

1 SCN概念

System Change Number. A database ordering primitive. The value of an SCN is the logical point in time at which changes are made to a database.

The system change number (SCN) is Oracle's clock - every time we commit, the clock increments. The SCN just marks a consistent point in time in the database.

A checkpoint is the act of writing dirty (modified blocks from the buffer cache to disk).

The database ALWAYS has transactions going on, ALWAYS. SMON and many other background processes are always doing work, the database (unless it is opened read only) is always doing transactions.

SCN是Oracle用来跟踪数据库内变化发生的先后顺序的机制，可以把它想象成一个高精度的时钟。每个Redo日志条目、UNDO data block、DATA block都会有SCN号。Oracle的Consistent-Read、Current-Read、Multiversion-Block都是依赖SCN实现。

“Consistent Read”就是前面介绍的“一致性读”，是指一个查询所看到的数据，是这个查询被提交的那一刻（SCN）系统的状态。

具体实现过程为：当一个查询语句被提交时，当时的SCN也被记录，这个查询用到的每个数据块的SCN都应该小于或等于提交时的SCN，这样的读取就是Current Read；如果数据块的SCN大于查询SCN，则说明这个数据块在查询发出后被修改过，已经不再是查询时的状态了。这时，系统会对这个数据块做一个拷贝，然后在拷贝上进行“回滚”操作，如果能够恢复到提交时的SCN状态，这个数据块就是一个“Consistent Version”的数据块，对它的读取就是“Consistent Read”；如果不能会督导提交时的SCN，就会抛出“ORA-01555 snapshot too old”。这就是Oracle的“Multi-versioning block”机制，Oracle就是利用这个机制平衡了“并发”和“数据一致”两个相互矛盾的诉求。

对于数据库的恢复操作，也是按照SCN顺序来重做日志中的内容，在单实例环境下，因为只有一个SCN产生器，所以对于按照SCN顺序重做日志就能够保证是按照改变发生的真正顺序来进行的，这就保证了准确的数据库恢复操作。

在RAC环境下，每个节点都有自己的SCN发生器，每个节点都会修改数据库，必须有某种机制保证这些SCN在时间排序上的精确，如果仅靠计算机的系统时钟肯定是不够的，必须引入额外的算法。

SCN号与oracle数据库恢复过程有着密切的关系，只有很好地理解了这层关系，才能深刻地理解恢复的原理，从而才能很好地解决这方面的问题。

SCN与CKPT进程在checkpoint发生时，将当时的SCN号写入数据文件头和控制文件，同时通知DBWR进程将数据块写到数据文件。

CKPT进程也会在控制文件中记录RBA(redo block address)，以标志Recovery需要从日志中哪个地方开始。与checkpoint相关的SCN号有四个，其中三个存在控制文件中，一个存放在数据文件头中，下面将分别介绍。

2 SCN分类

2.1 系统检查点scn (System Checkpoint SCN)

当一个检查点动作完成后，Oracle就把系统检查点的SCN存储到控制文件中。

```
SQL> alter session set nls_date_format='yyyy-mm-dd hh24:mi:ss';
```

Session altered.

```
SQL> SELECT
checkpoint_change#,scn_to_timestamp(checkpoint_change#)+0,current_scn,scn_to_timestamp(current_scn)+0
FROM v$database;
```

```
CHECKPOINT_CHANGE# SCN_TO_TIMESTAMP(CH CURRENT_SCN SCN_TO_TIMESTAMP(CU
-----
2354352 2018-07-31 10:51:19      2354499 2018-07-31 10:52:10
```

```
SQL> alter system checkpoint;
```

System altered.

```
SQL> SELECT
checkpoint_change#,scn_to_timestamp(checkpoint_change#)+0,current_scn,scn_to_timestamp(current_scn)+0
FROM v$database;
```

```
CHECKPOINT_CHANGE# SCN_TO_TIMESTAMP(CH CURRENT_SCN SCN_TO_TIMESTAMP(CU
-----
2354603 2018-07-31 10:52:58      2354941 2018-07-31 10:53:04
```

