

## 多线程

---

### 多线程基础

#### 聊聊线程和进程 (高)

进程是程序的一次执行过程，是系统运行程序的基本单位，是操作系统分配资源的最小单位。一个进程会有一个主线程

- 在 Java 中，当我们启动 main 函数时其实就是一个 JVM 的进程
- 而 main 函数所在的线程就是这个进程中的一个线程，也称主线程

线程是一个比进程更小的执行单位，一个进程在其执行的过程中可以产生多个线程

- 线程共享进程的**堆和方法区**资源
- 每个线程有自己的**程序计数器、虚拟机栈和本地方法栈**
- 线程切换的代码比进程小得多，线程也被称为轻量级进程

#### 进程通信 (中)

进程是分配系统资源的单位（包括内存地址空间），因此**各进程拥有的内存地址空间相互独立**。一个进程不能访问另一个进程的内存地址空间，所以就需要进程通信。

1. **共享内存(Shared memory)** : 在内存中划分出一块共享存储区,使得多个进程可以访问同一块内存空间, 不同进程可以及时看到对方进程中对共享内存中数据的更新。 **各个进程要互斥地访问共享内存**, 可以用互斥锁和信号量等实现互斥操作。这个的典型使用就是剪贴板.
2. **管道通信 (字符流)**: 操作系统建立管道连接两个软件, 再传输字符流. 这个管道是一种特殊的pipe文件, 管道大小固定, 采用半双工方式通信, **各进程互斥访问管道**. 数据以字符流的形式写入管道, 当管道写满时, 写进程的write()系统调用将被阻塞, 等待另一个进程将数据取走。当读进程将数据全部取走后, 管道变空, 此时读进程的read()系统调用将被阻塞, 第一个进程就可以开始写了。  
**(写满了才能读, 读完了才能写)**
3. **消息传递**:
  1. 直接通信方式: 每个进程会有一个消息缓冲队列, 其他进程想要通信, 就**要把发送的数据封装为一个消息, 然后放入目标进程的消息缓冲队列中**, 目标进程就能读取消息缓冲队列, 获得数据
  2. 间接通信方式: **进程发送消息时, 会发送到一个中间实体中, 称为信箱**. 消息的消息头中存放了发送进程ID和接收进程ID, 所以目标进程想知道消息, 直接从信箱取即可.
4. **socket套接字**: 更为通用的进程间通信机制, 可用于不同机器之间的进程间通信

## Java创建线程有几种方式 (高)

常规回答

- 继承Thread类, 重写run方法
- 实现 Runnable 接口
- 实现Callable接口

- 通过线程池创建

以上回答并未触及本质, 接下来聊聊本质

- Java创建线程有且只有一种方式, 就是继承Thread类重写run方法, 调用Thread类的start方法
- 实现Runnable和Callable的还是要把实现类对象传入Thread类中, 调用Thread类的start方法. 所以本质创建线程还是依靠Thread类中的start方法.
- 实现Runnable和Callable实际上是创建了一个线程任务. 然后调用Thread类中的start方法, start方法调用start0方法, start0是一个本地方法, 由C/C++编写, 用来进行系统调用创建新线程, 然后用新线程来执行线程任务.
- 这也说明了为何实现Runnable或者Callable后调用run方法不能开启新线程, 是因为开启新线程本质上需要调用Thread的start0方法.

## 什么是死锁 (中)

- 多线程争抢资源, 只有得到资源才能继续执行.
- 但是每个线程都持有一部分资源, 都等待对方释放资源.
- 多个线程互相僵持, 导致都无法运行

## Thread 类中的start() 和 run() 方法有什么区别 (中)

- start()方法被用来启动新创建的线程, 而且start()内部调用了run()方法
- 如果直接调用run(), 只会是在原来的线程中调用, 不会启动新线程, start()方法才会启动新线程

## 线程间通信方式有哪些？（中）

- Object类的wait()、notify()和notifyAll()方法. (必须基于synchronized.)
  - wait(): 使当前线程进入等待状态，直到其他线程调用该对象的notify()或notifyAll()方法
  - notify(): 唤醒在此对象监视器上等待的单个线程
  - notifyAll(): 唤醒在此对象监视器上等待的所有线程
- Lock和Condition接口. (Lock接口提供了比synchronized更灵活的锁机制，Condition接口则配合Lock实现线程间的等待/通知机制)
  - await(): 使当前线程进入等待状态，直到被其他线程唤醒
  - signal(): 唤醒一个在该Condition对象的条件队列等待的线程
  - signalAll(): 唤醒所有在该Condition对象的条件队列等待的线程
- 通过volatile关键字让多个线程共享变量
  - volatile关键字用于保证变量的可见性
  - 当一个变量被声明为volatile时，它会保证对该变量的写操作会立即刷新到主内存中，而读操作会从主内存中读取最新的值
- CountDownLatch, 可以让一个或多个线程等待其他线程完成操作
- CyclicBarrier, 可以让一组线程相互等待，直到所有线程都到达某个公共屏障点
- Semaphore, 信号量机制，可以控制同时访问特定资源的线程数量

