

故障排除: "log file sync"等待 (Doc ID 1626301.1)

文档内容

- 用途
- 排错步骤
 - 什么是'log file sync'等待事件?
 - 用户应该搜集那些信息, 来初步分析'log file sync'等待事件?
 - 什么原因造成了很高的'log file sync'等待?
 - 影响 LGWR 的 IO 性能问题
 - 比较'log file sync'和'log file parallel write'的平均等待时间。
 - 建议
 - 间歇性物理IO缓慢对 'log file sync' 等待事件的影响
 - 检查 LGWR trace
 - 建议
 - 检查在线重做日志是否足够大
 - 建议
 - 应用程序提交过多
 - 比较 user commit/rollback 同 user calls 比值的平均值
 - 建议
 - 其他可能相关的等待事件
 - Adaptive Log File Sync
 - Redo Synch Time Overhead
 - 已知问题
 - 其他诊断程序来帮助分析'log file sync'等待?
 - Data Guard 和'log file sync'
 - 其他问题故障排除
- 参考

适用于:

Oracle Cloud Infrastructure - Database Service - 版本 N/A 和更高版本
Oracle Database Cloud Exadata Service - 版本 N/A 和更高版本
Oracle Database Exadata Express Cloud Service - 版本 N/A 和更高版本
Oracle Database Cloud Schema Service - 版本 N/A 和更高版本
Oracle Database Backup Service - 版本 N/A 和更高版本
本文档所含信息适用于所有平台

用途

本文旨在帮助诊断存在大量'log file sync'等待事件的问题。

排错步骤

什么是'log file sync'等待事件?

当一个用户会话提交, 那个用户产生的所有redo需从内存刷新到重做日志文件, 使事务对数据库的修改永久化。

在提交时, 用户会话会通知 LGWR 把日志缓冲区中的信息写到重做日志文件 (当前所有未被写入磁盘的 redo 信息, 包括本次会话的 redo 信息)。当 LGWR 发现写操作完成后, 它会通知用户会话。当等待 LGWR 通知确认所有 redo 已经安全的保存到磁盘的过程时, 用户会话会等待'log file sync'。

用户会话显示等待'log file sync', 是用户会话通知 LGWR 和 LGWR 通知用户的写操作完成之间的时间。

注意在11.2及以上版本, LGWR会在默认的post/wait模式和polling模式之间自动切换。在polling模式下, 写操作的进展会维护在一个内存结构中, 等待'log file sync'的会话可以不时检查这个内存结构看是否自己所发起的事务对应的redo已经写入磁盘。这时, 等待时间会从用户会话通知 LGWR开始到会话自己发现对应的redo已经写入磁盘结束。

需要注意的是, 如果已有一个正在进行的同步, 其他需要提交的会话 (为了保存日志信息) 也需等待 LGWR, 进而也将等待'log file sync'?

用户应该搜集那些信息，来初步分析'log file sync'等待事件？

初步分析等待'log file sync'，下面的信息是有帮助的：

- 'log file sync'等待在可接受范围的类似时间的 AWR 报告，作为用于比较的性能基线
- 'log file sync'等待很严重期间的 AWR 报告
注：2 个报告应在 10-30 分钟之间。
- LGWR 日志文件（在12.2版本以上还需收集LGnn日志文件）
当日志写入较慢的时候，LGWR 日志文件将会显示警告信息

什么原因造成了很高的'log file sync'等待？

'log file sync'可以在用户会话通知 LGWR 写日志，和 LGWR 写完日志后通知用户会话（本地重做日志以及同步模式下的远程备库的日志），及用户会话被唤醒间来接受LGWR的通知或者poll LGWR的写入的日志的进展的任何一个点发生。
更多详情，请参照文档：

[Document:34592.1](#) WAITEVENT: "log file sync"

其中的最常见的原因：

- [影响 LGWR 的 I/O 性能问题](#)
- [过多的应用程序 commit](#)

这些原因以及如何解决它们详情概述如下：

影响 LGWR 的 IO 性能问题

我们在这里回答的主要问题是“是否 LGWR 写入磁盘慢？”，下面的步骤可以帮助确定是否是这个导致的。

比较'log file sync'和'log file parallel write'的平均等待时间。

等待事件'log file parallel write'表示 LGWR 正在等待写 redo 操作。该事件的持续时间就是等待 IO 操作部分的时间。关于'log file parallel write'的更多信息，请参阅：

