



图文 42 redo log是直接一条一条写入文件的吗？非也，揭秘redo log block！

713 人次阅读 2020-03-19 10:48:24

📱 手机观看

详情 评论

redo log是直接一条一条写入文件的吗？非也，揭秘redo log block！

- **如何提问：**每篇文章都有评论区，大家可以尽情留言提问，我会逐一答疑
- **如何加群：**购买狸猫技术窝专栏的小伙伴都可以加入狸猫技术交流群，一个非常纯粹的技术交流的地方

具体加群方式，请参见目录菜单下的文档：《MySQL专栏付费用户如何加群》（购买后可见）

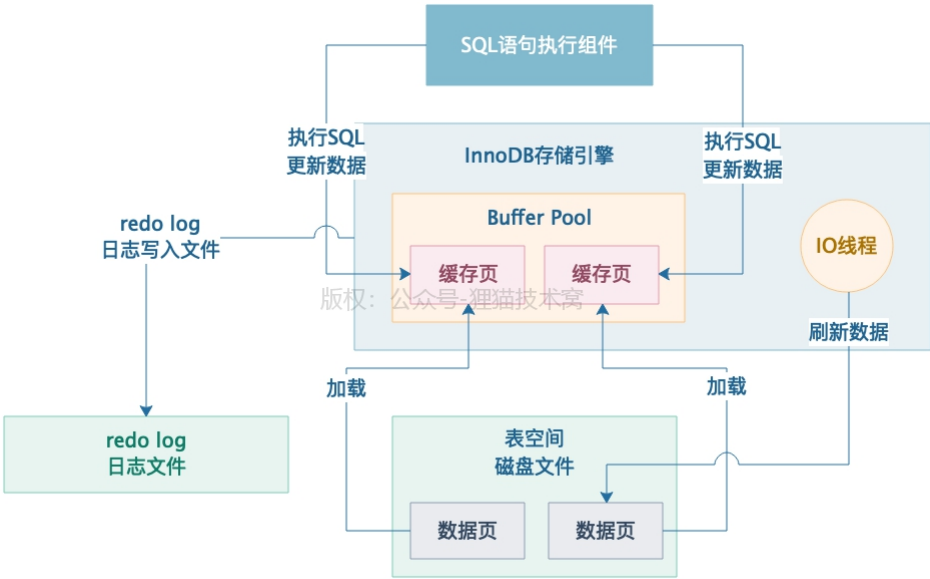
之前我们已经给大家讲解了redo log自己的一些基本的结构，今天我们就来讲解下一个问题，就是redo log是一条一条的直接就往磁盘文件里写入吗？

可能有一些朋友会认为就是如此简单粗暴的往磁盘文件里写，但其实没那么简单！

接下来几天我们就会揭秘一下这个redo log写磁盘的过程。

首先大家看下面的图，学习到现在，我想任何一个朋友一看下面的图就知道是怎么回事了。

平时我们执行CRUD的时候，从磁盘加载数据页到buffer pool的缓存页里去，然后对缓存页执行增删改，同时还会写redo log到日志文件里去，后续不定时把缓存页刷回磁盘文件里去，大概就是这个原理，如下图所示：



那么上次我们也介绍了一下每一条redo log长什么样子，说白了，他就是记录了：

表空间号+数据页号+数据页内偏移量+修改了几个字节的数据+实际修改数据

就是简简单单这么一条日志罢了

所以大家可以想一下，redo log就是按照上述格式，一条一条的直接就写入到磁盘上的日志文件里去了吗？



狸猫技术窝

进店逛逛

相关频道



从零开始带你成为MySQL实战优化高手
已更新60期

显然不是的！

其实MySQL内有另外一个数据结构，叫做**redo log block**，大概你可以理解为，平时我们的数据不是存放在数据页了的么，用一页一页的数据页来存放数据。

那么对于redo log也不是单行单行的写入日志文件的，他是用一个redo log block来存放多个单行日志的。

一个redo log block是512字节，这个redo log block的512字节分为3个部分，一个是12字节的header块头，一个是496字节的body块体，一个是4字节的trailer块尾

