第10讲 | 如何保证集合是线程安全的? ConcurrentHashMap如何实现高效地线程安全?

2018-05-26 杨晓峰





第10讲 | 如何保证集合是线程安全的? ConcurrentHashMap如何实现高效地线程安全? 杨晓峰

00:00 / 10:46

我在之前两讲介绍了Java集合框架的典型容器类,它们绝大部分都不是线程安全的,仅有的线程安全实现,比如Vector、Stack,在性能方面也远不尽如人意。幸好Java语言提供了并发包(java.util.concurrent),为高度并发需求提供了更加全面的工具支持。

今天我要问你的问题是,如何保证容器是线程安全的? ConcurrentHashMap如何实现高效地线程安全?

典型回答

Java提供了不同层面的线程安全支持。在传统集合框架内部,除了Hashtable等同步容器,还提供了所谓的同步包装器(Synchronized Wrapper),我们可以调用Collections工具类提供的包装方法,来获取一个同步的包装容器(如Collections.synchronizedMap),但是它们都是利用非常粗粒度的同步方式,在高并发情况下,性能比较低下。

另外,更加普遍的选择是利用并发包提供的线程安全容器类,它提供了:

- 各种并发容器,比如ConcurrentHashMap、CopyOnWriteArrayList。
- 各种线程安全队列(Queue/Deque),如ArrayBlockingQueue、SynchronousQueue。
- 各种有序容器的线程安全版本等。

具体保证线程安全的方式,包括有从简单的synchronize方式,到基于更加精细化的,比如基于分离锁实现的ConcurrentHashMap等并发实现等。具体选择要看开发的场景需求,总体来说,并发包内提供的容器通用场景,远优于早期的简单同步实现。

考点分析

谈到线程安全和并发,可以说是Java面试中必考的考点,我上面给出的回答是一个相对宽泛的总结,而且ConcurrentHashMap等并发容器实现也在不断演进,不能一概而论。

如果要深入思考并回答这个问题及其扩展方面,至少需要:

- 理解基本的线程安全工具。
- 理解传统集合框架并发编程中Map存在的问题,清楚简单同步方式的不足。
- 梳理并发包内,尤其是ConcurrentHashMap采取了哪些方法来提高并发表现。
- 最好能够掌握ConcurrentHashMap自身的演进,目前的很多分析资料还是基于其早期版本。

今天我主要是延续专栏之前两讲的内容,重点解读经常被同时考察的HashMap和ConcurrentHashMap。今天这一讲并不是对并发方面的全面梳理,毕竟这也不是专栏一讲可以介绍完整的,算是个开胃菜吧,类似CAS等更加底层的机制,后面会在Java进阶模块中的并发主题有更加系统的介绍。

知识扩展

1.为什么需要ConcurrentHashMap?

Hashtable本身比较低效,因为它的实现基本就是将put、get、size等各种方法加上"synchronized"。简单来说,这就导致了所有并发操作都要竞争同一把锁,一个线程在进行同步操作时,其他线程只能等待,大大降低了并发操作的效率。

前面已经提过HashMap不是线程安全的,并发情况会导致类似CPU占用100%等一些问题,那么能不能利用Collections提供的同步包装器来解决问题呢?

看看下面的代码片段,我们发现同步包装器只是利用输入Map构造了另一个同步版本,所有操作虽然不再声明成为synchronized方法,但是还是利用了"this"作为互斥的mutex,没有真正意义上的改进!

极客时间

所以, Hashtable或者同步包装版本, 都只是适合在非高度并发的场景下。

2.ConcurrentHashMap分析

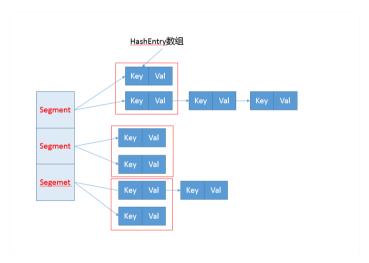
我们再来看看ConcurrentHashMap是如何设计实现的,为什么它能大大提高并发效率。

首先,我这里强调,ConcurrentHashMap的设计实现其实一直在演化,比如在Java 8中就发生了非常大的变化(Java 7其实也有不少更新),所以,我这里将比较分析结构、实现机制等方面,对比不同版本的主要区别。

