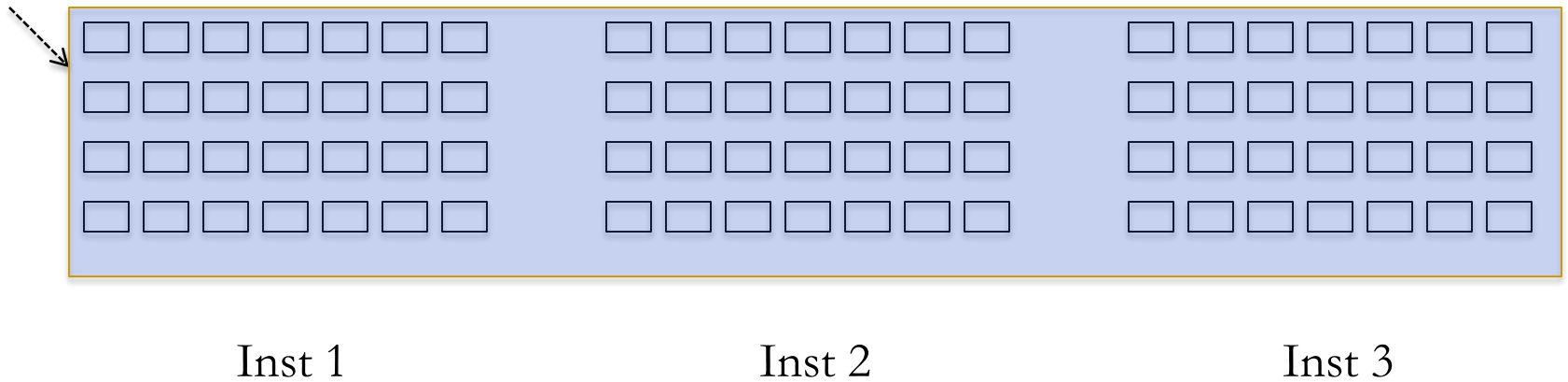
Demo: remaster demo. Refer sqldeveloper

DRM (11g)

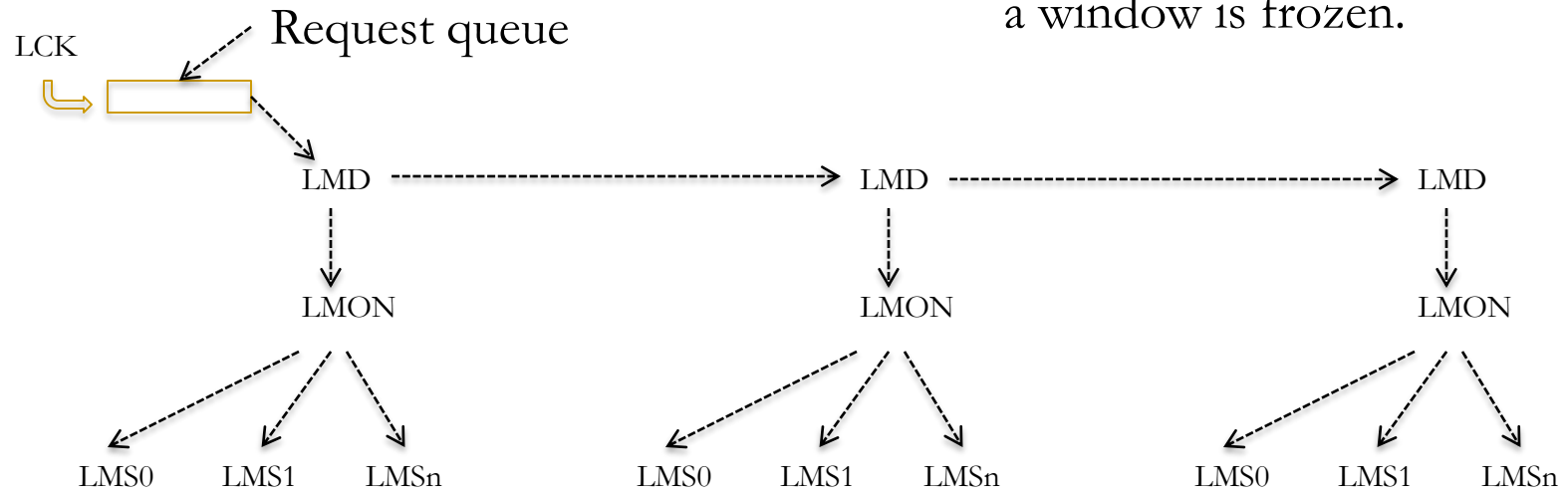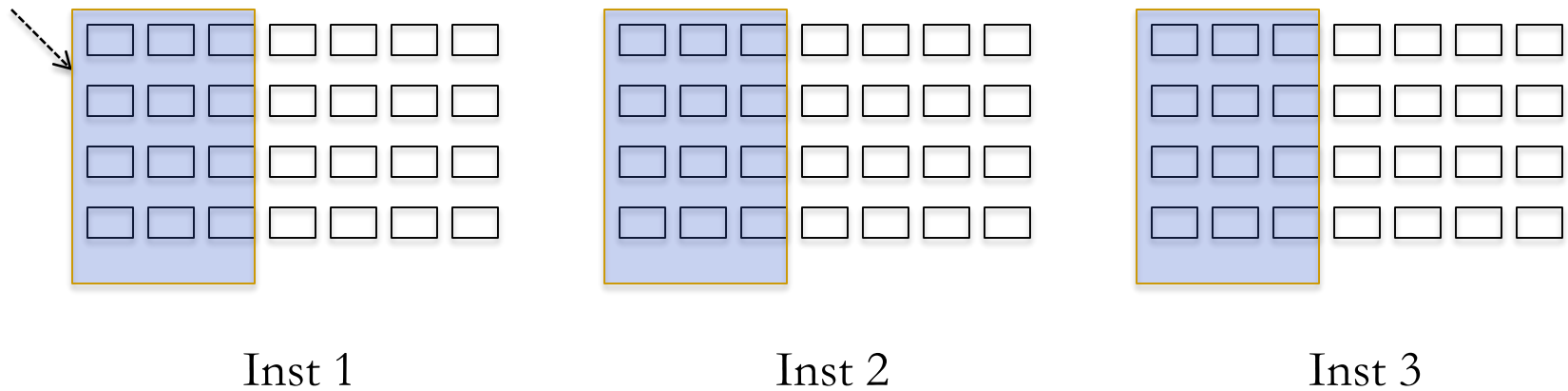In 11g, all resources are frozen during the reconfiguration.