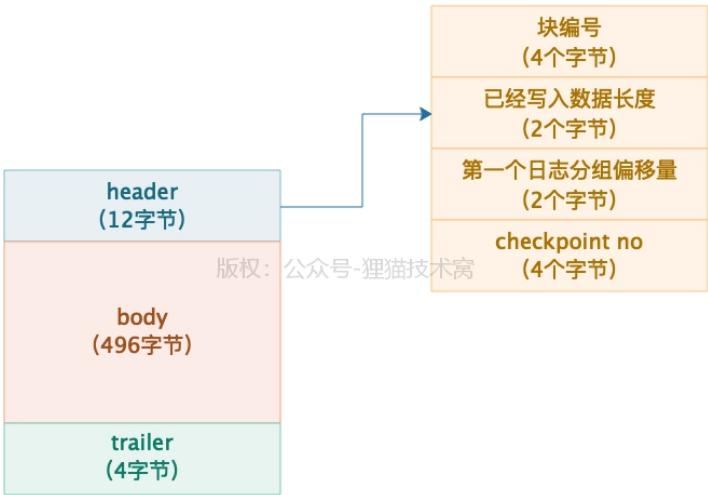
如下图所示



在这里面，12字节的header头又分为了4个部分。

- 1. 包括4个字节的block no，就是块唯一编号；
- 2. 2个字节的data length，就是block里写入了多少字节数据；
- 3. 2个字节的first record group。这个是说每个事务都会有多个redo log，是一个redo log group，即一组redo log。那么在这个block里的第一组redo log的偏移量，就是这2个字节存储的；
- 4. 4个字节的checkpoint on

我们看下图，这个header可以进行进一步的区分。

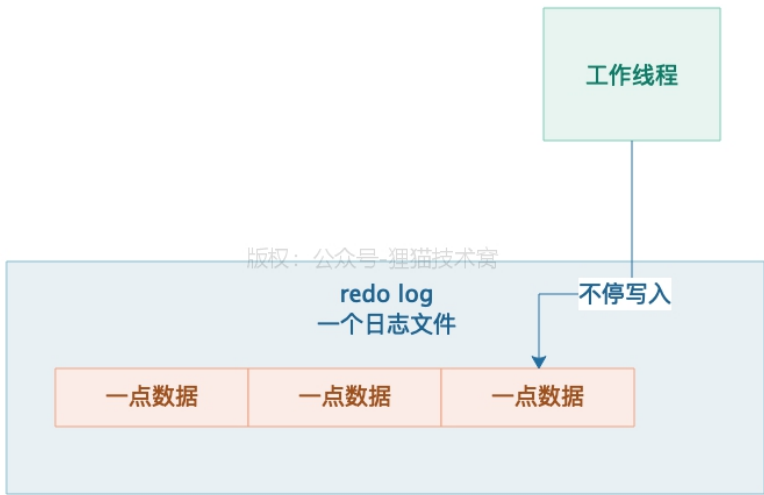


所以我们看到上图就知道，其实对于我们的redo log而言，他确实是不停的追加写入到redo log磁盘文件里去的，但是其实每一个redo log都是写入到文件里的一个redo log block里去的，一个block最多放496自己的redo log日志。

此时可能有人会有疑问了，到底一个一个的redo log block在日志文件里是怎么存在的？那么一条一条的redo log又是如何写入日志文件里的redo log block里去的呢？估计很多人都很奇怪这个问题。

所以我们接下来就给大家解答这个问题。

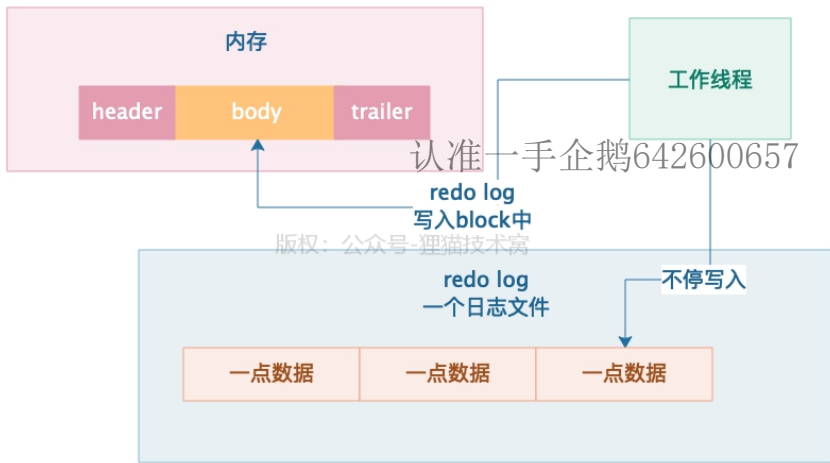
大家先想一下，假设你有一个redo log日志文件，平时我们往里面写数据，你大致可以认为是从第一行开始，从左往右写，可能会有很多行，比如下面这样子，你看看是不是你理解的那样？