早期ConcurrentHashMap, 其实现是基于:

- 分离锁,也就是将内部进行分段(Segment),里面则是HashEntry的数组,和HashMap类似,哈希相同的条目也是以链表形式存放。
- HashEntry内部使用volatile的value字段来保证可见性,也利用了不可变对象的机制以改进利用Unsafe提供的底层能力,比如volatile access,去直接完成部分操作,以最优化性能,毕竟Unsafe中的很多操作都是JVM intrinsic优化过的。

你可以参考下面这个早期ConcurrentHashMap内部结构的示意图,其核心是利用分段设计,在进行并发操作的时候,只需要锁定相应段,这样就有效避免了类似Hashtable整体同步的问题,大大提高了性能。



在构造的时候,Segment的数量由所谓的concurrentcyLevel决定,默认是16,也可以在相应构造函数直接指定。注意,Java需要它是2的幂数值,如果输入是类似15这种非幂值,会被自动调整到16之类2的幂数值。

具体情况,我们一起看看一些Map基本操作的遍码,这是JDK 7比较新的get代码。针对具体的优化部分,为方便理解,我直接注释在代码段里,get操作需要保证的是可见性,所以并没有什么同步逻辑。

```
public V get(Object key) {
    SegmentK,V > 5; // manually integrate access methods to reduce overhead
    HashEntryK,V>[] tab;
    int h = hash(key.hashCode());
    //#JEUGHENHERINGYENE
    long u = (((h >>> segmentMask) << SSHIFT) + SBASE;
    // U.SegmentJmED. 出于定位
    // JPJRAsafe周期进行Volatile access
    if ((s = (SegmentK,V)UNSAFE.getObjectVolatile(segments, u)) != null &&
        (tab = s.table) != null) {
        //###
    }
    return null;
}
```

而对于put操作,首先是通过二次哈希避免哈希冲突,然后以Unsafe调用方式,直接获取相应的Segment,然后进行线程安全的put操作:

ATT often AART

其核心逻辑实现在下面的内部方法中:

```
final V put(K key, int hash, V value, boolean onlyIfAbsent) {
         // scanAndLockForPut会去查找是否有key相同Node
         // 无论如何,确保获取锁
          HashEntry<K,V> node = tryLock() ? null :
             scanAndLockForPut(key, hash, value);
          V oldValue:
          try {
             HashEntry<K,V>[] tab = table;
             int index = (tab.length - 1) & hash;
             HashEntry<K,V> first = entryAt(tab, index);
             for (HashEntryck.V> e = fir8::) {
                if (e != null) {
                   Kk;
                    // 更新已有value...
                 else {
                   // 放置HashEntry到特定位置,如果超过阈值,进行rehash
                    // ...
         } finally {
             unlock();
          return oldValue;
```

所以,从上面的源码清晰的看出,在进行并发写操作时:

- ConcurrentHashMap会获取再入锁,以保证数据一致性,Segment本身就是基于ReentrantLock的扩展实现,所以,在并发修改期间,相应Segment是被锁定的。
- 在最初阶段,进行重复性的扫描,以确定相应key值是否已经在数组里面,进而决定是更新还是放置操作,你可以在代码里看到相应的注释。重复扫描、检测冲突是ConcurrentHashMap的常见技巧。
- 我在专栏上一讲介绍HashMap时,提到了可能发生的扩容问题,在ConcurrentHashMap中同样存在。不过有一个明显区别,就是它进行的不是整体的扩容,而是单独 对Segment进行扩容,细节就不介绍了。

另外一个Map的size方法同样需要关注,它的实现涉及分离锁的一个副作用。

试想,如果不进行同步,简单的计算所有Segment的总值,可能会因为并发put,导致结果不准确,但是直接锁定所有Segment进行计算,就会变得非常昂贵。其实,分离锁也限制了Map的初始化等操作。

所以,ConcurrentHashMap的实现是通过重试机制(RETRIES_BEFORE_LOCK,指定重试次数2),来试图获得可靠值。如果没有监控到发生变化(通过对比Segment.modCount),就直接返回,否则获取锁进行操作。

下面我来对比一下,在Java 8和之后的版本中,ConcurrentHashMap发生了哪些变化呢?