# 线程池

## ThreadPoolExecutor构造函数(高)

```
1 ThreadPoolExecutor(int corePoolSize,
2                     int maximumPoolSize,
3                     Long keepAliveTime,
4                     TimeUnit unit,
5                     BlockingQueue<Runnable>
6                     workQueue,
7                     ThreadFactory threadFactory,
8                     RejectedExecutionHandler
9                     handler)
```

- `corePoolSize` 指定线程池的核心线程数(必须大于0), 核心线程就是一直在线程池里的长久存活的线程
- `maximumPoolSize` 指定线程池中的最大线程数(最大线程数>核心线程数), 临时线程用完销毁
- `keepAliveTime` 指定临时线程空闲时的存活时间
- `unit` 指定临时线程存活时间的单位
- `workQueue` 指定任务队列 (不能为null). 当提交的任务数超过核心线程数后, 再提交的任务就存放在工作队列
- `threadFactory` 指定哪个线程工厂创建线程 (不能为null)
- `handler` 指定线程忙, 任务队列满的时候, 新任务来了怎么办 (不能为null)

临时线程什么时候创建啊?

- 新任务提交时发现核心线程都在忙, 任务队列也满了, 并且还可以创建临时线程, 此时才会创建临时线程。

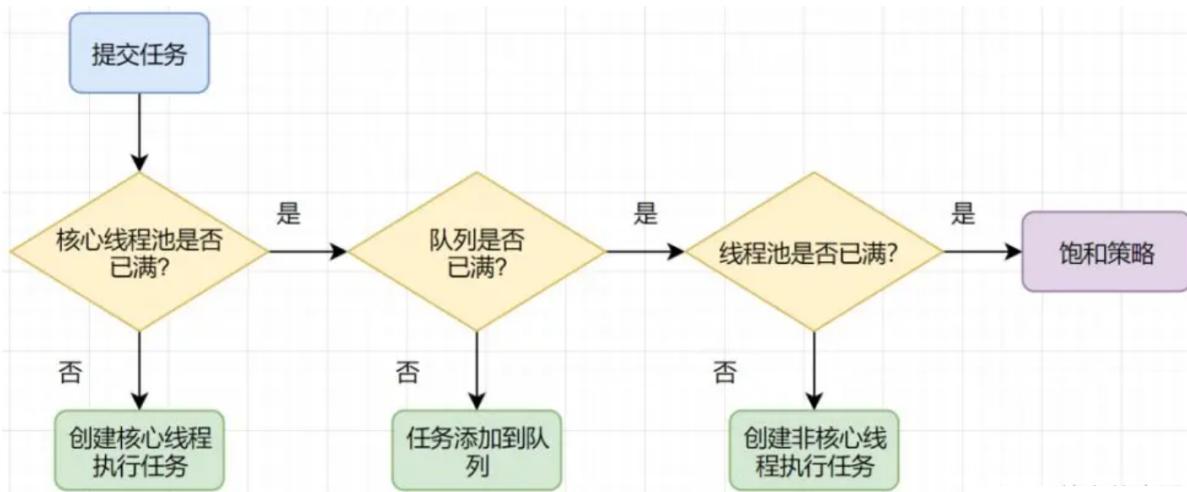
- 这样是最大限度避免创建线程

什么时候会开始拒绝任务？

- 核心线程和临时线程都在忙，任务队列也满了，新的任务过来的时候才会开始任务拒绝。

## 提交一个新任务到线程池时，具体的执行流程 / 线程池的工作流程 (高)

- 当我们提交任务，线程池会根据corePoolSize大小创建若干任务数量线程执行任务
- 当任务的数量超过corePoolSize数量，后续的任务将会进入阻塞队列阻塞排队
- 当阻塞队列也满了之后，那么将会继续创建(maximumPoolSize-corePoolSize)个数量的线程来执行任务，如果任务处理完成，maximumPoolSize-corePoolSize额外创建的线程等待keepAliveTime之后被自动销毁
- 如果达到maximumPoolSize，阻塞队列还是满的状态，那么将根据不同的拒绝策略对应处理



## 线程池线程数如何设置? (高)

- CPU密集型: 任务需要大量计算, 很少阻塞, CPU一直处于忙碌状态. CPU核数 + 1
- IO密集型: 任务需要频繁的IO操作(与磁盘, 网络交互), CPU经常等待IO完成. CPU核数 \* 2

`Runtime.getRuntime().availableProcessors();` 获取CPU最大核心数

## 线程池中submit() 和execute()方法有什么区别 (中)

- 两个方法都可以向线程池提交任务
- execute()只能提交Runnable任务, submit()既可以提交Runnable任务, 又能提交Callable任务
- submit()可以返回持有计算结果的Future对象.

