# 锁

## 锁的分类

**偏向锁**： 通常只有一个线程在临界区执行，临界区即同步代码块内。

**轻量级锁**： 可以有多个线程交替进入临界区，在竞争不激烈的时候，稍微自旋等待一下就能获得锁。

**重量级锁**：也是我最为期待的锁，那就是出现了激烈的竞争，只好让我们去阻塞休息了。

## Lock实现原理

lock实现过程中的几个关键词：计数值、双向链表、CAS+自旋

ReentrayLock继承于AbstractQueuedSynchronizer（简称AQS），主要代码都是在AQS

锁的存储结构就两个东西:"双向链表" + "int类型状态"

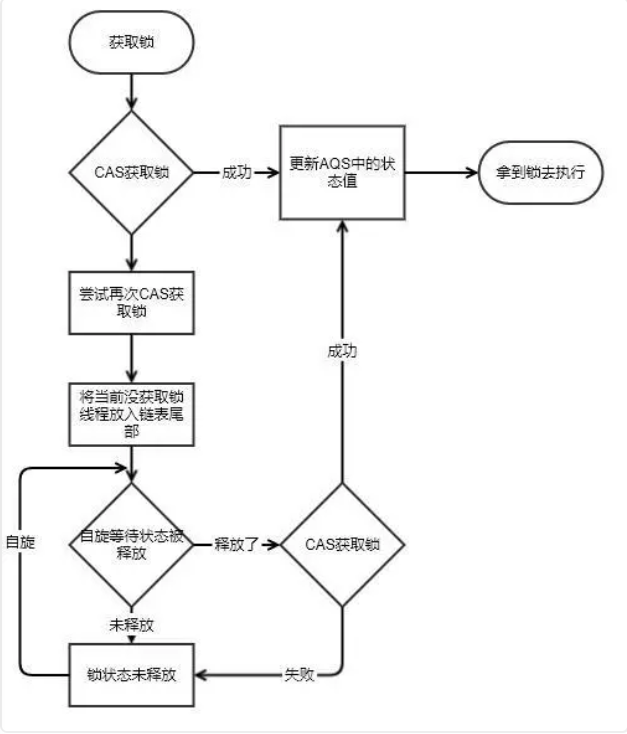
总结：

lock的存储结构：一个int类型状态值（值为01用于锁的状态变更），一个双向链表（用于存储等待中的线程）

lock获取锁的过程：本质上是通过CAS来获取状态值修改，如果当场没获取到，会将该线程放在线程等待链表中。

lock释放锁的过程：修改状态值，调整等待链表。

可以看到在整个实现过程中，lock大量使用CAS+自旋。因此根据CAS特性，lock建议使用在低锁冲突的情况下。目前java1.6以后，官方对synchronized做了大量的锁优化（偏向锁、自旋、轻量级锁）。因此在非必要的情况下，建议使用synchronized做同步操作。



## Synchronized实现原理

synchronized可以保证方法或者代码块在运行时，同一时刻只有一个方法可以进入到临界区，同时它还可以保证共享变量的内存可见性

synchronized实现同步的基础：

普通同步方法，锁是当前实例对象

静态同步方法，锁是当前类的class对象

同步方法块，锁是括号里面的对象

利用javap工具查看生成的class文件信息来分析Synchronize的实现可以看出，同步代码块是使用monitorenter和monitorexit指令实现的,同步方法（在这看不出来需要看JVM底层实现）依靠的是方法修饰符上的ACCSYNCHRONIZED实现。

Java对象头和monitor是实现synchronized的基础

synchronized用的锁是存在Java对象头里的，那么什么是Java对象头呢？Hotspot虚拟机的对象头主要包括两部分数据：Mark Word（标记字段）、Klass Pointer（类型指针）

什么是Monitor？我们可以把它理解为一个同步工具，也可以描述为一种同步机制，它通常被描述为一个对象。 与一切皆对象一样，所有的Java对象是天生的Monitor，每一个Java对象都有成为Monitor的潜质，因为在Java的设计中 ，每一个Java对象自打娘胎里出来就带了一把看不见的锁，它叫做内部锁或者Monitor锁。

