RBA 和实例恢复关系

Redo Byte Address (RBA)参考文档

Recent entries in the redo thread of an Oracle instance are addressed using a 3-part redo byte address, or RBA. An RBA is comprised of:

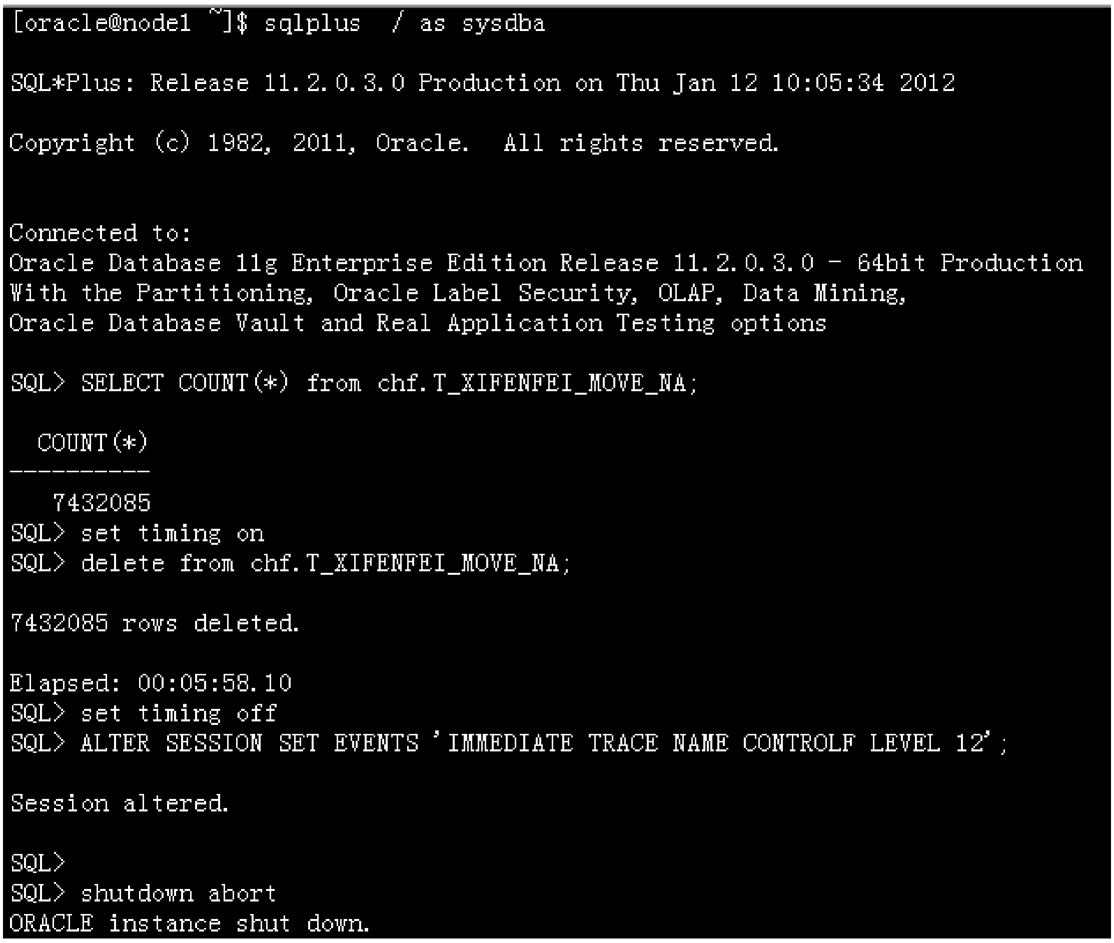
- the log file sequence number (4 bytes)
- the log file block number (4 bytes)
- the byte offset into the block at which the redo record starts (2 bytes)

RBAs are not necessarily unique within their thread, because the log file sequence number may be reset to 1 in all threads if a database is opened with the RESETLOGS option.

RBAs are used in the following important ways.
With respect to a dirty block in the buffer cache, **the low RBA** is the address of the redo for the first change that was applied to the block since it was last clean, and **the high RBA** is the address of the redo for the most recent change to have been applied to the block.
Dirty buffers are maintained on the buffer cache checkpoint queues in low RBA order. The checkpoint RBA is the point up to which DBWn has written buffers from the checkpoint queues if incremental checkpointing is enabled -- otherwise it is the RBA of last full thread checkpoint. The checkpoint RBA is copied into the checkpoint progress record of the controlfile by the checkpoint heartbeat once every 3 seconds. Instance recovery, when needed, begins from the checkpoint RBA recorded in the controlfile. The target RBA is the point up to which DBWn should seek to advance the checkpoint RBA to satisfy instance recovery objectives.
The on-disk RBA is the point up to which LGWR has flushed the redo thread to the online log files. DBWn may not write a block for which the high RBA is beyond the on-disk RBA. Otherwise transaction recovery (rollback) would not be possible, because the redo needed to undo a change is always in the same redo record as the redo for the change itself.
The term sync RBA is sometimes used to refer to the point up to which LGWR is required to sync the thread. However, this is not a full RBA -- only a redo block number is used at this point.

