

What's the Point of Oracle Checkpoints?

Harald van Breederode
Oracle University
4-DEC-2008

ORACLE

About Me

- **Senior Principal DBA Trainer – Oracle University**
- **25 years Unix Experience**
- **12 years Oracle DBA Experience**
- **Oracle8i, 9i, 10g and 11g OCP**
- **Oracle10g OCM**
- **DBA Certification Exam Team Reviewer**
- **DBA Curriculum Development Reviewer**
- **Enterprise Linux Certification Exam Global Lead**
- **Visually Impaired (Legally Blind)**



ORACLE

Agenda

- **Introduction**
- **Buffer Cache Basics**
- **Redo and the Buffer Cache**
- **What is a Checkpoint?**
- **Types of Checkpoint**
- **What is Driving your Checkpoints?**
- **Sizing Online Redo Log Files**
- **Conclusion**
- **Questions & Answers**

Buffer Cache Basics

The buffer cache component structures are:

- **Buffers**
 - Each buffer may hold an image of one data block at any one time
- **Buffer headers**
 - Store metadata about contents of the buffers
 - Act as cache management structures
- **Buffer pools**
 - Collection of buffers used for the same purpose and managed accordingly
- **Working set**
 - All or part of a buffer pool
 - Assigned to a DBWn process

Buffer Management

- **Cached buffers managed by doubly linked lists:**
- **REPL**
 - Buffers containing block images being used
- **REPL-AUX**
 - Buffers ready to be used for I/O or CR build
- **WRITE and CKPT-Q**
 - Dirty Buffers requiring I/O
- **WRITE-AUX**
 - Dirty Buffers with I/O in progress
- **Touch count is used to decide the initial insertion location in the REPL chain**
- **AUX lists avoid wasteful scanning**

Redo and the Buffer Cache

- **Block modification dirties the buffer containing the block image and generates redo**
- **A buffer becomes dirty at a particular RBA which is a point in the redo stream**
- **Redo written by LGWR makes the corresponding part of the redo log file “active”**
- **Dirty block images written by DBWn makes the corresponding part of the redo log file “inactive”**
- **Redo is always written prior to the corresponding block images**
- **Size of active redo in the log file influences instance and crash recovery time**
- **Trade-off between performance and recovery time**

Buffer Cache I/O

- **Servers look for an available buffer on REPL-AUX then read a data block into selected buffer**
 - Buffer gets moved from REPL-AUX to REPL
 - If block is modified, buffer is added to CKPT-Q
 - Servers move dirty buffers to WRITE during free buffer search
- **DBWn writes dirty buffer contents to database**
 - Buffer gets moved from WRITE to WRITE-AUX
 - Once block written:
 - Buffer is moved back to REPL-AUX
 - Buffer taken off CKPT-Q
- **DBWn writes upon request**
 - Make free buffers
 - Checkpoint

What is a Checkpoint?

- **A synchronization event at a specific point in time**
- **Causes some or all dirty block images to be written to the database thereby guaranteeing that blocks dirtied prior to that point in time get written**
- **Brings administration up to date**
- **Several types of checkpoint exist**

Types of Checkpoint

- **Full Checkpoint**
- **Thread Checkpoint**
- **File Checkpoint**
- **Object “Checkpoint”**
- **Parallel Query Checkpoint**
- **Incremental Checkpoint**
- **Log Switch Checkpoint**

Note: Some checkpoints can be logged to the alert log by setting `log_checkpoints_to_alert` to true

Full Checkpoint

- **Writes block images to the database for all dirty buffers from all instances**
- **Statistics updated:**
 - DBWR checkpoints
 - DBWR checkpoint buffers written
 - DBWR thread checkpoint buffers written
- **Caused by:**
 - Alter system checkpoint [global]
 - Alter database close
 - Shutdown
- **Controlfile and datafile headers are updated**
 - CHECKPOINT_CHANGE#

Question

**Does an Oracle instance
take a full checkpoint after
a log switch?**

Answer

As always, it depends!

