

18 | 如何用硬件同步原语（CAS）替代锁？

2019-09-03 李玥

消息队列高手课

[进入课程 >](#)



讲述：李玥

时长 12:03 大小 13.80M



你好，我是李玥。上节课，我们一起学习了如何使用锁来保护共享资源，你也了解到，使用锁是有一定性能损失的，并且，如果发生了过多的锁等待，将会非常影响程序的性能。

在一些特定的情况下，我们可以使用硬件同步原语来替代锁，可以保证和锁一样的数据安全性，同时具有更好的性能。

在今年的 NSDI（NSDI 是 USENIX 组织开办的关于网络系统设计的著名学术会议）上，伯克利大学发表了一篇论文《[Confluo: Distributed Monitoring and Diagnosis Stack for High-speed Networks](#)》，这个论文中提到的 Confluo，也是一个类似于消息队列的流数据存储，它的吞吐量号称是 Kafka 的 4~10 倍。对于这个实验结论我个人不是很认同，因为它设计的实验条件对 Kafka 来说不太公平。但不可否认的是，Confluo 它的这个设计思路是一个创新，并且实际上它的性能也非常好。

Confluo 是如何做到这么高的吞吐量的呢？这里面非常重要的一个创新的设计就是，它使用硬件同步原语来代替锁，在一个日志上（你可以理解为消息队列中的一个队列或者分区），保证严格顺序的前提下，实现了多线程并发写入。


今天，我们就来学习一下，如何用硬件同步原语（CAS）替代锁？

什么是硬件同步原语？

为什么硬件同步原语可以替代锁呢？要理解这个问题，你要首先知道硬件同步原语是什么。

硬件同步原语（Atomic Hardware Primitives）是由计算机硬件提供的一组原子操作，我们比较常用的原语主要是 CAS 和 FAA 这两种。

CAS（Compare and Swap），它的字面意思是：先比较，再交换。我们看一下 CAS 实现的伪代码：

 复制代码

```
1 << atomic >>
2 function cas(p : pointer to int, old : int, new : int) returns bool {
3     if *p ≠ old {
4         return false
5     }
6     *p ← new
7     return true
8 }
```

它的输入参数一共有三个，分别是：

p: 要修改的变量的指针。

old: 旧值。


new: 新值。

返回的是一个布尔值，标识是否赋值成功。

通过这个伪代码，你就可以看出 CAS 原语的逻辑，非常简单，就是先比较一下变量 p 当前的值是不是等于 old，如果等于，那就把变量 p 赋值为 new，并返回 true，否则就不改变

变量 `p`，并返回 `false`。

这是 CAS 这个原语的语义，接下来我们看一下 FAA 原语（Fetch and Add）：

 复制代码

```
1 << atomic >>
2 function faa(p : pointer to int, inc : int) returns int {
3     int value <- *location
4     *p <- value + inc
5     return value
6 }
```

FAA 原语的语义是，先获取变量 `p` 当前的值 `value`，然后给变量 `p` 增加 `inc`，最后返回变量 `p` 之前的值 `value`。

讲到这儿估计你会问，这两个原语到底有什么特殊的呢？

上面的这两段伪代码，如果我们用编程语言来实现，肯定是无法保证原子性的。而原语的特殊之处就是，它们都是由计算机硬件，具体说就是 CPU 提供的实现，可以保证操作的原子性。

我们知道，**原子操作具有不可分割性，也就不存在并发的**问题。所以在某些情况下，原语可以用来替代锁，实现一些即安全又高效的并发操作。

CAS 和 FAA 在各种编程语言中，都有相应的实现，可以来直接使用，无论你是使用哪种编程语言，它们底层的实现是一样的，效果也是一样的。

接下来，还是拿我们熟悉的账户服务来举例说明一下，看看如何使用 CAS 原语来替代锁，实现同样的安全性。

CAS 版本的账户服务

假设我们有一个共享变量 `balance`，它保存的是当前账户余额，然后我们模拟多个线程并发转账的情况，看一下如何使用 CAS 原语来保证数据的安全性。


这次我们使用 Go 语言来实现这个转账服务。先看一下使用锁实现的版本：

```
1 package main
2
3 import (
4     "fmt"
5     "sync"
6 )
7
8 func main() {
9     // 账户初始值为 0 元
10    var balance int32
11    balance = int32(0)
12    done := make(chan bool)
13    // 执行 10000 次转账，每次转入 1 元
14    count := 10000
15
16    var lock sync.Mutex
17
18    for i := 0; i < count; i++ {
19        // 这里模拟异步并发转账
20        go transfer(&balance, 1, done, &lock)
21    }
22    // 等待所有转账都完成
23    for i := 0; i < count; i++ {
24        <-done
25    }
26    // 打印账户余额
27    fmt.Printf("balance = %d \n", balance)
28 }
29 // 转账服务
30 func transfer(balance *int32, amount int, done chan bool, lock *sync.Mutex) {
31    lock.Lock()
32    *balance = *balance + int32(amount)
33    lock.Unlock()
34    done <- true
35 }
```