[Document:34583.1](#) WAITEVENT: "log file parallel write" Reference Note

结合事件“log file sync”看同步操作消耗在 IO 的时间，由此推断，有多少处理时间消耗在 CPU 上。

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
log file sync	868,667	62,544	72	58.29	Commit

Foreground Wait Events

- s - second, ms - millisecond - 1000th of a second
- Only events with Total Wait Time (s) >= .001 are shown
- ordered by wait time desc, waits desc (idle events last)
- %Timeouts: value of 0 indicates value was < .5%. Value of null is truly 0

Event	Waits	%Time -outs	Total Wait Time (s)	Avg wait (ms)	Waits /txn	% DB time
log file sync	868,667	0	62,544	72	1.01	58.29

Background Wait Events

- ordered by wait time desc, waits desc (idle events last)
- Only events with Total Wait Time (s) >= .001 are shown
- %Timeouts: value of 0 indicates value was < .5%. Value of null is truly 0

Event	Waits	%Time -outs	Total Wait Time (s)	Avg wait (ms)	Waits /txn	% bg time
log file parallel write	93,144	0	5,495	59	0.11	40.31

上面的例子显示了'log file sync' 和 'log file parallel write' 都有很高的等待时间

如果'log file sync'的时间消耗在'log file parallel write'上的比例高，那么大部分的等待时间是由于 IO（等待 redo 写入）。应该检查 LGWR 在 IO 方面的性能。作为一个经验法则，'log file parallel write'平均时间超过 20 毫秒，意味着 IO 子系统有问题。（当然这个时间对于拥有更多cache的现代的存储系统以及 SSD, NVRAM来说会更小）

- 与系统管理员一起检查重做日志所在的文件系统以及本地卷的位置，以提高 IO 性能。
- 不要把重做日志放在需要额外计算的较老的或者复杂的RAID上，比如 RAID-5或者RAID-6或者只有很少cache部件的存储上
- 不要把重做日志放在上一代或者较老的Solid State Disk (SSD)磁盘上
虽然通常情况下，SSD 写入性能好于平均水平，他们可能会遇到写峰值,从而导致大量的严重'log file sync'等待并引发数据库性能不稳定或者hung住(关于这一点您需要详尽的测试，因为我们也碰到一些即使SSD的写性能不稳定但数据库性能仍然可以接受的系统)(Engineered Systems (Exadata, SuperCluster 和 Oracle Database Appliance) 除外，因为在这些系统上已经为使用SSD来存放重做日志而做了额外的优化)
- 监控其他可能需要写到相同路径的进程，确保该磁盘具有足够的带宽，足以应付所要求的容量。如果不能满足,增加硬盘或者更快的磁盘设备或者把这些文件分散到不同的磁盘设备。
- 确保 LOG_BUFFER 不要太大，一个非常大的 log_buffer 的不利影响就是刷新需要更长的等待时间。当缓冲区满了的时候，它必须将所有数据写入到重做日志文件。LGWR 将一直等待,直到最后的 I/O 完成。

间歇性物理IO缓慢对 'log file sync' 等待事件的影响

LGWR倾向于做很多小的IO操作，而不是大块的IO操作。大部分的磁盘配置并不能在这种场景下很好的工作，可能会发生间歇性物理IO缓慢。但是从平均等待时间来看，IO等待的时间并不长（它们仅仅是不能很好地处理这种小而多的IO负载），磁盘设备提供商据此断定没有磁盘问题。因为系统里还有其它的IO操作，这些正常的IO操作的等待时间很短，所有这些IO操作平均起来的等待时间并不长，这就掩盖了间歇性物理IO缓慢（IO异常值）的问题。但是间歇性物理IO缓慢的问题会造成很高的'log file sync'，虽然平均的'log file parallel write'是处于正常性能的范围。

如果你发现系统的'log file sync'很高，但是'log file parallel write'是处于正常的范围，那么这可能是由于间歇性物理IO缓慢导致的。可以检查AWR的部分，也可以使用一些像OSWatcher一样的工具(参照 [Document 301137.1](#))来确定是否存在间歇性物理IO缓慢。如果可以确定存在间歇性物理IO缓慢，那么你需要与磁盘提供商一起来解决这个问题。