2.2 数据文件检查点scn (Datafile Checkpoint SCN)

当一个检查点动作完成后，Oracle就把每个数据文件的scn单独存放在控制文件中。

```
SQL> col name for a50
```

```
SQL> SELECT NAME,Checkpoint_change#,scn_to_timestamp(Checkpoint_change#)+0, checkpoint_time
FROM v$datafile;
```

```
NAME                                CHECKPOINT_CHANGE# SCN_TO_TIMESTAMP(CH CHECKPOINT_TIME
-----
/u01/app/oracle/oradata/orcl/system01.dbf      2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/flasharchive01.dbf 2354603 2018-07-31 10:52:58 2018-07-31
10:53:00
/u01/app/oracle/oradata/orcl/sysaux01.dbf      2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
```

/u01/app/oracle/oradata/orcl/undotbs01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/test01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/users01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00

6 rows selected.

2.3 开始scn (Start SCN)

Oracle把这个检查点的scn存储在每个数据文件的文件头中，这个值称为开始scn，因为它用于在数据库实例启动时，检查是否需要执行数据库恢复（实例恢复）。

```
SQL> SELECT NAME,Checkpoint_change#,scn_to_timestamp(Checkpoint_change#)+0, checkpoint_time
FROM v$datafile_header;
```

NAME	CHECKPOINT_CHANGE# SCN_TO_TIMESTAMP(CH CHECKPOINT_TIME
/u01/app/oracle/oradata/orcl/system01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/flasharchive01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/sysaux01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/undotbs01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/test01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/users01.dbf	2354603 2018-07-31 10:52:58 2018-07-31 10:53:00

6 rows selected.

执行检查点，并将表空间离线，离线的表空间对应的数据文件头部的SCN无法取到，而离线的表空间对应的数据文件存在控制文件头部的SCN保留在表空间离线的那一时间。

```
SQL> alter tablespace users offline;
```

Tablespace altered.

```
SQL> SELECT NAME,Checkpoint_change#,checkpoint_time FROM v$datafile_header;
```

NAME	CHECKPOINT_CHANGE# CHECKPOINT_TIME
/u01/app/oracle/oradata/orcl/system01.dbf	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/flasharchive01.dbf	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/sysaux01.dbf	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/undotbs01.dbf	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/test01.dbf	2354603 2018-07-31 10:53:00

0

6 rows selected.

```
SQL> SELECT NAME,Checkpoint_change# ,scn_to_timestamp(Checkpoint_change#)+0, checkpoint_time
FROM v$datafile;
```

NAME	CHECKPOINT_CHANGE#	SCN_TO_TIMESTAMP(CH	CHECKPOINT_TIME
/u01/app/oracle/oradata/orcl/system01.dbf	2354603	2018-07-31 10:52:58	2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/flasharchive01.dbf	2354603	2018-07-31 10:52:58	2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/sysaux01.dbf	2354603	2018-07-31 10:52:58	2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/undotbs01.dbf	2354603	2018-07-31 10:52:58	2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/test01.dbf	2354603	2018-07-31 10:52:58	2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/users01.dbf	2355644	2018-07-31 10:58:30	2018-07-31 10:58:30

6 rows selected.

2.4 结束scn (Stop SCN)

每个数据文件的终止scn都存储在控制文件中，在数据库正常运行的情况下，对可读写的，online的数据文件，该SCN号为NULL。

```
SQL> SELECT NAME,status,enabled,Checkpoint_change# ,checkpoint_time,last_change# FROM v$datafile;
```

NAME	STATUS	ENABLED	CHECKPOINT_CHANGE#
CHECKPOINT_TIME	LAST_CHANGE#		
/u01/app/oracle/oradata/orcl/system01.dbf	SYSTEM	READ WRITE	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/flasharchive01.dbf	ONLINE	READ WRITE	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/sysaux01.dbf	ONLINE	READ WRITE	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/undotbs01.dbf	ONLINE	READ WRITE	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/test01.dbf	ONLINE	READ WRITE	2354603 2018-07-31 10:53:00
/u01/app/oracle/oradata/orcl/users01.dbf	OFFLINE	DISABLED	2355644 2018-07-31 10:58:30

6 rows selected.