这个例子中，我们让账户的初始值为 0，然后启动多个协程来并发执行 10000 次转账，每次往账户中转入 1 元，全部转账执行完成后，账户中的余额应该正好是 10000 元。

如果你没接触过 Go 语言，不了解协程也没关系，你可以简单地把它理解为进程或者线程都可以，这里我们只是希望能异步并发执行转账，我们并不关心这几种“程”他们之间细微的差别。

这个使用锁的版本，反复多次执行，每次 balance 的结果都正好是 10000，那这段代码的安全性是没问题的。接下来我们看一下，使用 CAS 原语的版本。

 复制代码

```
1 func transferCas(balance *int32, amount int, done chan bool) {
2     for {
3         old := atomic.LoadInt32(balance)
4         new := old + int32(amount)
5         if atomic.CompareAndSwapInt32(balance, old, new) {
6             break
7         }
8     }
9     done <- true
10 }
```

这个 CAS 版本的转账服务和上面使用锁的版本，程序的总体结构是一样的，主要的区别就在于，“异步给账户余额 +1”这一小块儿代码的实现。

那在使用锁的版本中，需要先获取锁，然后变更账户的值，最后释放锁，完成一次转账。我们可以看一下使用 CAS 原语的实现：

首先，它用 for 来做了一个没有退出条件的循环。在这个循环的内部，反复地调用 CAS 原语，来尝试给账户的余额 +1。先取得账户当前的余额，暂时存放在变量 old 中，再计算转账之后的余额，保存在变量 new 中，然后调用 CAS 原语来尝试给变量 balance 赋值。我们刚刚讲过，CAS 原语它的赋值操作是有前置条件的，只有变量 balance 的值等于 old 时，才会将 balance 赋值为 new。


我们在 for 循环中执行了 3 条语句，在并发的环境中执行，这里面会有两种可能情况：

一种情况是，执行到第 3 条 CAS 原语时，没有其他线程同时改变了账户余额，那我们是安全变更账户余额的，这个时候执行 CAS 的返回值一定是 true，转账成功，就可以退出循环了。并且，CAS 这一条语句，它是一个原子操作，赋值的安全性是可以保证的。

另外一种情况，那就是在这个过程中，有其他线程改变了账户余额，这个时候是无法保证数据安全的，不能再进行赋值。执行 CAS 原语时，由于无法通过比较的步骤，所以不会执行赋值操作。本次尝试转账失败，当前线程并没有对账户余额做任何变更。由于返回值为 false，不会退出循环，所以会继续重试，直到转账成功退出循环。

这样，每一次转账操作，都可以通过若干次重试，在保证安全性的前提下，完成并发转账操作。

其实，对于这个例子，还有更简单、性能更好的方式：那就是，直接使用 FAA 原语。

 复制代码

```
1 func transferFaa(balance *int32, amount int, done chan bool) {  
2     atomic.AddInt32(balance, int32(amount))  
3     done <- true  
4 }
```

FAA 原语它的操作是，获取变量当前的值，然后把它做一个加法，并且保证这个操作的原子性，一行代码就可以搞定了。看到这儿，你可能会想，那 CAS 原语还有什么意义呢？

在这个例子里面，肯定是使用 FAA 原语更合适，但是我们上面介绍的，使用 CAS 原语的方法，它的适用范围更加广泛一些。类似于这样的逻辑：先读取数据，做计算，然后更新数据，无论这个计算是什么样的，都可以使用 CAS 原语来保护数据安全，但是 FAA 原语，这个计算的逻辑只能局限于简单的加减法。所以，我们上面讲的这种使用 CAS 原语的方法并不是没有意义的。

另外，你需要知道的是，这种使用 CAS 原语反复重试赋值的方法，它是比较耗费 CPU 资源的，因为在 for 循环中，如果赋值不成功，是会立即进入下一次循环没有等待的。如果线程之间的碰撞非常频繁，经常性的反复重试，这个重试的线程会占用大量的 CPU 时间，随之系统的整体性能就会下降。

缓解这个问题的一个方法是使用 Yield()，大部分编程语言都支持 Yield() 这个系统调用，Yield() 的作用是，告诉操作系统，让出当前线程占用的 CPU 给其他线程使用。每次循环结束前调用一下 Yield() 方法，可以在一定程度上减少 CPU 的使用率，缓解这个问题。你也可以在每次循环结束之后，Sleep() 一小段时间，但是这样做的代价是，性能会严重下降。