原文：https://mp.weixin.qq.com/s/nU0M3QVE209S79bCrrwrSg

## atomic的实现原理

synchronized或者lock同步数据的操作过于繁重，concurrent包下的atomic提供我们这么一种轻量级的数据同步的选择。

AtomicInteger的实现原理为

他的值是存在一个volatile的int里面。volatile只能保证这个变量的可见性。不能保证他的原子性。

如何保证原子性：自旋 + CAS（乐观锁）。在这个过程中，通过compareAndSwapInt比较更新value值，如果更新失败，重新获取旧值，然后更新。

CAS相对于其他锁，不会进行内核态操作，有着一些性能的提升。但同时引入自旋，当锁竞争较大的时候，自旋次数会增多。cpu资源会消耗很高。

Java 8做出的改进和努力

在Java 8中引入了4个新的计数器类型，LongAdder、LongAccumulator、DoubleAdder、DoubleAccumulator。他们都是继承于Striped64。

在LongAdder 与AtomicLong有什么区别？

Atomic\*遇到的问题是，只能运用于低并发场景。因此LongAddr在这基础上引入了分段锁的概念。大概就是当竞争不激烈的时候，所有线程都是通过CAS对同一个变量（Base）进行修改，当竞争激烈的时候，会将根据当前线程哈希到对于Cell上进行修改（多段锁）。

## volatile实现原理

volatile的两大特性：禁止重排序、内存可见性

volatile是一种类型修饰符，被volatile声明的变量表示随时可能发生变化，每次使用时，都必须从变量i对应的内存地址读取，编译器对操作该变量的代码不再进行优化。它是基于底层c++实现的。

## CAS与Synchronized的关系

Synchronized属于悲观锁，Synchronized关键字会让没有得到锁资源的线程进入BLOCKED状态，而后在争夺到锁资源后恢复为RUNNABLE状态，这个过程中涉及到操作系统用户模式和内核模式的转换，代价比较高。

CAS属于乐观锁，它会让当前线程处于忙循环，一直去争取获取锁。

什么时候使用CAS锁

atomic类，以及lock类的底层实现都会设计到CAS锁。在Java1.6以上的版本,synchronized转变为重量级锁之前，也会采用CAS机制。

CAS的缺点：

1.CPU开销较大：在并发量比较高的情况下，如果许多线程反复尝试更新某一个变量，却又一直更新不成功，循环往复，会给CPU带来很大的压力。

2.不能保证代码块的原子性：CAS机制所保证的只是一个变量的原子性操作，而不能保证整个代码块的原子性。比如需要保证3个变量共同进行原子性的更新，就不得不使用Synchronized了。

3.ABA问题：这是CAS机制最大的问题所在。

## 锁的原理和实现方式

AQS

CHL: AQS 里将阻塞线程封装到一个内部类 Node 里。并维护一个 CHL Node FIFO 队列。 CHL队列是一个非阻塞的 FIFO 队列，也就是说往里面插入或移除一个节点的时候，在并发条件下不会阻塞，而是通过自旋锁和 CAS 保证节点插入和移除的原子性。实现无锁且快速的插入。

EXCLUSIVE：排它

SHARED：共享

1、什么是锁？

一种保护机制，在多线程的情况下，保证操作数据的正确性与一致性

2、锁有哪些分类？

悲观锁、乐观锁、独占锁、共享锁、公平锁、非公平锁、分布式锁、自旋锁

3、谈谈悲观锁乐观锁：

指的是看待并发同步的角度，一般结合数据库将，以MySQL为例，悲观锁主要是表锁，行锁和间隙锁，叶锁，读锁，因为这些锁在被触发时会引起线程阻塞，所以叫悲观锁；而乐观锁其实在MySQL中本身不存在，但是MySQL提供了MVCC的机制，支持乐观锁机制