## Executors中的常用线程池 (高)

`static ExecutorService newCachedThreadPool()` 线程数量随着任务增加而增加, 如果线程任务执行完毕且空闲了一段时间则会被回收掉。 (全是临时线程, 没有核心线程)

`static ExecutorService newFixedThreadPool(int nThreads)` 创建固定线程数量的线程池, 如果某个线程因为执行异常而结束, 那么线程池会补充一个新线程替代它。 (全是核心线程, 没有临时线程)

`static ExecutorService newSingleThreadExecutor()` 创建只有一个线程的线程池对象, 如果该线程出现异常而结束, 那么线程池会补充一个新线程。 (只有一个核心线程)

```
static ScheduledExecutorService  
newScheduledThreadPool(int corePoolSize)
```

创建一个线程池，可以实现在给定的延迟后运行任务，或者定期执行任务。(使用 ScheduledExecutorService类接受返回值, 调用schedule()方法来实现延迟执行, 调用scheduleAtFixedRate()方法或者 scheduleWithFixedDelay()方法实现定期执行) 既有核心线程, 又有临时线程

注: Executors的底层其实也是基于线程池的实现类 ThreadPoolExecutor创建线程池对象的。

## Executors创建线程池会存在什么问题 (高)

Executors有 newCachedThreadPool()  
newFixedThreadPool(int nThreads)  
newSingleThreadExecutor() newScheduledThreadPool(int corePoolSize) 几大方法创建线程池.

Executors用起来比较方便, 但是在大型分布式系统中直接使用会产生一些问题.

newFixedThreadPool() 和 newSingleThreadExecutor() 固定线程数量的线程池和单个线程的线程池, 线程数量不会溢出, 但是任务队列有可能溢出. 允许的任务队列长度为Integer.MAX\_VALUE, 可能堆积大量任务, 导致OOM.

newCachedThreadPool() 和 newScheduledThreadPool() 线程数量随着任务增加而增加, 线程数量有可能溢出. 线程数最大为 Integer.MAX\_VALUE, 可能会创建大量线程, 从而导致oom.

阿里巴巴开发手册中强制规定, 线程池不允许使用Executors创建, 而是通过ThreadPoolExecutor方式创建, 这样可以使编程者更加明确线程池运行规则, 避免资源耗尽.

## 线程池的拒绝策略 (中)

线程池的拒绝策略主要有四种：

1. AbortPolicy：抛出RejectedExecutionException异常。
2. CallerRunsPolicy：调用者线程自己执行任务。
3. DiscardPolicy：直接丢弃任务，不处理。
4. DiscardOldestPolicy：丢弃队列中最旧的任务，然后重新尝试提交当前任务。
5. 自定义拒绝策略，通过实现RejectedExecutionHandler接口，并重写rejectedExecution方法来实现自定义逻辑

# 锁

---

## 锁基础

### 常用锁的概念 (高)

- 乐观锁和悲观锁
  - 悲观锁和独占锁是一个意思，它假设一定会发生冲突，因此获取到锁之后会阻塞其他等待线程。
    - 这么做的好处是简单安全，但是挂起线程和恢复线程都需要转入内核态进行，这样做会带来很大的性能开销。
    - 悲观锁的代表是 synchronized。
    - 然而在真实环境中，大部分时候都不会产生冲突。悲观锁会造成很大的浪费。
  - 而乐观锁不一样，它假设不会产生冲突，先去尝试执行某项操作，失败了再进行其他处理（一般都是不断循环重试）。

- 这种锁不会阻塞其他的线程，也不涉及上下文切换，性能开销小。
- 代表实现是 CAS。
- 公平锁和非公平锁
  - 非公平锁性能更好，因为在线程切换间隙会产生短暂的延时，所以非公平锁此时就可以去抢锁，效率更高。但是如果一直有线程来抢锁，而某个线程一直没抢到锁，就有可能造成某些线程的饥饿。
  - 公平锁因为需要排队拿锁，所以效率低，但是可以保证只要排队了就能拿到锁，不会产生饥饿现象。
- 可重入锁和非可重入锁
  - 可重入锁又名递归锁，是指在同一个线程在外层方法获取锁的时候，再进入该线程的内层方法会自动获取锁（前提锁对象得是同一个对象或者class），不会因为之前已经获取过还没释放而阻塞。Java中ReentrantLock和synchronized都是可重入锁，可重入锁的一个优点是可一定程度避免死锁。
- 阻塞锁和非阻塞锁
  - 阻塞锁：没拿到锁就阻塞等待。
  - 非阻塞锁：没拿到锁直接返回。
- 自旋锁和互斥锁
  - 自旋锁：当没拿到锁时，不会释放CPU，而是一直处于活跃状态（while循环），一直循环判断锁是否被释放。循环时间一长十分消耗CPU资源。
  - 互斥锁：没拿到锁就释放CPU，进入阻塞等待状态，直到被唤醒。
- 读写锁
  - 写锁就是一种独占锁/排他锁，避免并发修改问题。写-写互斥
  - 读锁是一种共享锁，多个读线程可以共享一把锁。读-读不互斥，读-写互斥