**It did so up to and
including Oracle8 but has
not done so since Oracle8i**

Thread Checkpoint

- **Writes block images to the database for all dirty buffers from one instance**
- **Statistics updated:**
 - DBWR checkpoints
 - DBWR checkpoint buffers written
 - DBWR thread checkpoint buffers written
- **Caused by:**
 - Alter system checkpoint local
- **Controlfile and datafile headers are updated**
 - CHECKPOINT_CHANGE#

File Checkpoint

- **Writes block images to the database for all dirty buffers for all files of a tablespace from all instances**
- **Statistics updated:**
 - DBWR tablespace checkpoint buffers written
 - DBWR checkpoint buffers written
 - DBWR checkpoints
- **Caused by:**
 - Alter tablespace XXX offline
 - Alter tablespace XXX begin backup
 - Alter tablespace XXX read only
- **Controlfile and datafile headers are updated**
 - CHECKPOINT_CHANGE#

Parallel Query Checkpoint

- **Writes block images to the database for all dirty buffers belonging to objects accessed by the query from all instances**
- **Statistics updated:**
 - **DBWR checkpoint buffers written**
 - **DBWR checkpoints**
- **Caused by:**
 - **Parallel Query**
 - **Parallel Query component of PDML or PDDL**
 - **Mandatory for consistency**

Object “Checkpoint”

- **Writes block images to the database for all dirty buffers belonging to an object from all instances**
- **Statistics updated:**
 - **DBWR object drop buffers written**
 - **DBWR checkpoints**
- **Caused by:**
 - **Drop table XXX**
 - **Drop table XXX purge**
 - **Truncate table XXX**
- **Mandatory for media recovery purposes**

Incremental Checkpoint

- **Writes the contents of “some” dirty buffers to the database from CKPT-Q**
- **Block images written in SCN order**
- **Checkpoint RBA updated in SGA**
- **Statistics updated:**
 - **DBWR checkpoint buffers written**
- **Controlfile is updated every 3 seconds by CKPT**
 - **Checkpoint progress record**

Definition of “Some”

- **Every 3 seconds CKPT calculates the checkpoint target RBA based on:**
 - The most current RBA
 - `log_checkpoint_timeout`
 - `log_checkpoint_interval`
 - `fast_start_mttr_target`
 - `fast_start_io_target`
 - 90% of the size of the smallest online redo log file
- **All buffers dirtied prior to the time corresponding to the target RBA are written to the database**

Log Switch Checkpoint

- **Writes the contents of “some” dirty buffers to the database**
- **Statistics updated:**
 - DBWR checkpoints
 - DBWR checkpoint buffers written
 - background checkpoints started
 - background checkpoints completed
- **Controlfile and datafile headers are updated**
 - **CHECKPOINT_CHANGE#**

Checkpoint Administration

- **Useful checkpoint administration views:**
 - **V\$INSTANCE_RECOVERY**
 - **V\$SYSSTAT**
 - **V\$DATABASE**
 - **V\$INSTANCE_LOG_GROUP**
 - **V\$THREAD**
 - **V\$DATAFILE**
 - **V\$DATAFILE_HEADER**

Question

What is the point of Oracle Checkpoints?

Answer

The point of Oracle checkpoints is to synchronize all datafiles, some datafiles or some objects to a point in time for consistency, performance and recoverability purposes, but you must control them before they control you!

What is driving your Checkpoints?

V\$INSTANCE_RECOVERY

- **WRITES_MTTR**
 - fast_start_mttr_target
- **WRITES_LOGFILE_SIZE**
 - 90% of the smallest online redo log file
- **WRITES_LOG_CHECKPOINT_SETTINGS**
 - log_checkpoint_timeout
- **WRITES_OTHER_SETTINGS**
 - fast_start_io_target
- **WRITES_AUTOTUNE**
 - 10g self tuning checkpoints
- **WRITES_FULL_THREAD_CKPT**
 - Manual checkpoints

Sizing Online Redo Log Files

- Incremental checkpoint should not be driven by the size of the online redo log files
- They **MUST** be sized correctly
- Use Redo log file size advisor:
 - `OPTIMAL_LOGFILE_SIZE` in `V$INSTANCE_RECOVERY`
- Alternatively use:
 - `WRITES_LOGFILE_SIZE` in `V$INSTANCE_RECOVERY`
- Use `archive_lag_target` to control the log switch frequency
- On RAC databases, logs must be sized properly on all instances
- **Note: Online redo log files can only be too small, never too large!**

Conclusion and Summary

If performance is critical then be careful with the following, especially during peak hours:

- **Dropping or truncating objects**
- **Making tablespaces read only**
- **Placing the database or tablespaces in backup mode**
- **Parallel query on objects that have DML run against them**
- **Size of your redo log files**
- **Note: Be aware that on RAC systems these issues become global**

Q U E S T I O N S
&
A N S W E R S

And Finally

Thank you for your kind attention!

For a copy of my demonstration please email me at:

Harald.van.Breederode@oracle.com