4、什么是MVCC？

只有在InnoDB引擎下存在，MVCC是为了实现事务的隔离性，通过版本号，避免同一数据在不同事务间的竞争，所说的乐观锁只在事务级别未提交锁和已提交锁时才会生效

5、具体MVCC机制有什么？

多版本并发控制，保证数据操作在多线程过程中，保证事务隔离的机制，可以降低锁竞争的压力，保证较高的并发量。在每开启一个事务时，会生成一个事务的版本号，被操作的数据会生成一条新的数据行（临时），但是在提交前对其他事务是不可见的，对于数据的更新操作成功，会将这个版本号更新到数据的行中，事务提交成功，将新的版本号更新到此数据行中，这样保证了每个事务操作的数据，都是互不影响的，也不存在锁的问题

6、在多个事务操作同一条数据的并发过程中，谁先成功?

mysql判断，先提交的先成功

7、谈谈事务。

事务常说一系列操作作为一个整体要么都成功要么都失败，主要特性acid，事务的的实现主要依赖两个log redo-log,undo-log,每次事务都会记录数据修改前的数据undo-log，修改后的数据放入redo-log,提出成功则使用redo-log 更新到磁盘，失败则使用undo-log将数据恢复到事务之前的数据

8、再谈谈独占锁、共享锁：

独占锁即是持有锁的线程只有一个，共享锁则可以有多个线程

9、共享的意义在于？

共享锁是为了提高程序的效率，举个例子：数据的操作有读写之分，对于写的操作加锁，保证数据正确性，而对于读的操作如果不加锁，在写读操作同时进行时，读的数据有可能不是最新数据，如果对读操作加独占锁，面对读多写少的程序肯定效率很低，所有就出现了共享锁，对于读的的操作就使用共享的概念，但是对于写的操作则是互斥的，保证了读写的数据操作都一致，在java中上述的锁叫读写锁

10、读写锁的机制是什么？

在Java中读写锁（ReadWritelock）的机制是基于AQS的一种实现，保证读读共享、读写互斥、写写互斥，如果要说机制的话，得从AQS说起：

AQS是Java实现的一种锁机制，互斥锁、读锁写锁、条件产量、信号量、栅栏都是AQS的衍生物，主要工作基于CHL队列，volatile关键字修饰状态state，线程通过CAS去修改状态符，成功了则获取成功，失败了就进队列等待，等待唤醒。

AQS中还有一个很重要的概念就是自旋：在等待唤醒的时候，经常会使用自旋（while(!cas())）的方式，不停地尝试获取锁，直到被其他线程获取成功

**共享与独占的区别在于：**

CHL队列中的节点的模式是EXCLUSIVE还是SHARED，当一个线程成功修改了stat状态，表示获取了锁，如果线程所在的节点为SHARED，将开始一个读锁传递的过程，从头结点，向队列后续节点传递唤醒，直到队列结束或者遇到了EXCLUSIVE的节点，等待所有激活的读操作完成，然后进入到独享模式

公平与非公平的区别就在于线程第一次获取锁时，也就是执行修改stat操作时，是进队列还是直接修改状态，这是基本的工作机制

11、Java除了AQS还有其他的锁支持么？

在java中，synchronized关键字，是语言自带的，也叫内置锁，synchronized关键字，被synchronized修饰的方法或者代码块，在同一时间内，只允许一个线程执行，是独享锁

12、synchronized的实现机制？

实现有两部分：monitor对象，线程，工作机制还是线程抢占对象使用权，对象有自己的对象头，存储了对象的很多信息，其中有一个是标识被哪个线程持有，对比AQS，线程从修改state，变为修改monitor的对象头，线程的等待区域动 AQS中的队列，变为monitor对象中的某个区域

13、详细介绍内置锁。

锁一直是围绕线程安全来实现的，比如独占锁，它在内存里面的操作是怎么样的

这个地方涉及到一个概念，内存模型（这个和jvm不要混淆，

The Java memory model used internally in the JVM divides memory between thread stacks and the heap. This diagram illustrates the Java memory model from a logic perspective）