- 总体结构上,它的内部存储变得和我在专栏上一讲介绍的HashMap结构非常相似,同样是大的桶(bucket)数组,然后内部也是一个个所谓的链表结构(bin),同步的粒度要更细致一些。
- 其内部仍然有Segment定义,但仅仅是为了保证序列化时的兼容性而已,不再有任何结构上的用处。
- 因为不再使用Segment,初始化操作大大简化,修改为lazy-load形式,这样可以有效避免初始开销,解决了老版本很多人抱怨的这一点。
- 数据存储利用volatile来保证可见性。
- 使用CAS等操作,在特定场景进行无锁并发操作。
- 使用Unsafe、LongAdder之类底层手段,进行极端情况的优化。

先看看现在的数据存储内部实现,我们可以发现Key是final的,因为在生命周期中,一个条目的Key发生变化是不可能的;与此同时val,则声明为volatile,以保证可见性。

```
Satic class NodecK,V> implements Map.EntrycK,V> {
    fmal int hash;
    fmal K key;
    volatile V val;
    volatile NodecK,V> next;
```

```
//-
}
```

我这里就不再介绍get方法和构造函数了,相对比较简单,直接看并发的put是如何实现的。

```
final V putVal(K key, V value, boolean onlyIfAbsent) { if (key == null || value == null) throw new NullPointerException();
  int hash = spread(key.hashCode());
  int binCount = 0:
  for (Node<K,V>[] tab = table;;) {
      Node<K,V> f; int n, i, fh; K fk; V fv;
     if (tab == null || (n = tab.length) == 0)
         tab = initTable();
      else if ((f = tabAt(tab, i = (n - 1) & hash)) == null) {
         // 利用CAS去进行无锁线程安全操作,如果bin是空的
         if (casTabAt(tab, i, null, new Node<K,V>(hash, key, value)))
            break:
      else if ((fh = f.hash) == MOVED)
         tab = helpTransfer(tab, f);
      else if (onlyIfAbsent // 不加锁, 进行检查
            && fh == hash
            && ((fk = f.key) == key || (fk != null && key.equals(fk)))
            && (fv = f.val) != null)
         return fv;
      else {
         V oldVal = null;
          synchronized (f) {
               // 细粒度的同步修改操作...
            }
         // Bin超过阈值,进行树化
          if (binCount != 0) {
           if (binCount >= TREEIFY_THRESHOLD)
               treeifyBin(tab, i);
            if (oldVal != null)
               return oldVal;
   addCount(1L, binCount);
   return null;
```

初始化操作实现在initTable 里面,这是一个典型的CAS使用场景,利用volatile的sizeCtl作为互斥手段:如果发现竞争性的初始化,就spin在那里,等待条件恢复;否则利用CAS设置排他标志。如果成功则进行初始化;否则重试。

请参考下面代码:

```
private final Node<K,V>[] initTable() {
  Node<K,V>[] tab; int sc;
  while ((tab = table) == null || tab.length == 0) {
      // 如果发现中突,进行spin等待
      if ((sc = sizeCtl) < 0)
        Thread.vield():
      // CAS成功返回true,则进入真正的初始化逻辑
      else if (U.compareAndSetInt(this, SIZECTL, sc, -1)) {
         try {
            if ((tab = table) == null || tab.length == 0) {
               int n = (sc > 0) ? sc : DEFAULT_CAPACITY;
               @SuppressWarnings("unchecked")
               Node<K,V>[] nt = (Node<K,V>[])new Node<?,?>[n];
                table = tab = nt;
                sc = n - (n >>> 2);
        } finally {
            sizeCtl = sc;
         }
  }
  return tab:
```

当bin为空时,同样是没有必要锁定,也是以CAS操作去放置。

你有沒有注意到,在同步逻辑上,它使用的是synchronized,而不是通常建议的ReentrantLock之类,这是为什么呢?现代JDK中,synchronized已经被不断优化,可以不再过分

担心性能差异,另外,相比于ReentrantLock,它可以减少内存消耗,这是个非常大的优势。

与此同时,更多细节实现通过使用Unsafe进行了优化,例如tabAt就是直接利用getObjectAcquire,避免间接调用的开销。

```
datic final <K,V> NodecK,V> tabAt(NodecK,V>[] tab, int i) {
    return (NodecK,V>)U.getObjectAcquire(tab, ((long)1 << ASHIFT) + ABASE);
}</pre>
```

再看看,现在是如何实现size操作的。阅读代码你会发现,真正的逻辑是在sumCount方法中, 那么sumCount做了什么呢?