# Java中有哪些常用的锁，在什么场景下使用？（高）

- synchronized
  - synchronized关键字可以用于方法或代码块。
  - 当一个线程进入synchronized代码块或方法时，它会获取关联对象的锁；
  - 当线程离开该代码块或方法时，锁会被释放
  - 如果其他线程尝试获取同一个对象的锁，它们将被阻塞，直到锁被释放
- ReentrantLock
  - ReentrantLock是一个显式的锁类，提供了比synchronized更高级的功能，如可中断的锁等待、定时锁等待、公平锁选项等
  - ReentrantLock使用lock()和unlock()方法来获取和释放锁
  - 公平锁按照线程请求锁的顺序来分配锁，保证了锁分配的公平性，但可能增加锁的等待时间
  - 非公平锁不保证锁分配的顺序，可以减少锁的竞争，提高性能，但可能造成某些线程的饥饿
- 读写锁ReadWriteLock
  - ReadwriteLock允许多个线程同时读取共享资源，但只允许一个线程写入共享资源
  - 读写锁通常用于读取远多于写入的情况，以提高并发性
  - 读读不互斥, 读写互斥, 写写互斥

# **synchronized**

## **说一说自己对于 synchronized 关键字的了解 (高)**

- synchronized关键字解决的是多个线程之间访问资源的同步性，synchronized关键字可以保证被它修饰的方法或者代码块在任意时刻只能有一个线程执行
- 在 Java 早期版本中，synchronized属于重量级锁，效率低下，因为锁监视器（monitor）是依赖于底层的操作系统来实现的。如果要挂起或者唤醒一个线程，都需要操作系统帮忙完成，而操作系统实现线程之间的切换时需要从用户态转换到内核态，这个状态之间的转换需要相对比较长的时间，这也是为什么早期的synchronized 效率低的原因.
- Java 官方对从 JVM 层面对synchronized 较大优化，如自旋锁、适应性自旋锁、锁消除、锁粗化、偏向锁、轻量级锁等技术来减少锁操作的开销。所以现在的 synchronized 锁效率也优化得很不错了。

## **怎么使用 synchronized 关键字 (高)**

- 修饰非静态方法, 锁的是this(当前对象), 也就是一个对象用一把锁
- 修饰静态方法, 锁的是类.class , 也就是类的所有对象公用一把锁
- 修饰代码块, 括号里写什么就锁什么.

## **synchronized原理 (高)**

- 每个 Java 对象 在内存中分为三部分：对象头（Header）、实例数据（Instance Data）、对齐填充（Padding）

- 对象头中的 Mark Word：存储对象的哈希码、锁状态、GC 分代年龄等。
- 每个对象的Mark Word会关联一个 Monitor锁监视器, 锁监视器的结构如下

```

○ 1 class ObjectMonitor {
2     void*      _owner;           // 持有锁的线程
3     intptr_t   _recursions;    // 重入次数
4     Objectwaiter* _waitSet;    // 等待队列（调
用 wait() 的线程）
5     Objectwaiter* _EntryList;  // 阻塞队列（竞
争锁失败的线程）
6 };

```

在jvm命令层面, `synchronized` 同步语句块的实现使用的是`monitorenter` 和 `monitorexit` 指令, 其中 `monitorenter` 指令指向同步代码块的开始位置, `monitorexit` 指令则指明同步代码块的结束位置

- 当多个线程同时访问该方法, 那么这些线程会先被放进对象的锁池, 此时线程处于blocking状态
- 当一个线程获取到了实例对象的监视器 (monitor) 锁, 那么就可以进入running状态, 执行方法, 此时 `owner` 设置为当前线程, 重入次数加1表示当前对象锁被一个线程获取
- 当running状态的线程调用wait()方法, 那么当前线程释放 monitor对象, 进入waiting状态, `owner` 变为null, 重入次数减1, 同时线程进入等待池, 直到有线程调用notify()方法唤醒该线程, 则该线程重新获取monitor对象
- 如果当前线程执行完毕, 那么也释放monitor对象, 进入waiting状态, `owner` 变为null, `count` 减1

## **synchronized可重入如何实现 (高)**

Monitor锁监视器中的重入次数字段, 重入一次加一次, 释放一次减一次, 减到0则释放完毕.