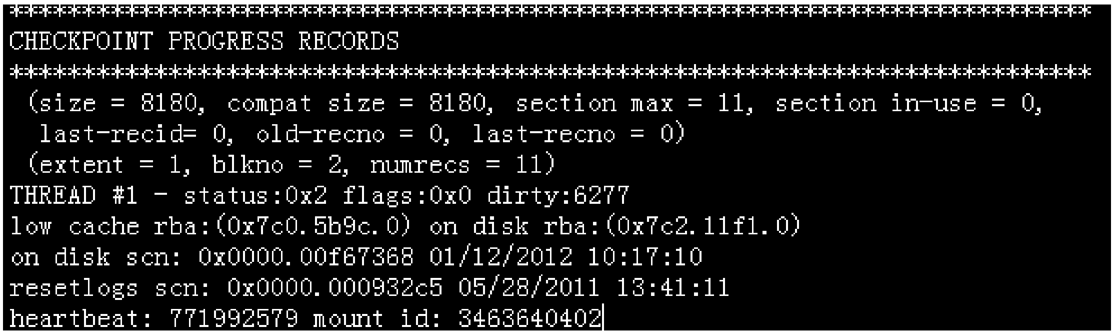
实例恢复模拟

删除表数据，跟踪控制文件，abort 数据库



```
[oracle@node1 ~]$ sqlplus  / as sysdba

SQL*Plus: Release 11.2.0.3.0 Production on Thu Jan 12 10:05:34 2012

Copyright (c) 1982, 2011, Oracle.  All rights reserved.


Connected to:
Oracle Database 11g Enterprise Edition Release 11.2.0.3.0 - 64bit Production
With the Partitioning, Oracle Label Security, OLAP, Data Mining,
Oracle Database Vault and Real Application Testing options

SQL> SELECT COUNT(*) from chf.T_XIFENFEI_MOVE_NA;

  COUNT(*)
----------
   7432085
SQL> set timing on
SQL> delete from chf.T_XIFENFEI_MOVE_NA;

7432085 rows deleted.

Elapsed: 00:05:58.10
SQL> set timing off
SQL> ALTER SESSION SET EVENTS 'IMMEDIATE TRACE NAME CONTROLF LEVEL 12’;

Session altered.

SQL>
SQL> shutdown abort
ORACLE instance shut down.
```

强制关闭数据库查看 controlfile 跟踪文件记录

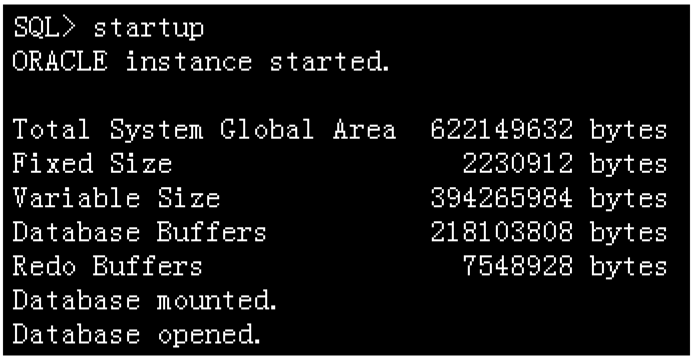


```
***************************************************************************
CHECKPOINT PROGRESS RECORDS
***************************************************************************
 (size = 8180, compat size = 8180, section max = 11, section in-use = 0,
  last-recid= 0, old-recno = 0, last-recno = 0)
 (extent = 1, blkno = 2, numrecs = 11)
THREAD #1 - status:0x2 flags:0x0 dirty:6277
low cache rba:(0x7c0.5b9c.0) on disk rba:(0x7c2.11f1.0)
on disk scn: 0x0000.00f67368 01/12/2012 10:17:10
resetlogs scn: 0x0000.000932c5 05/28/2011 13:41:11
heartbeat: 771992579 mount id: 3463640402
```

这里根据这里的提示，启动时候，需要恢复从 low cache rba:(0x7c0.5b9c.0) 到 on disk rba:(0x7c2.11f1.0)，一共有 6277 个脏数据块需要恢复