我们发现,虽然思路仍然和以前类似,都是分而治之的进行计数,然后求和处理,但实现却基于一个奇怪的CounterCell。 难道它的数值,就更加准确吗?数据一致性是怎么保证的?

```
flatic final class CounterCell {
  volatile long value;
  CounterCell(long x) { value = x; }
}
```

其实,对于CounterCell的操作,是基于java.util.concurrent.atomic.LongAdder进行的,是一种JVM利用空间换取更高效率的方法,利用了<u>Striped64</u>内部的复杂逻辑。这个东西非常小众,大多数情况下,建议还是使用AtomicLong,足以满足绝大部分应用的性能需求。

今天我从线程安全问题开始,概念性的总结了基本容器工具,分析了早期同步容器的问题,进而分析了Java 7和Java 8中ConcurrentHashMap是如何设计实现的,希望ConcurrentHashMap的并发技巧对你在日常开发可以有所帮助。

一课一结

关于今天我们讨论的题目你做到心中有数了吗?留一个道思考题给你,在产品代码中,有没有典型的场景需要使用类似ConcurrentHashMap这样的并发容器呢?

请你在留言区写写你对这个问题的思考,我会选出经过认真思考的留言,送给你一份学习鼓励金,欢迎你与我一起讨论。

你的朋友是不是也在准备面试呢?你可以"请朋友读",把今天的题目分享给好友,或许你能帮到他。



徐金铎

需要注意的一点是,1.8以后的锁的颗粒度,是加在链表头上的,这个是个思路上的突破。 作者回复

是的 雷霹雳的爸爸

2018-05-26

2018-05-26

2018-05-28

今天这个纯粹知识盲点, 纯赞, 源码也得不停看

极密时间

最近用ConcurrentHashMap的场景是,由于系统是一个公共服务,全程异步处理。最后一环节需要http rest主动响应接入系统,于是为了定制化需求,利用netty写了一版异步http clinet。 甘在绥存tcp链接时田到了 具在接行CP链接时用到了 各球的一个间板改起了各條物和偏向物。 自接物介人理解的是在3的一种应用方式。并发包中的原子类是典型的应用。 偏向物介人理解的是基础物的优化。在RentrantLock中用于实现已获取完锁的的线程重入问题。 不知道理解的是否有误差。欢迎指正探讨。谢谢 正确,互相交流 偏向锁,侧重是低竞争场器的优化。 夫掉可能不必要的同步 期待unsafe和cas的文音 RESE 2018-07-04 通过分段加锁segment,一个hashmap里有若干个segment,每个segment里有若干个桶,桶里存放K-V形式的链表,put数据时通过key哈希得到该元素要添加到的segment,然后对segment进行加锁。然后在哈希,计量得到给元素要添加到的桶。然后偏压桶中的链表。