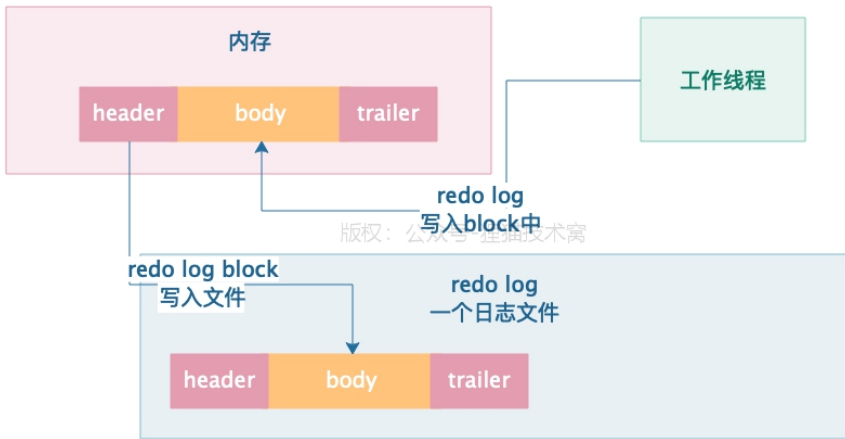
好，那么所以现在既然如此，假设你要写第一个redo log了，是不是应该起码是先在内存里把这个redo log给弄到一个redo log block数据结构里去？

然后似乎你应该是等内存里的一个redo log block的512字节都满了，再一次性把这个redo log block写入磁盘文件？

如下图所示



然后其实按照我们所说的，一个redo log block就是512字节，那么是不是真正写入的时候，把这个redo log block的512字节的数据，就写入到redo log文件里去就可以了？那么redo log文件里就多了一个block，如下图所示。

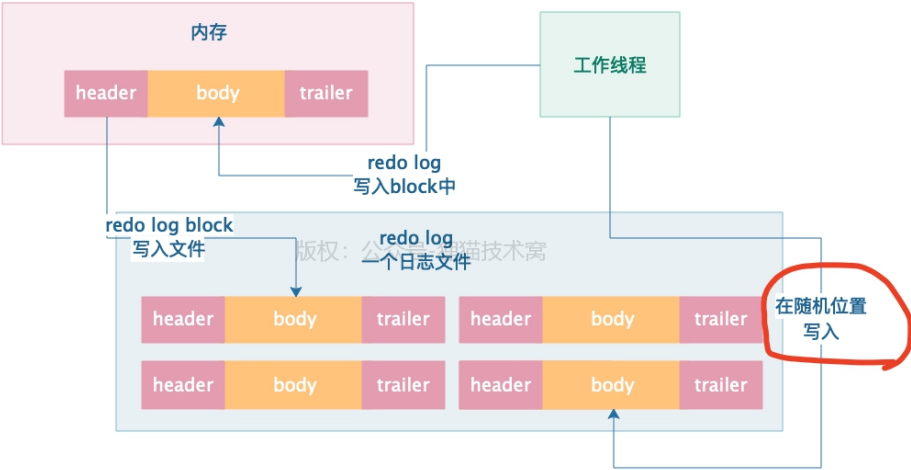


所以大家看到上图演示之后，对于这个所谓的redo log和redo log block的关系，以及redo log block如何进入日志文件，日志文件里是如何存放一个又一个的redo log block的，应该都很清楚了！

其实有一定开发经验的朋友都知道，写文件的时候，可以按照字节，一个字节一个字节的写入的，文件里存放的东西就是很多很多字节，依次排开，然后其中可能512个字节组合起来，就固定代表了一个redo log block。

这其实就是任何一个中间件系统，数据库系统，底层依赖磁盘文件存储数据的一个共同的原理，所以大家也不用把这个复杂数据写入磁盘文件想象的太复杂了。

那么如果依次在磁盘文件里的末尾追加不停的写字节数据，就是磁盘顺序写；但是假设现在磁盘文件里已经有很多很多的redo log block了，此时要在磁盘里某个随机位置找到一个redo log block去修改他里面几个字节的数据，这就是磁盘随机写，看下图。



好了，今天把redo log block的数据结构和他与磁盘文件的关系都讲的很清楚了，明天我们继续讲解redo log buffer，就是redo log是如何通过一个内存缓冲数据结构之后，再进入到磁盘文件的！

End

专栏版权归公众号狸猫技术窝所有


未经许可不得传播，如有侵权将追究法律责任

狸猫技术窝精品专栏及课程推荐：

- [《从零开始带你成为消息中间件实战高手》](#)
- [《21天互联网Java进阶面试训练营》（分布式篇）](#)
- [《互联网Java工程师面试突击》（第1季）](#)
- [《互联网Java工程师面试突击》（第3季）](#)
- [《从零开始带你成为JVM实战高手》](#)

认准一手企鵝642600657

Copyright © 2015-2020 深圳小鵝網絡技術有限公司 All Rights Reserved. [粵ICP備15020529號](#)

 小鵝通提供技術支持