所以，这种方法它只适合于线程之间碰撞不太频繁，也就是说绝大部分情况下，执行 CAS 原语不需要重试这样的场景。

小结

这节课我们一起学习了 CAS 和 FAA 这两个原语。这些原语，是由 CPU 提供的原子操作，在并发环境中，单独使用这些原语不用担心数据安全问题。在特定的场景中，CAS 原语可以替代锁，在保证安全性的同时，提供比锁更好的性能。


接下来，我们用转账服务这个例子，分别演示了 CAS 和 FAA 这两个原语是如何替代锁来使用的。对于类似：“先读取数据，做计算，然后再更新数据”这样的业务逻辑，可以使用 CAS 原语 + 反复重试的方式来保证数据安全，前提是，线程之间的碰撞不能太频繁，否则太多重试会消耗大量的 CPU 资源，反而得不偿失。

思考题

这节课的课后作业，依然需要你去动手来写代码。你需要把我们这节课中的讲到的账户服务这个例子，用你熟悉的语言，用锁、CAS 和 FAA 这三种方法，都完整地实现一遍。每种实现方法都要求是完整的，可以执行的程序。

因为，对于并发和数据安全这块儿，你不仅要明白原理，熟悉相关的 API，会正确地使用，是非常重要的。在这部分写出的 Bug，都比较诡异，不好重现，而且很难调试。你会发现，你的数据一会儿是对的，一会儿又错了。或者在你开发的电脑上都正确，部署到服务器上又错了等等。所以，熟练掌握，一次性写出正确的代码，这样会帮你省出很多找 Bug 的时间。

验证作业是否正确的方法是，你反复多次执行你的程序，应该每次打印的结果都是：

 复制代码

```
1 balance = 10000
```

欢迎你把代码上传到 GitHub 上，然后在评论区给出访问链接。如果你有任何问题，也可以在评论区留言与我交流。

感谢阅读，如果你觉得这篇文章对你有一些启发，也欢迎把它分享给你的朋友。

消息队列高手课

从源码角度全面解析 MQ 的设计与实现

李玥

京东零售技术架构部资深架构师



新版升级：点击「 请朋友读」，20位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 17 | 如何正确使用锁保护共享数据，协调异步线程？

下一篇 19 | 数据压缩：时间换空间的游戏

精选留言 (20)

写留言






微微一笑

2019-09-03

老师好，实现了下CAS,代码连接：<https://github.com/shenyachen/JKSJ/blob/master/study/src/main/java/com/jksj/study/casAndFaa/CASThread.java>。

对于FAA，通过查找资料，jdk1.8在调用sun.misc.Unsafe#getAndAddInt方法时，会根据系统底层是否支持FAA，来决定是使用FAA还是CAS。

展开

作者回复:   



4



一步

2019-09-03

NodeJS中，没有发现有关操作CpU原语CAS或者FAA的实现的

作者回复: 可以试试这个: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Atomics

1

3



QQ怪

2019-09-03

MutxLock: <https://github.com/xqq1994/algorithm/blob/master/src/main/java/com/test/concurrency/MutxLock.java>

CAS、FFA:

<https://github.com/xqq1994/algorithm/blob/master/src/main/java/com/test/concurrency/CAS.java...>

展开

作者回复: 1 1 1

1

1



姜戈

2019-09-03

JAVA中的FAA和CAS: FAA就是用CAS实现的。

```
public final int getAndAddInt(Object var1, long var2, int var4) {  
    int var5;  
    do {...
```

展开

1

1



明日

2019-09-03

Java实现: <https://gist.github.com/imgaoxin/a2b09715af99b993e30b44963cebc530>

作者回复: transfer2要放在循环中，否则有可能转账失败。

另外，transfer1中，虽然一个简单的加法不会引起任何异常，但总是把unlock放到finally中是一个好习惯。



leslie

2019-09-03

打卡：老师一步步剥离一层层拨开实质-又涨知识了，期待老师的下节课。



张三

2019-09-03

复习了一下Java中的原子类，对应到go里边的CAS实现中的for循环是自旋，还有就是要注意ABA问题吧。

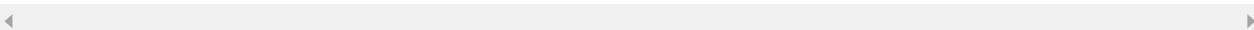


张三

2019-09-03

Java里边有支持FAA这种CPU指令的实现吗？以前没听说

作者回复: 在java中，可以看一下java.util.concurrent.atomic.AtomicLong#getAndAdd



王莹

2019-09-09

代码工程

<https://github.com/dlutsniper/wy-ja-lock/tree/master/src/main/java/wy/ja/lock/demo>

试验耗时的环节，深刻体会JIT的强大，执行次数越多，耗时均值越低
JIT吗？执行越多速度越快？ ...