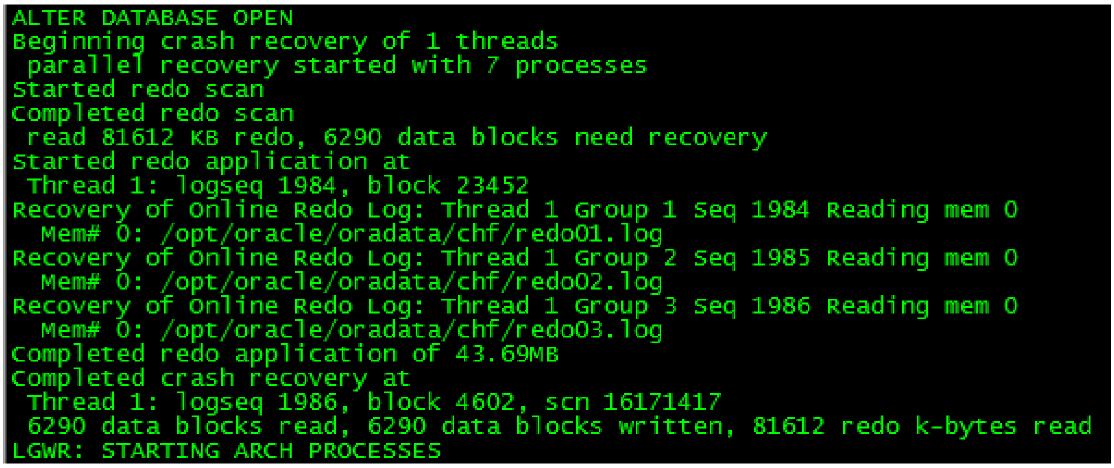
	RBA 信息	Log Sequence	Blcok Number
Low Cache RBA	0x7c0.5b9c.0	0x7c0 = 1984	5b9c=23452
On Disk RBA	0x7c2.11f1.0	0x7c2=1986	11f1=4593

启动数据库



```
SQL> startup
ORACLE instance started.

Total System Global Area  622149632 bytes
Fixed Size                  2230912 bytes
Variable Size             394265984 bytes
Database Buffers          218103808 bytes
Redo Buffers                7548928 bytes
Database mounted.
Database opened.
```

查看 alert 日志



```
ALTER DATABASE OPEN
Beginning crash recovery of 1 threads
 parallel recovery started with 7 processes
Started redo scan
Completed redo scan
 read 81612 KB redo, 6290 data blocks need recovery
Started redo application at
 Thread 1: logseq 1984, block 23452
Recovery of Online Redo Log: Thread 1 Group 1 Seq 1984 Reading mem 0
  Mem# 0: /opt/oracle/oradata/chf/redo01.log
Recovery of Online Redo Log: Thread 1 Group 2 Seq 1985 Reading mem 0
  Mem# 0: /opt/oracle/oradata/chf/redo02.log
Recovery of Online Redo Log: Thread 1 Group 3 Seq 1986 Reading mem 0
  Mem# 0: /opt/oracle/oradata/chf/redo03.log
Completed redo application of 43.69MB
Completed crash recovery at
 Thread 1: logseq 1986, block 4602, scn 16171417
 6290 data blocks read, 6290 data blocks written, 81612 redo k-bytes read
LGWR: STARTING ARCH PROCESSES
```

从这里可以看出一共恢复了 6290(在跟踪控制文件和 abort 之间，可能还有数据发生改变 6290-6277)，恢复过程中读取的 redo 日志为 1984,1985,1986

相关说明

1.关于 heartbeat 和 checkpoint

在这次的删除过程中我没有执行 commit，而是直接 abort 数据库。整个删除过程执行了近 6 分钟，控制文件的心跳每三秒进行一次，心跳是把 low cache rba 记录到了控制文件中，而没有真正的把全部的脏数据写入到磁盘(在发生了 checkpoint 时候，会把相关的脏数据写入到磁盘，而这里的控制文件的 heartbeat 和 checkpoint 是两回事，checkpoint 一般是在切换日志，数据文件正常离线，执行 begin backup 命令时发生，昨晚晚上后面的一个困惑就是上面的英文描述，让我把这两者搞混淆了)

2.三种 rba 解释

low rba ：在 buffer cache 中的数据块第一次数据改变所对应的 RAB。
high rba ：在 buffer cache 中的数据块最近一次数据改变时所对应的 RAB。
on-disk rba：是 lgwr 写日志文件的最末位置的地址。

3.实例恢复过程解释

实例恢复的时候，是从控制文件 heartbeat 记录的 low rba 开始读 redo log 数据(会多读取一点，因为 heartbeat 是每三秒执行一次，假设在 2.9 秒的时候，数据库异常 down 了，控制文件中记录的还是 2.9 秒前的 low rba，这个时候，从该点开始读取 redo)，恢复到 on-disk rba，而不是 high rba(high rba 一般情况下会大于 on-disk rba，但是因为 high rba 比 on-disk rba 多的部分记录在 redo log buffer 中，在实例恢复的时候，因为其未被记录到 redo log file 中，所以不能被恢复，其实也没有必要恢复，因为该数据肯定是没有 commit 或者 rollback)

在此分析过程中感谢**上海-select.简单**(736053407)和**北京-边城**(156157989)指点