2.5 与检查点无关的SCN

2.5.1 日志文件SCN(Lowest system change number (SCN) in the log)

```
SQL> select group#,sequence#,status,first_change#,first_time from v$log;
```

GROUP#	SEQUENCE#	STATUS	FIRST_CHANGE#	FIRST_TIME
1	1	INACTIVE	2174098	2018-07-30 17:44:19
2	2	CURRENT	2252659	2018-07-31 00:05:09
3	0	UNUSED	0	

记录日志时，自动记录最新的SCN 到日志中。因此，每一条日志都包含了一个时间。当DBWn 进程启动时，将依照日志记录写入一段数据。这一段日志记录的时间段必定有一个最早时间和一个最新（迟）时间。最早的时间我们称为LOW SCN（即上次清空日志后的第一条记录），最新时间我们称为NEXT SCN（即下次日志清空前的最近一条记录）。这里，由CKPT 进程将NEXT SCN 的值同步更新到START SCN、SYSTEM CHECKPOINT SCN 和DATAFILE CHECKPOINT SCN 中。以确保所有文件的一致性。

2.5.2 当前current scn

```
SQL> select to_char(dbms_flashback.get_system_change_number) from dual;
```

```
TO_CHAR(DBMS_FLASHBACK.GET_SYSTEM_CHANGE  
-----  
2356944
```

```
SQL> select to_char(current_scn) system_scn from v$database;
```

```
SYSTEM_SCN  
-----  
2356963
```

```
SQL> select to_char(TIMESTAMP_TO_SCN(SYSDATE)) from dual;
```

```
TO_CHAR(TIMESTAMP_TO_SCN(SYSDATE))  
-----  
2356977
```

```
SQL> SELECT TO_CHAR(SCN_TO_TIMESTAMP(2356977)) from dual;
```

```
TO_CHAR(SCN_TO_TIMESTAMP(2356977))  
-----  
31-JUL-18 11.06.25.000000000 AM
```

SCN实际是和系统时间挂钩的，通过SYSDATE可以自动生成SCN号。

3 数据库运行期间SCN值

在数据库打开并运行之后，控制文件中的系统检查点、控制文件中的数据文件检查点scn和每个数据文件头中的开始scn都是相同的。控制文件中的每个数据文件的结束scn都为null。

```
SQL> SELECT checkpoint_change# FROM v$database;
```

```
CHECKPOINT_CHANGE#
```

```
-----  
2358164
```

```
SQL> SELECT NAME,checkpoint_change#,last_change# FROM v$datafile;
```

```
NAME                                CHECKPOINT_CHANGE# LAST_CHANGE#
```

```
-----  
/u01/app/oracle/oradata/orcl/system01.dbf      2358164  
/u01/app/oracle/oradata/orcl/flasharchive01.dbf 2358164  
/u01/app/oracle/oradata/orcl/sysaux01.dbf      2358164  
/u01/app/oracle/oradata/orcl/undotbs01.dbf     2358164  
/u01/app/oracle/oradata/orcl/test01.dbf        2358164  
/u01/app/oracle/oradata/orcl/users01.dbf       2358164
```

6 rows selected.

```
SQL> SELECT NAME,checkpoint_change# FROM v$datafile_header;
```

```
NAME                                CHECKPOINT_CHANGE#
```

```
-----  
/u01/app/oracle/oradata/orcl/system01.dbf      2358164  
/u01/app/oracle/oradata/orcl/flasharchive01.dbf 2358164  
/u01/app/oracle/oradata/orcl/sysaux01.dbf      2358164  
/u01/app/oracle/oradata/orcl/undotbs01.dbf     2358164  
/u01/app/oracle/oradata/orcl/test01.dbf        2358164  
/u01/app/oracle/oradata/orcl/users01.dbf       2358164
```

6 rows selected.