展开 ∨



青舟

2019-09-04

<https://github.com/qingzhou413/geektime-mq.git>

做了1000万次加法，

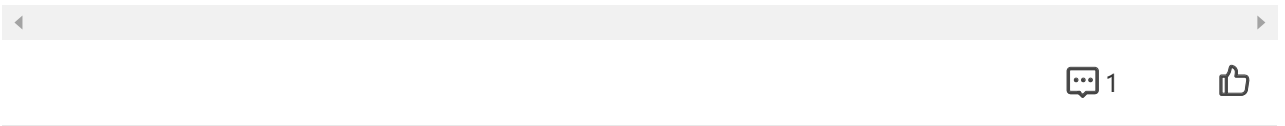
Lock: 380ms

CAS: 200ms

FAA: 280ms

展开 ∨

作者回复: java里面有AtomicInteger等等封装好的类, 包含了CAS和FAA等原子操作, 可以直接使用的。



Randy Liu

2019-09-04

用户硬件同步原语来代替锁的效果, 确实是一个好思路



许童童

2019-09-03

老师讲得很好, 对于我这种基础薄弱的, 长见识了, 感谢老师。



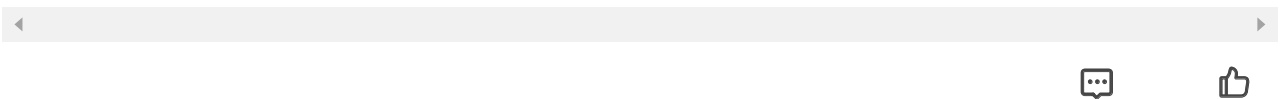
潘政宇

2019-09-03

go的语法太奇特了, 代码中的done作用是什么啊

展开 ▾

作者回复: 可以参考一下: <https://blog.golang.org/pipelines>



linqw

2019-09-03

最近出差都落下了好些, 找个空闲时间把这些实现下

展开 ▾



刘天鹏

2019-09-03

对于 “CAS 和 FAA 在各种编程语言中, 都有相应的实现”

我想起之前在操作数据的时候, 为了并发的修改用户数据(Money), 当时的做法是

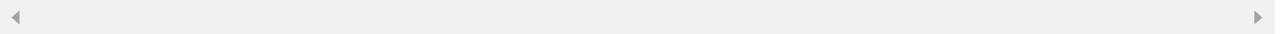
先Select出用户的Money

再对Money做操作

最后在Update的时候多附加一个条件 (Money=OldMoney)

展开 ▾

作者回复: 其实很多实现方法和技巧都是相通的



💬 1



刘天鹏

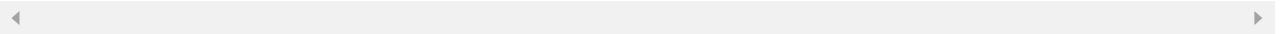
2019-09-03

<https://gist.github.com/liutianpeng/6f72bca647be41705d68736a79246c2f>

用Golang实现的版本,其实老师都已经实现了, Test和Benchmark都做了, 果然操作时间
FAA < CAS < Mutex

另外还有一个问题 atomic.LoadXXX 这组函数是什么作用 (或者说为了解决什么问题...
展开 ▾

作者回复: 就是确保这个赋值也是一个原子操作。



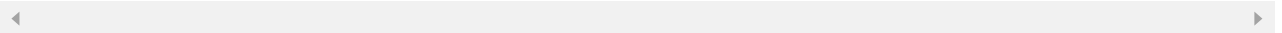
ponymm

2019-09-03

“CAS 和 FAA 在各种编程语言中, 都有相应的实现, 可以来直接使用, 无论你是使用哪种编程语言, 它底层使用的系统调用是一样的, 效果也是一样的。” 李老师这句话有点小问题: car,faa并不是通过系统调用实现的, 系统调用的开销不小, cas本来就是为了解决性能, 不会走系统调用。事实上是在用户态直接使用汇编指令就可以实现

展开 ▾

作者回复: 感谢你指出错误, 我已经联系编辑在文稿中改正了。



2019-09-03

yiald感觉还是不是等一定周期使用是不是更好, 如果系统线程多, 线程的频繁切换带来的开销也不小, go的协程会好些

展开 ▾



白小白

2019-09-03

打卡打卡！晚上回家做作业！

展开 ▾



书策稠浊

2019-09-03

看完，先抢个沙发，晚点上链接。

展开 ▾