## **JDK1.6 之后的 synchronized 底层做了哪些优化? (高)**

**java的线程模型是1对1的, 加锁需要调用操作系统的底层原语 mutex, 所以每次切换线程都需要操作系统切换到内核态, 开销很大. 这也是之前synchronized的问题所在, jdk1.6对其进行了优化, 从无锁到偏向锁到轻量级锁到重量级锁**

### **锁的四种状态**

JDK 1.6 引入了偏向锁和轻量级锁, 从而让锁拥有了四个状态: 无锁状态 (unlocked) 、偏向锁状态 (biasable) 、轻量级锁状态 (lightweight locked) 和重量级锁状态 (inflated) 。 虚拟机对象头里锁标志位, 就记录了这4中状态.

### **偏向锁**

#### **为什么引入偏向锁**

大多数时候是不存在锁竞争的, 常常是一个线程多次获得同一个锁, 因此如果每次都要竞争锁会增大很多没有必要付出的代价, 为了降低获取锁的代价, 才引入的偏向锁。

#### **偏向锁的过程**

- 当锁对象第一次被线程获得的时候, 使用 CAS 操作将线程 ID 记录到对象头的MarkWord中, 这个线程以后每次进入这个锁相关的同步块就不需要再进行任何同步操作。
- 当有另外一个线程去尝试获取这个锁对象时, 偏向状态就宣告结束, 此时撤销偏向后恢复到未锁定状态或者轻量级锁状态。

## 轻量级锁

### 为什么引入轻量级锁

轻量级锁考虑的是竞争锁对象的线程不多，而且线程持有锁的时间也不长的情景。因为阻塞线程需要CPU从用户态转到内核态，代价较大，如果刚刚阻塞不久这个锁就被释放了，那这个代价就有点得不偿失了，因此这个时候就干脆不阻塞这个线程，让它自旋这等待锁释放。

### 什么是轻量级锁

轻量级锁是相对于传统的重量级锁而言，它使用自旋 + CAS 操作来避免重量级锁使用互斥量的开销。

长时间的自旋会使CPU一直空转，浪费CPU，所以这里自旋是适应性自旋，自旋时间由上一个线程自旋的时间决定的。

### 轻量级锁什么时候升级为重量级锁

- 线程自旋的次数到了阈值，另外一个线程还没释放锁，那么就膨胀为重量级锁。
- 如果有两条以上的线程争用同一个锁，那轻量级锁就不再有效，要膨胀为重量级锁。
- 所谓的重量级锁，就是要阻塞其他线程，这里就是操作系统层面的阻塞了。需要从用户态切换到内核态，开销较大。所以只会在锁竞争大的时候进行。

## 锁消除

锁消除是指对于被检测出不可能存在竞争的共享数据的锁进行消除。

锁消除主要是通过逃逸分析来支持，如果堆上的共享数据不可能逃逸出去被其它线程访问到，那么就可以把它们当成私有数据对待，也就可以将它们的锁进行消除。

## 锁粗化

如果一系列的连续操作都对同一个对象反复加锁和解锁，频繁的加锁操作就会导致性能损耗。

比如连续使用StringBuffer的append() 方法就属于这类情况。如果虚拟机探测到由这样的一串零碎的操作都对同一个对象加锁，将会把加锁的范围扩展（粗化）到整个操作序列的外部。对于上一节的示例代码就是扩展到第一个 append() 操作之前直至最后一个 append() 操作之后，这样只需要加锁一次就可以了。

## Lock和synchronized异同？(高)

- 相同点
  - 两者都能保证线程安全，都是阻塞式的同步
  - 两者都是可重入锁。“可重入锁”指的是自己可以再次获取自己的内部锁。比如一个线程获得了某个对象的锁，此时这个对象锁还没有释放，当其再次想要获取这个对象的锁的时候还是可以获取的，如果是不可重入锁的话，就会造成死锁。同一个线程每次获取锁，锁的计数器都自增 1，所以要等到锁的计数器下降为 0 时才能释放锁。
- 不同点
  - synchronized 是java关键字，是依赖于jvm实现的。  
ReentrantLock 是 JDK 层面实现的一个接口，也就是 API 层面，需要 lock() 和 unlock() 方法配合 try/finally 语句块来完成。
  - synchronized 无法判断是否获取到了锁， ReentrantLock 可以判断是否获取到了锁(tryLock())。
  - synchronized 会自动释放锁， lock 必须要手动释放锁！如果不释放锁，死锁