在完全关闭数据库的过程中，系统会执行一个检查点动作，这时所有数据文件的结束scn都会设置成数据文件头中的那个开始scn的值。在数据库重新启动的时候，Oracle将文件头中的那个开始scn与数据库文件检查点scn进行比较，如果这两个值相互匹配，oracle接下来还要比较数据文件头中的开始scn和控制文件中数据文件的结束scn。如果这两个值也一致，就意味着所有数据块多已经提交，所有对数据库的修改都没有在关闭数据库的过程中丢失，因此这次启动数据库的过程也不需要任何恢复操作，此时数据库就可以打开了。当所有的数据库都打开之后，存储在控制文件中的数据文件结束scn的值再次被更改为null，这表示数据文件已经打开并能够正常使用了。

4 SCN与数据库启动/关闭

4.1 SCN与数据库启动

4.1.1 检查是否需要介质恢复

在数据库启动过程中，当System Checkpoint SCN、Datafile Checkpoint SCN和Start SCN 号都相同时，数据库可以正常启动，不需要做media recovery。三者当中有一个不同时，则需要做media recovery。

4.1.2 检查是否需要实例恢复

如果在启动的过程中，End SCN 号为NULL，则需要做instance recovery，ORACLE 在启动过程中首先检查是否需要media recovery，然后再检查是否需要instance recovery。

4.2 SCN与数据库关闭

如果数据库正常关闭的话，将会触发一个checkpoint，同时将数据文件的END SCN号设置为相应数据文件的Start SCN号，当数据库启动时，发现它们是一致的，则不需要做instance recovery，在数据库正常启动后，ORACLE会将END SCN号设置为NULL，如果数据库异常关闭的话，则END SCN号将为NULL。

4.2.1 数据库正常关闭

```
SQL> shutdown immediate
Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> startup mount
ORACLE instance started.
```

Total System Global Area 788529152 bytes

Fixed Size 8625656 bytes

Variable Size 566231560 bytes

Database Buffers 209715200 bytes

Redo Buffers 3956736 bytes

Database mounted.

```
SQL> SELECT checkpoint_change# FROM v$database;
```

CHECKPOINT_CHANGE#

```
-----
2359407
```

```
SQL> SELECT NAME,checkpoint_change#,last_change# FROM v$datafile;
```

NAME	CHECKPOINT_CHANGE#	LAST_CHANGE#
------	--------------------	--------------

```

-----
/u01/app/oracle/oradata/orcl/system01.dbf      2359407      2359407
/u01/app/oracle/oradata/orcl/flasharchive01.dbf      2359407      2359407
/u01/app/oracle/oradata/orcl/sysaux01.dbf      2359407      2359407
/u01/app/oracle/oradata/orcl/undotbs01.dbf      2359407      2359407
/u01/app/oracle/oradata/orcl/test01.dbf      2359407      2359407
/u01/app/oracle/oradata/orcl/users01.dbf      2359407      2359407

```

6 rows selected.

```
SQL> SELECT NAME,checkpoint_change# FROM v$datafile_header;
```

```

NAME                                CHECKPOINT_CHANGE#
-----
/u01/app/oracle/oradata/orcl/system01.dbf      2359407
/u01/app/oracle/oradata/orcl/flasharchive01.dbf      2359407
/u01/app/oracle/oradata/orcl/sysaux01.dbf      2359407
/u01/app/oracle/oradata/orcl/undotbs01.dbf      2359407
/u01/app/oracle/oradata/orcl/test01.dbf      2359407
/u01/app/oracle/oradata/orcl/users01.dbf      2359407

```

6 rows selected.