检查 LGWR trace

尽管'log file parallel write'的平均等待时间可能在一个合理的区间范围内，在峰值时刻写操作时间还是可能会很长进而影响'log file sync'的等待时间。从 10.2.0.4 开始如果写操作超过 500 毫秒我们会在 LGWR 的 trace 中写警告信息。这个阈值很高所以就算没有警告也不代表没有问题。警告信息如下：

```
*** 2011-10-26 10:14:41.718
Warning: log write elapsed time 21130ms, size 1KB
(set event 10468 level 4 to disable this warning)

*** 2011-10-26 10:14:42.929
Warning: log write elapsed time 4916ms, size 1KB
(set event 10468 level 4 to disable this warning)
```

注意：上面的峰值如果时间间隔的很远,可能不会对'log file parallel wait'有大的影响。但是，如果有 100 个会话等待'log file parallel wait'完成，'log file sync'总等待可能就会很高，因为等待时间将被乘以会话的个数 100。因此，值得探讨日志写 IO 高峰的原因。

请参阅：

[Document:601316.1](#) LGWR Is Generating Trace file with "Warning: Log Write Time 540ms, Size 5444kb" In 10.2.0.4 Database

建议

- 与系统管理员一起检查其他正在发生的可能会导致 LGWR 写磁盘峰值的操作
- 当 LGWR 进程慢的时候，对其进行 Truss 操作会帮助确定时间消耗在什么地方

注意：这些警告信息对于预防潜在问题的发生很有帮助。就算平均等待时间没问题，通过找到 I/O 性能峰值点，DBA 可以知道 LGWR 会间歇性的遇到性能问题，进而在它引发更大问题前将其解决。

检查在线重做日志是否足够大

每次重做日志切换到下一个日志时，会执行'log file sync'操作，以确保下一个日志开始之前信息都写完。标准建议是日志切换最多 15 至 20 分钟一次。如果切换比这更频繁，那么将发生更多的'log file sync'操作，意味着更多的会话等待。

- 检查 alert.log 日志文件切换的时间

```
Thu Jun 02 14:57:01 2011
Thread 1 advanced to log sequence 2501 (LGWR switch)
Current log# 5 seq# 2501 mem# 0: /opt/oracle/oradata/<SID>/redo05a.log
Current log# 5 seq# 2501 mem# 1: /opt/oracle/logs/<SID>/redo05b.log
Thu Nov 03 14:59:12 2011
Thread 1 advanced to log sequence 2502 (LGWR switch)
Current log# 6 seq# 2502 mem# 0: /opt/oracle/oradata/<SID>/redo06a.log
Current log# 6 seq# 2502 mem# 1: /opt/oracle/logs/<SID>/redo06b.log
Thu Nov 03 15:03:01 2011
Thread 1 advanced to log sequence 2503 (LGWR switch)
```

```
Current log# 4 seq# 2503 mem# 0: /opt/oracle/oradata/<SID>/redo04a.log
Current log# 4 seq# 2503 mem# 1: /opt/oracle/logs/<SID>/redo04b.log
```

在上面的例子中，我们看到每 2 到 4 分钟进行日志切换，这比建议值的5倍还高。

- 您也可以检查 AWR 报告日志切换的平均时间

Instance Activity Stats - Thread Activity		
Statistics identified by '(derived)' come from sources other than SYSSTAT		
Statistic	Total	per Hour
log switches (derived)	15	29.98

在上面的例子中基于 AWR 中的信息，每小时有 29.98 次重做日志切换：每 2 分钟切换一次。这个比每 15-20 分钟切换一次的建议值要高,并将影响前台进程需要等待'log file sync'完成的时间，因为发起同步操作的开销比必要的多。

建议

增加redo logs的大小

[Document:602066.1](#) How To Maintain and/or Add Redo Logs
[Document:779306.1](#) How To Add/Increase The Size Of Redo Log Files In Rac Environment?