- 相比 `synchronized`, `ReentrantLock` 增加了一些高级功能.  
公平锁, 等待可中断, 锁可以绑定多个条件.
- 可以实现公平锁, 通过构造函数传参创建公平锁
  - 等待可中断: 对于 `synchronized` 来说, 线程获取不到锁只能傻等, 而对于 `Lock` 来说, 正在等待的线程可以选择放弃等待, 改为处理其他事情, 这就可以避免出现死锁的情况. 使用 `lockInterruptibly()` 方法能够响应中断, 即中断线程的等待状态。例如, 当两个线程A、B同时通过 `lock.lockInterruptibly()` 获取锁时, 假若此时线程A获取到了锁, 而线程B进入等待状态, 那么线程B就可调用 `interrupt()` 方法中断线程B的等待过程。  
(`interrupt()` 方法只能中断阻塞过程中的线程)
  - 可实现选择性通知 (锁可以绑定多个条件) :  
`synchronized` 关键字与 `wait()` 和 `notify()`/`notifyAll()` 方法相结合可以实现等待/通知机制。 `ReentrantLock` 类当然也可以实现, 但是需要借助于 `Condition` 接口与 `newCondition()` 方法。  
`Condition` 接口有 `await()` 使一个线程等待 `signal()` 唤醒一个线程 `signalAll()` 唤醒所有线程。一个 `ReentrantLock` 对象可以同时绑定多个 `Condition` 对象。

## CAS

### CAS了解吗 (高)

CAS是一种乐观锁, 它抱着乐观的态度认为自己一定可以成功。

- 当多个线程同时使用CAS操作一个变量时, 只有一个会胜出, 并成功更新, 其余均会失败

- 失败的线程不会被挂起，仅是被告知失败，并且允许再次尝试，当然也允许失败的线程放弃操作

cas算法的过程是这样的，cas包括有三个值：

- v表示要更新的变量
- e表示预期值，就是旧的值
- n表示新值

更新时，判断只有e的值等于v变量的当前旧值时，才会将n新值赋给v，更新为新值。否则，则认为已经有其他线程更新过了，则当前线程更新失败

## CAS原理 (高)

- 原子类的底层就是调用了unsafe类，unsafe类来做CAS操作。
- unsafe类中的compareAndSwap，调用了native修饰的方法，其底层是使用了汇编指令cmpxchg(如果是x86的话)，它能够实现指令级的原子操作。(这里的原子操作是由硬件保证的)
- 纯软件方法是无法保证cas操作的原子性的。

## 乐观锁存在哪些问题？ (高)

- 1、自旋CAS的方式如果长时间不成功，会给CPU带来很大的开销
- 2、一次性只能保证一个共享变量的原子性(多个可以通过AtomicReference来处理或者使用锁synchronized实现)
- 3、ABA问题，但是ABA的问题大部分场景下都不影响并发的最终效果

## 什么是ABA问题？ABA问题怎么解决？ (高)

在做cas操作时

- 线程1获取值为A
- 线程2修改值为B
- 线程3修改值为A
- 线程1获取值还是A, 以为该值从未被修改过, 这就是ABA问题

解决aba问题, 解决ABA问题的常见方法是使用带有版本号的CAS, 比如用 `AtomicStampedReference` 类.

- 该类的原理是新增一个版本号, 每次操作给版本号一个新值. 进行 cas操作时不仅要数值和之前一样, 版本号也要一样.
- 当数值一样, 但是版本号不同时, 说明产生了aba问题. 此时值不会被修改.

ABA的问题大部分场景下都不影响并发的最终效果, 所以在很多场景下其实可以不用解决

## AQS

## AQS是什么？ AQS的原理是什么？ (高)

AQS是juc中的核心工具, 用于构建锁和其他并发工具类（`ReentrantLock`、`Semaphore`、`CountDownLatch` 等）

核心设计:

- **等待队列**: 双向 FIFO 队列, 存储等待线程的节点 (Node) , 处理线程的排队与阻塞唤醒逻辑
- **state (状态变量)** : 表示同步状态 (如锁的持有次数、信号量剩余资源数)
- **cas操作**: 通过 CAS 操作保证修改state状态的原子性

基本思想:

- 线程去获取共享资源, 能获取到就直接设置该线程为该资源的有效工作线程, 并且将共享资源设置为锁定状态 (通过cas操作修改 state状态)
- 如果被请求的共享资源被占用, 那么就将暂时获取不到锁的线程加入到等待队列中. 等待被唤醒后重新获取共享资源.