正常关闭时，Start SCN=End SCN，不需要恢复。

4.2.2 数据库异常关闭

```
SQL> alter database open;
```

Database altered.

```
SQL> create table stu.t_test(id number,name varchar2(20));
```

Table created.

```
SQL> insert into stu.t_test values(1,'Alen');
```

1 row created.

```
SQL> commit;
```

Commit complete.

```
SQL> insert into stu.t_test values(2,'LS');
```

1 row created.

SQL> shutdown abort;

ORACLE instance shut down.

SQL> startup mount

ORACLE instance started.

Total System Global Area 788529152 bytes

Fixed Size 8625656 bytes

Variable Size 566231560 bytes

Database Buffers 209715200 bytes

Redo Buffers 3956736 bytes

Database mounted.

SQL> SELECT checkpoint_change# FROM v\$database;

CHECKPOINT_CHANGE#

2359410

SQL> SELECT NAME,checkpoint_change#,last_change# FROM v\$datafile;

NAME CHECKPOINT_CHANGE# LAST_CHANGE#

/u01/app/oracle/oradata/orcl/system01.dbf 2359410
/u01/app/oracle/oradata/orcl/flasharchive01.dbf 2359410
/u01/app/oracle/oradata/orcl/sysaux01.dbf 2359410
/u01/app/oracle/oradata/orcl/undotbs01.dbf 2359410
/u01/app/oracle/oradata/orcl/test01.dbf 2359410
/u01/app/oracle/oradata/orcl/users01.dbf 2359410

6 rows selected.

SQL> SELECT NAME,checkpoint_change# FROM v\$datafile_header;

NAME CHECKPOINT_CHANGE#

/u01/app/oracle/oradata/orcl/system01.dbf 2359410
/u01/app/oracle/oradata/orcl/flasharchive01.dbf 2359410
/u01/app/oracle/oradata/orcl/sysaux01.dbf 2359410
/u01/app/oracle/oradata/orcl/undotbs01.dbf 2359410
/u01/app/oracle/oradata/orcl/test01.dbf 2359410
/u01/app/oracle/oradata/orcl/users01.dbf 2359410

6 rows selected.

异常关闭时，End SCN为空。

未提交的事务回滚：

```
SQL> alter database open;
```

Database altered.

```
SQL> select *from stu.t_test;
```

```
ID NAME
```

```
-----  
1 Alen
```

查看Alert信息，数据库打开时需进行实例恢复：

```
alter database open  
2018-07-31T11:38:19.854154+08:00  
Ping without log force is disabled:  
instance mounted in exclusive mode.  
2018-07-31T11:38:19.960625+08:00  
Beginning crash recovery of 1 threads  
2018-07-31T11:38:20.115977+08:00  
Started redo scan  
2018-07-31T11:38:20.836838+08:00  
Completed redo scan  
read 458 KB redo, 68 data blocks need recovery  
2018-07-31T11:38:20.964336+08:00  
Started redo application at  
Thread 1: logseq 2, block 249899, offset 0
```

5 讨论

1 为什么需要System checkpoint SCN号与Datafile Checkpoint SCN号？

为什么ORACLE会在控制文件中记录System checkpoint SCN号的同时，还需要为每个数据文件记录Datafile Checkpoint SCN号？

原因有二：

- 1) 对只读表空间，其数据文件的Datafile Checkpoint SCN、Start SCN和END SCN号均相同。这三个SCN在表空间处于只读期间都将被冻结。
- 2) 如果控制文件不是当前的控制文件，则System checkpoint会小于Start SCN或END SCN号。记录这些SCN号，可以区分控制文件是否是当前的控制文件。

```
SQL> alter tablespace users read only;
```

```
SQL> SELECT NAME,checkpoint_change#,last_change# FROM v$datafile;
```

```
SQL> SELECT NAME,checkpoint_change# FROM v$datafile_header;
```

```
SQL> SELECT checkpoint_change# FROM v$database;
```