应用程序提交过多

在这种情况下，要回答的问题是“是否应用程序 commit 过于频繁？”。

如果是，那么过多的 commit 活动可能会导致性能问题，因为把 redo 从日志缓冲区刷新到重做日志可能会导致等待'log file sync'。

如果'log file sync'的平均等待时间比'log file parallel write'高很多，这意味着大部分时间等待不是由于等待 redo 的写入，因而问题的原因不是 IO 慢导致。剩余时间是 CPU 活动，而且是过多的 commit 导致的最常见的竞争。

此外，如果'log file sync'的平均等待时间低，但等待次数高，那么应用程序可能 commit 过于频繁。

比较 user commit/rollback 同 user calls 比值的平均值

在 AWR 或 Statspack 报告中，如果每次 commit/rollback 的平均 user calls ("user calls/(user commits+user rollbacks)") 小于 30，表明 commit 过于频繁

Instance Activity Stats			
Ordered by statistic name			
Statistic	Total	per Second	per Trans
user calls	4,962,433	2,755.00	5.76
user commits	861,662	478.37	1.00

在上面的例子中，我们看到，平均每 5.76 次 user calls 就会有一次 commit, 大约高出建议值 5 倍。基于经验，我们期望 user calls/user commit 至少是 25。当然，这取决于应用程序设计。

建议

- 如果有很多短事务，看是否可能把这些事务组合在一起，从而减少 commit 操作。因为每一个 commit 都必须收到相关 REDO 已写到磁盘上的确认，额外的 commit 会显著的增加开销。虽然 Oracle 可以将某些 commit 组合在一起，通过事务的批处理来减少commit的总体数量还是可以带来非常有益的效果。
- 看看是否有操作可以使用 COMMIT NOWAIT 选项（务必在使用前应明白语义）。
- 看看是否有操作可以安全地使用 NOLOGGING/ UNRECOVERABLE 选项完成。

其他可能相关的等待事件

检查 AWR 报告，看是否有跟 LGWR 相关的,显示占用了显著数量时间的其他事件,因为这可能会给出导致这个问题的一个线索。前台和后台事件都应该进行检查。

例如下面的 AWR 显示某些其他前台和后台等待事件等待高，意味着传输重做日志到远程位置的问题，这可能会导致 fore gorund 进程等待"log file sync"。

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
log file sync	32,108	32,902	1025	94.09	Commit
log file switch completion	35	1,008	28813	2.88	Configuration
DB CPU		511		1.46	
buffer busy waits	874	267	305	0.76	Concurrency
control file sequential read	473,353	105	0	0.30	System I/O

After 'log file sync', the next wait event is 'log file switch completion'. Each switch takes an average of 28813ms to complete.

Background Wait Events

- ordered by wait time desc, waits desc (idle events last)
- Only events with Total Wait Time (s) >= .001 are shown
- %Timeouts: value of 0 indicates value was < .5%. Value of null is truly 0

Event	Waits	%Time-outs	Total Wait Time (s)	Avg wait (ms)	Waits /txn	% bg time
LNS wait on SENDREQ	22,290	0	3,426	154	0.71	46.77
LGWR-LNS wait on channel	324,024	91	3,334	10	10.32	45.51

High percentage of time is spent on events related to transporting redo logs.

Adaptive Log File Sync

11.2 中引入了 Adaptive Log File sync, 由参数 `_use_adaptive_log_file_sync` 控制, 在 11.2.0.1 和 11.2.0.2 默认设置为 false。从 11.2.0.3 开始默认是 true。当启用时, Oracle 可以在两种方法之间切换:

- Post/wait, 传统发布写重做日志完成的方法
- Polling, 一种新的方法, 其中前台进程会检查 LGWR 是否已写完成。

更多关于这个特性的信息, 请参阅:

[Document 1541136.1](#) Waits for "log file sync" with Adaptive Polling vs Post/Wait Choice Enabled

Redo Synch Time Overhead

统计值 'redo synch time overhead' 在 11.2.0.4 和 12c 被引入, 记录了理想的 log file sync 时间以及真正的 log file sync 时间的差值。如果这个差值很小, 说明很大的 log file sync 可能是由于离群 (就是偏离平均等待时间很大的 log file parallel write 等待) 的 log file parallel write 等待导致的。

已知问题

关于 'log file sync' 的已知 bug, 请参照:

[Document 34592.1](#) WAITEVENT: "log file sync" Reference Note

下列文件可以帮助特定环境下的已知问题:

[Document 1548261.1](#) High Waits for 'Log File Sync': Known Issue Checklist for 11.2

[Document 1229104.1](#) LOG FILE SYNC WAITS SPIKES DURING RMAN ARCHIVELOG BACKUPS

[Document 1541136.1](#) Waits for "log file sync" with Adaptive Polling vs Post/Wait Choice Enabled

Sparc:

[Document 1523164.1](#) SPARC: Reducing High Waits on 'log file sync' on Oracle Solaris SPARC by Increasing Priority of Log Writer

[Document 1205673.1](#) 'Log File Sync' problem on a Sun Server: A Typical Source for LOGFILE SYNC Performance Problems

[Document 10318123.8](#) Bug 10318123 - Solaris: LGWR regularly stalls for 3 seconds at a time

AIX:

[Document 1318709.1](#) AIX: Things To Check When Seeing Long Log File Sync Time in 11.2

Veritas/Symantec ODM:

[Document 1278149.1](#) Intermittent Long 'log file sync' Waits on Veritas ODM, LGWR Posting Long Write Times, I/O Portion of Wait Minimal

[Document 13551402.8](#) High "log file parallel write" and "log file sync" after upgrading 11.2 with Veritas/Symantec ODM

SUSE Linux:

[Document 1942134.1](#) Higher Log File Synch After OS Upgrade from SUSE Linux Enterprise Server SLES11 SP1 to SLES11 SP3

Exadata:

关于不要把 redo 放置在 SSD/Flash 存储上 (除了 Engineered System) 的信息请参考以下文档的第 11 页

<http://www.oracle.com/technetwork/database/exadata/exadata-smart-flash-cache-366203.pdf>

其他诊断程序来帮助分析'log file sync'等待?

下面的脚本着眼于与'log file sync'等待有关的重要参数, 'log file sync'等待直方图数据和相关信息。

[Document:1064487.1](#) Script to Collect Log File Sync Diagnostic Information (lfsdiag.sql)

Data Guard 和'log file sync'

当使用 Data Guard SYNC 和 commit WAIT 默认配置时, 可能需要更多的时间。在 Data Guard 环境中, 虽然上述调整步骤仍然适用, 网络写和 RFS/redo 写入备用重做日志的时间也需要加以考虑。下面的技术手册解释了'Log File Sync'如何适用于 Data Guard:

[Document:387174.1](#) MAA - Data Guard Redo Transport and Network Best Practices.

在主数据库和备用数据库的设置信息可以通过从以下文档的脚本来收集:

[Document:241374.1](#) Script to Collect Data Guard Primary Site Diagnostic Information

[Document:241438.1](#) Script to Collect Data Guard Physical Standby Diagnostic Information

其他问题故障排除

对于排除其他性能问题的指导, 请参阅:

[Document:1377446.1](#) Troubleshooting Performance Issues

参考

[NOTE:1064487.1](#) - Script to Collect Log File Sync Diagnostic Information (lfsdiag.sql)

[NOTE:1278149.1](#) - Intermittent Long 'log file sync' Waits on Veritas ODM, LGWR Posting Long Write Times, I/O Portion of Wait Minimal

[NOTE:1318709.1](#) - AIX: Long "log file sync" Wait Time in 11.2. : Things To Check

[NOTE:34592.1](#) - WAITEVENT: "log file sync" Reference Note

[NOTE:241374.1](#) - Script to Collect Data Guard Primary Site Diagnostic Information for Version 9i

[NOTE:779306.1](#) - How To Add/Increase The Size Of Redo Log Files In Rac Environment?

[NOTE:13551402.8](#) - Bug 13551402 - High "log file parallel write" and "log file sync" after upgrading 11.2 with Veritas/Symantec ODM

[NOTE:34583.1](#) - WAITEVENT: "log file parallel write" Reference Note

[NOTE:241438.1](#) - This document has been replaced with more recent information on this topic. Please refer to more recent documentation.

[NOTE:1377446.1](#) - * Troubleshooting Performance Issues

[NOTE:601316.1](#) - "Warning: log write time" Messages Seen in LGWR Trace Files

[NOTE:10318123.8](#) - Bug 10318123 - LGWR regularly stalls for 3 seconds at a time

[NOTE:387174.1](#) - MAA - Data Guard Redo Transport and Network Best Practices

[NOTE:1205673.1](#) - 'Log File Sync' problem on a Sun Server: Discussion of a Potential Source for LOGFILE SYNC Performance Problems plus references

[NOTE:1229104.1](#) - LOG FILE SYNC WAITS SPIKES DURING RMAN ARCHIVELOG BACKUPS

[NOTE:602066.1](#) - How To Maintain and/or Add Redo Logs

Didn't find what you are looking for?