LCK

Request queue

LMD - - - - - - - - - - - -> LMD - - - - - - - - - - - -> LMD

LMON   LMON   LMON

LMS0 LMS1 LMSn   LMS0 LMS1 LMSn   LMS0 LMS1 LMSn

Resources

Inst 1   Inst 2   Inst 3

DRM (12c)

In 12c, only set of resources in a window is frozen.

LCK

Request queue

LMD - - - - - - - - - - - → LMD - - - - - - - - - - - → LMD

LMON   LMON   LMON

LMS0   LMS1   LMSn   LMS0   LMS1   LMSn   LMS0   LMS1   LMSn

Resources

Inst 1   Inst 2   Inst 3

Resource names

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

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

RESOURCE_NAME
------------------------------
[0x15f29][0x0],[TM][ext 0x3,0x0]  <- hrdb1 PDB - GSTEST table -object_id=89897
[0x15f29][0x0],[TM][ext 0x4,0x0]  <- hrdb2 PDB - GSTEST table -object_id=89897
```
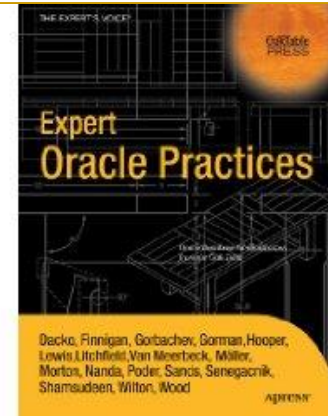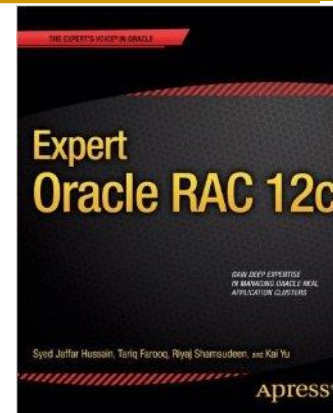
- Comes handy if you are debugging RAC trace files, to identify the PDB generating the errors.

THANK YOU

- *Email: rshamsud@orainternals.com*

- *Blog : orainternals.wordpress.com*

- *Web: www.orainternals.com*