## 同步状态及相关方法

- AQS 使用 **int 类型的成员变量 state 表示同步状态**, state 变量由 volatile 修饰, 用于表示当前临界资源的获取情况
- 状态信息 state 可以通过compareAndSetState() cas机制进行操作 (操作资源的占用和释放)

## 等待队列

- 等待队列是一个虚拟的双向队列 (虚拟的双向队列即不存在队列实例, 仅存在结点之间的关联关系)
- AQS 是将每条请求共享资源的线程封装成一个 CLH 锁队列的一个结点 (Node) 来实现锁的分配。

- 在 CLH 同步队列中，一个节点表示一个线程，它保存着线程的引用 (thread) 、当前节点在队列中的状态 (state) 、前驱节点 (prev) 、后继节点 (next)
- 当获取资源失败，节点会通过CAS自旋加入队列尾部
- 当资源释放时(使state=0, 资源变为可获取状态)，AQS 从队列中唤醒线程，来获取资源

## 简单总结描述

AQS就是用volatile修饰共享变量state表示同步状态，线程通过CAS去修改state，state从0修改为1则成功获取资源，否则修改失败则进入等待队列(双向链表实现的CLH虚拟双向队列)，等待资源释放被重新唤醒。

## 讲一下aqs的独占和共享模式 (中)

aqs有独占和共享两种模式.

- 共享模式就是多个线程可以共享资源，适用于多个线程协作的场景.
- 比如说Semaphore或者CountDownLatch.
  - 对于Semaphore来说state就代表为许可证数量. 比如当某个资源能同时被3个线程访问，那state初始化就为3, 获取时state递减，减到0后新线程就无法访问共享资源了，就得进aqs等待队列去等待了. 当线程释放资源时state递增.
  - CountDownLatch也是同理, state表示剩余任务数, 调用countDown()时递减, state归零时唤醒所有await()线程.
- 独占模式就是只允许单线程独占资源，适用于互斥场景，比如锁，也就是ReentrantLock.

- 独占模式通过acquire() 与 tryAcquire()获取资源, 对应 ReentrantLock的lock和tryLock().
- 当state为0表示资源空闲, 加锁时, cas修改state, 修改为1, 表示加锁成功, 加锁失败的就进入aqs等待队列.

## ReentrantLock

### ReentrantLock底层原理 (高)

ReentrantLock底层是用AQS实现的, 主要思想就是获取锁时通过自旋+cas去修改aqs的state, 获取不到就进入aqs的等待队列中.

- state 初始值为 0, 表示未锁定状态。
- A 线程 lock() 时, 会调用AQS的 tryAcquire() 方法, cas将state值从0修改为1.
- 此后, 其他线程再 tryAcquire() 时就会失败, 然后进入aqs的等待队列中, 被挂起.
- 当 A 线程 unlock() 到 state=0 (即释放锁) 为止, 触发AQS的唤醒机制
- AQS从队列头部开始遍历, 然后唤醒队头节点的下一节点对应的线程, 让其来拿锁.

### ReentrantLock如何实现可重入 (高)

用aqs的state充当锁计数器, 重入一次加一次, 释放一次减一次, 减到0则释放完毕.

# Volatile

---

Volatile是Java虚拟机提供轻量级的同步机制

## 三大特性 (高)

- 禁止进行指令重排 (有序性)
- 保证可见性, 指线程之间的可见性, 一个线程修改的状态对另一个线程是可见的. 线程每次获取volatile修饰的变量时, 都能获取到最新值. (volatile本质是在告诉jvm当前变量在工作内存(cache缓存/寄存器)中的值是不确定的, 需要从主存中读取) **本质是禁用CPU缓存, 每次去主存取值**
- 不保证原子性

## Volatile和synchronized 的区别 (高)

- volatile本质是在告诉jvm当前变量在工作内存(CPU缓存/寄存器)中的值是不确定的, 需要从主存中读取. synchronized则是锁定当前变量, 只有当前线程可以访问该变量, 其他线程被阻塞住
- volatile仅能使用在变量级别; synchronized则可以使用在变量、方法、和类级别的。
- volatile仅能实现变量的修改可见性, 并不能保证原子性; synchronized则可以保证变量的修改可见性和原子性
- volatile不会造成线程的阻塞; synchronized可能会造成线程的阻塞

## 为什么volatile不能保证线程安全? (高)

很简单呀，可见性不能保证操作的原子性.

## 如果不使用lock和synchronized, 怎么保证num++的原子性? (高)

可以使用AtomicInteger原子类.

# ThreadLocal

---

## 什么是 ThreadLocal / ThreadLocal 有什么用? (高)

对于多线程共享的资源, 为了避免线程安全问题, 可以加锁处理. 也可以用ThreadLocal.

ThreadLocal可以理解为线程本地变量(不共享), 他会在每个线程都创建一个副本, 那么在线程之间访问内部副本变量就行了, 每个线程各用各的, 这做到了线程之间互相隔离, 解决了线程安全问题.