替每些新维节点到桶中 分段计算两次,两次结果相同则返回,否则对所以段加锁重新计算 1.8 put CAS 加锁 DBC AS MAN Segment加锁,segment数量与桶数量一致; 首先判断容器是否为空,为空则进行初始化利用volatile的sizectl作为互斥手段,如果发现竞争性的初始化,就暂停在那里,等待条件恢复,否则利用CAS设置排他标志 (U.compareAndSwapInt(link)、SIZECTL、sc. -1));否则重试 对key hash计算得到该key存放的桶位置,判断该桶是否为空,为空则利用CAS设置新节点 否则使用synchronize加锁,遍历桶中数据,替换或新增加点到桶中 最后判断是否需要转为红黑树,转换之前判断是否需要扩容 利用LongAdd累加计算 coder± 您说的synchronized被改进很多很多了,那么在我们平常使用中,就用这个synchronized完成一些同步操作是不是OK? ◆◆ 作者同复 2018-05-28 通常是的, 前提是JDK版本需要新一点 mongo 请教老师: putVal方法的第二个If分支,为什么要用tabAt? 我的认识里直接数组下标寻址tab[i=(n-1) & hash]也是一个原子操作,不是吗? tabAt里面的getObjectVolatle () 方法跟直接用数组下标ab[i=(n-1) & hash]寻址有什么区别? 作者同复 2018-05-28 这个有volatile load语义 hansc 从1.5有并发包,到1.6对synchronized的改进,到1.7的并发map的分段锁,再到1.8的cas+synchronized。 作者回复 2018-05-28 嗯嗯,有的历史我也未必知道 専箔 这期内容太难,分寸不好把握 看8的concurenthashmap源码感觉挺困难,网上的博文帮助也不大,尤其是扩容这部分(似乎文章中没提) 求问杨大有没有什么窍门,或者有什么启发性的paper或文章 可以泛化成,长期对lock free实现多个状态修改的问题比较困惑,希望得到启发 本文尽量梳理了相对比较容易理解的部分;扩容细节我觉得是个加分项,不是每个人都会在乎那么深入;窍门,可以考虑画图辅助理解,我是比较笨的类型,除了死磕,不会太多窍门…… 2018-06-08 这里有个地方想跟老师交流一下想法,从文中"所以,ConcurrentHashMap 的实现是通过重试机制(RETRIES_BEFORE_LOCK,指定重试次数 2),来试图获得可靠值。如果没有监控到发生变化(通过对比 Segment modCount),就直接返回。否则萃取锁进行操作。"可以看出,在高井发的情况下。"size()"方法只是返回"近似信",而我的问题是:既然只是一个近似值,为啥要用这种"重试,分段锁"的复杂做法去计算这个值?直接在不加锁的情况下返回segment 的size 岂不是更简单?我能理解d K开发者想尽一切努力在高性能地返回最精确的数值,但这个"精确"度无法置化啊,对于调用方来说。这个值依然是不可靠的啊,所以,在我看来。这种做法校监很小(可能是我也比较赚的)。或者有些设计上的要点我没有领悟出来,希望老师指点一下。 作者回复 2018-06-08 这个是在代价可接受情况下,尽量准确,就像含金量90%和99.9%,99.999%,还是有区别的,虽然不是百分百 老师你好,tabAt里面的getObjectVolatle()方法跟直接用数组下标tab[i=(n-1) & hash]寻址有什么区别,这个我也不懂,volitile不是已经保证内存可见性吗? 作者回复 2018-05-31 volatile保证的是数组,不是数组元素 Hesher 并发包用的很少,这一节内容的前置知识比较多,对于使用经验少的人来说貌似是有点难了。问题很好,正好可以见识一下各种使用场景,不过留言大部分是针对内容的难点提问,而真正回答 问题的还没有出现。 2018-05-28

后面并发部分会详细分析

	2018-05-
纷师嫉烦讲讲自旋锁,偏向锁的特点和区别吧,一直不太清楚。 作者回复	
	2018-05-
好,后面有章节	
	2018-05-
哪在课程里讲到同步包装类比较低效,不太适合离并发的场景,那想请教一下老师,在list接口的实现类中。在高井发的场景下,选择哪种实现类比较 它起装类又很低效,CopyonwriteArrayList又是以快赔的形式来实现的,在频繁写入数据的时候,其实也很低效,那这个类型该怎么选择比较好? 作者回复	
目前并发list好像就那一个,我觉得不必拘泥于list,不还有queue之类,看场景需要的真是list吗	2018-05-
nongo	
情教老师:putVal方法中,什么情况下会进入else if ((fh=f.hash) == MOVED)分支?是进行扩容的时候吗?nextTable是做什么用的? 作者回复	2018-05-
我理解是的,判断是个ForwardingNode,resize正在进行; nexttable是扩容时的临时过渡	2018-05-
ansc	
sgi环境中多线程启动bundle,扫描注解配置缓存起来。	2018-05-
对于我这种菜鸟来说,应该来一期讲讲volatile��	2018-07-
Inswer	
Insafe?	2018-07-
hawn	
7/4 D ***	2018-07-
图师,什么只有bin为空的时候才使用cas,其他地方用synchronized 呢? azindes	
	2018-06
8师的内容讲的丰富 深入浅出 希望提高一下朗读人的要求吧 每节课都感觉有读错的 英文读不准就不说了 互斥读成互拆听的实在是别扭 作者回复	
哈,抱歉,我反馈一下,主摘也辛苦,不一定是职业码农	2018-06
响,78条,54.X.以 下,工图也十百,小 上定年工的4	
ENää的时候,Segment 的数量由所谓的 concurrentcyLevel 决定,默认是 16! 并发数不是越多越好吗? 作者回复	2018-06-
	2018-06-
都是有代价的,取个折衷 を できました こうしゅう アンドラ アンドラ アンドラ アンドラ アンドラ アンドラ アンドラ アンドラ	
	2018-06
歼 几个词感觉第一次见,大哭 ◆◆ 作者回复	2018-06
嗯…哪几个?回头根据反馈补充下	
在天上飞	2018-06
奶能解释下可重入锁吗? 作者回复	2018-06
马上就有章节介绍了	
liaozhe	2018-05
多额,看到ConcurrentHashMap中有定义N CPU,想问问跟CPU什么关系? 作者回复	
那个是Runtime.availableProcessors	2018-05
Seorge	
A VALUE OF THE STATE OF THE STA	2018-05
文中说的UNSAFE是什么意思? 作者回复	2018-05
作者回复 JDK内部的一个基础类sun.misc.Unsafe,如果公共api能解决的问题,不建议使用它,如果是做底层开发可能会用到	
作者回复	2018-05

作者回复

基本正确,cas只用在部分场景; 事后看容易啊,说比做容易,��

极客cq

2018-05-29

不从实现上只从圈形结构上看,ConcurrentHashMap和HashMap一样,不过是将buckets换成了segment,然后加锁方式从整map,下沉到segment(buckets)上,这样简单理解正确

作者回复

有一定道理, 但seg和buck范围不完全对等

Kyle

2018-05-29 2018-05-28

之前用JavaFX做一个客户端IM工具的时候,我将拉来的未被读取的用户聊天信息用ConcurrentHashMap存储(同时异步存储到Sqlite),Key存放用户id,Value放未读取的聊天消息列表。 因为我考虑到存消早和滤消息是由两个线程并发处理的,这两个线程共同操作一个ConcurrentHashMap。可能是挖没处理好,最后直到我离职了还有消息重复。乱呼的问题。请问我这种应用 场票有什么问题吗?

Ethan

请问老师以后会不会有讲线程中断的不同处理呢?这一块一直对我来讲都比较抽象而且不好测试

作者回复

2018-05-28

2018-05-28

2018-05-27

2018-05-28

2018-05-26

2018-05-28

2018-05-28

会有,我也记录下

lorancechen

2018-05-28 rpc调用内部,客户端会记录每次请求的unique id,用于匹配返回的数据应该响应哪个请求。高并发情况下,应该使用concurrenthashmap做这种id到回调的记录。

如果多线程中已经在上层代码使用了读写锁进行访问控制,底层集合是否就可以使用HashMap,而没有必要使用线程安全的容器了?

自己做同步当然也是个选择

可不可以用AI让JVM更智能?

作者回复

2018-05-28

需要更具体一些,没有一劳永逸的方法,比如,有用机器学习调优jvm运行参数的; jvm比较运行场景太多,SE既可以跑在嵌入式,也可以到高性能服务器,缩小范围才好进行

灰飞灰猪不会灰飞.烟灭

2018-05-26

jdk7和8的区别感觉就是加锁不一样了,其他的没看懂。 老师,synchronized和lock是不是都是在类字节码种携带自旋锁和偏向锁啊? 他两底层区别是啥呢?我知道lock里面维护了一个双向链表

作者回复

我列出来的区别不止吧; 存在锁升降级的问题,后面有章节介绍

2018-05-26

有些本地缓存就是基于它实现的。

CUZZ.

2018-05-26

有点不懂。。

作者回复

请问具体哪个方面? 并发工具后面会有专门分析

极等时间		

极等时间		