ThreadLocal两大特点:

- 线程间资源隔离
- 单个线程内资源共享

## ThreadLocal原理 (高)

每个线程内有一个ThreadLocalMap类型的成员变量, 用来存储资源对象, 通过当前线程就能拿到ThreadLocalMap

ThreadLocalMap又包含了一个Entry数组，Entry本身是一个弱引用，他的key是指向ThreadLocal的弱引用，Entry具备了保存key value键值对的能力。

- 调用set方法，就是以ThreadLocal自己作为key,资源对象作为value,放入当前线程的ThreadLocalMap集合中
- 调用get方法，就是以ThreadLocal自己作为key,到当前线程中查找关联的资源值
- 调用remove方法，就是以ThreadLocal自己作为key,移除当前线程关联的资源值

**不同的线程之间threadlocal这个key值是一样，但是不同的线程所拥有的ThreadLocalMap是独一无二的**

## threadlocal内存泄漏问题? (高)

### 为什么会内存泄漏

如果说一个线程用完就释放，那么没有问题。但是实际开发中为了避免资源浪费，往往会大量使用线程池，线程池的核心线程是不会释放的，会重复利用。那么就会有一条线 GCRoots -> ... -> Thread -> ThreadLocalMap -> Entry。

当线程池某个Thread调用threadlocalset方法往ThreadLocalMap放数据，放完不删除。该线程之后并不会被销毁，而是会重新放回线程池中。那么此时ThreadLocalMap中的数据无法被gc回收，就有可能有内存泄漏问题。

### 为什么key是弱引用

假设key为强引用

- 在业务代码中执行了 ThreadLocal instance = null 操作，想清理掉这个 ThreadLocal 实例

- 但是在 ThreadLocalMap 的 Entry 中强引用了 ThreadLocal 实例(key)
- 虽然在业务代码中把 ThreadLocal 实例置为了 null，但是依然有这个引用链(entry->ThreadLocal )的存在

GC 在垃圾回收的时候会进行可达性分析，它会发现这个 ThreadLocal 对象依然是可达的，所以对于这个 ThreadLocal 对象不会进行垃圾回收，这样的话就造成了内存泄漏的情况。所以Entry 中对key是弱引用。

一旦将 ThreadLocal 变量置为 null，在下一次 GC 时，ThreadLocal 实例会被回收，Entry 的 key 变为 null

此时调用 ThreadLocal 的 set()、get() 或 remove() 方法时，会触发清理 key 为 null 的 Entry，释放 value 的强引用

### value是如何清除的 / 如何避免value内存泄漏

ThreadLocalMap 中使用的 key 为 ThreadLocal 的弱引用，而 value 是强引用。所以，如果 ThreadLocal 没有被外部强引用的情况下，在垃圾回收的时候，key 会被清理掉，而 value 不会被清理掉。

这样一来，ThreadLocalMap 中就会出现 key 为 null 的 Entry。假如我们不做任何措施的话，value 永远无法被 GC 回收，这个时候就可能会产生内存泄露。ThreadLocalMap 实现中已经考虑了这种情况，在调用 set()、get()、remove() 方法的时候，会清理掉 key 为 null 的记录。使用了 ThreadLocal 又不再调用 get(),set(),remove()方法，那么就会导致内存泄漏，所以使用完 ThreadLocal 方法后 最好手动调用 remove() 方法

## 为什么要清除掉threadlocal? (高)

- ThreadLocalMap 使用 ThreadLocal 的弱引用作为 key
- 如果一个 ThreadLocal 不存在外部强引用时, Key(ThreadLocal)势必会被 GC 回收, 这样就会导致 ThreadLocalMap 中 key 为 null, 而 value 还存在着强引用 (value 可能会被其他线程所引用), 只有 thread 线程退出以后, value 的强引用链条才会断掉
- 但如果当前线程再迟迟不结束的话, 这些 key 为 null 的 Entry 的 value 就会一直存在一条强引用链: ThreadRef->Thread-> ThreadLocalMap->Entry -> value 永远无法回收, 造成内存泄漏。

ThreadLocal 正确的使用方法: 每次使用完 ThreadLocal 都调用它的 remove() 方法清除数据。

- **key 被回收**: ThreadLocal 实例被回收后, Entry 的 key 变为 null, 此时 Entry 成为 **无效条目**。
- **主动清理机制**: 调用 ThreadLocal 的 set()、get() 或 remove() 方法时, 会触发清理 key 为 null 的 Entry, 释放 value 的强引用。

## ThreadLocal 与 synchronized 的不同之处 (高)

- 作用不同
  - synchronized 主要是针对对象或者类的, 主要是防止多个线程同时执行一个代码块产生的问题;
  - ThreadLocal 主要针对变量的, 主要是防止共享变量在多线程下产生的安全问题。
- 机制不同

- synchronized 是采用锁的机制，保证一次只有一个线程执行代码，在前一个线程执行代码期间其他线程都处于等待状态
- ThreadLocal是采用拷贝变量副本的形式来让每个线程维护各自的变量，保证每个线程只会访问到自己变量的副本, 多个线程同时访问也没有问题.