内存模型是JVM用来区别线程栈和堆的内存方式，每个线程在运行的时候，所操作的数据存储空间有两个，一个是主内存，一个是工作内存

主内存其实就是jvm中堆，工作内存就是线程的栈，每次的数据操作，都是从主内存中把数据读到工作内存中，然后在工作内存中进行各种处理，如果进行了修改，会把数据回写到主内存，然后其他线程又进行同样的操作，就这样数据在工作内存和主内存，不断拷贝与刷新，在高并发下，就出现了数据不一致的问题，导致了多线程的不安全性；整个操作过程还牵涉到CPU，高速缓存等概念

14、内存模型还有哪些可以说说？

happen-befor 原则，Volatile 关键字（线程的可见性），内存屏障

（1）happen-befor原则定义了内存模型执行过程中的定律，就像1+1 = 2，不可能被打破的jvm的运行机制都依赖于这个原则，是jvm的“宪法”

（2）volatile关键字：提供内存可见性和禁止内存重排序

volatile使用场景：

两个条件如下：

对变量的写操作不依赖当前值，如多线程下执行a++，是无法通过volatile保证结果准确性的

该变量没有包含在具有其他变量的不变式中

普通变量和volatile变量的异同：

对于普通变量：读操作会优先读取工作内存的数据，如果工作内存中不存在，则从主内存中拷贝一份数据到工作内存中；写操作只会修改工作内存的副本数据，这种情况下，其它线程就无法读取变量的最新值

对于volatile变量（表层）：读操作时JMM会把工作内存中对应的值设为无效，要求线程从主内存中读取数据；写操作时JMM会把工作内存中对应的数据刷新到主内存中，这种情况下，其它线程就可以读取变量的最新值

内存可见性保证原理（底层实现）：volatile变量的内存可见是基于内存屏障（Memory Barrier）实现的。

内存屏障：又称内存栅栏，是一个CPU命令，在程序运行时，为了提高执行性能，编译器和处理器会对指令进行重排序，JMM为了保证在不同的编译器和CPU上有相同的结果（可见性），通过插入特定类型的内存屏障来禁止特定类型的编译器重排序和处理重排序，插入一条内存屏障会告诉编译器和CPU：不管什么指令都不能和这条内存屏障指令重排序（禁止重排序）

Synchronized锁实现方式使用的是对象，每个对象都有头和尾

**ReentrantLock的调用过程**

经过观察ReentrantLock把所有Lock接口的操作都委派到一个Sync类上，该类继承了AbstractQueuedSynchronizer

锁实现（加锁）

简单说来，AbstractQueuedSynchronizer会把所有的请求线程构成一个CLH队列，当一个线程执行完毕（lock.unlock()）时会激活自己的后继节点，但正在执行的线程并不在队列中，而那些等待执行的线程全部处于阻塞状态，经过调查线程的显式阻塞是通过调用LockSupport.park()完成，而LockSupport.park()则调用sun.misc.Unsafe.park()本地方法，再进一步，HotSpot在Linux中中通过调用pthread\_mutex\_lock函数把线程交给系统内核进行阻塞。与synchronized相同的是，这也是一个虚拟队列，不存在队列实例，仅存在节点之间的前后关系。令人疑惑的是为什么采用CLH队列呢？原生的CLH队列是用于自旋锁，但Doug Lea把其改造为阻塞锁。   
当有线程竞争锁时，该线程会首先尝试获得锁，这对于那些已经在队列中排队的线程来说显得不公平，这也是非公平锁的由来，与synchronized实现类似，这样会极大提高吞吐量。   
如果已经存在Running线程，则新的竞争线程会被追加到队尾，具体是采用基于CAS的Lock-Free算法，因为线程并发对Tail调用CAS可能会导致其他线程CAS失败，解决办法是循环CAS直至成功。

Synchronized

为了保证共享数据在同一时刻只被一个线程使用，我们有一种很简单的实现思想，就是在共享数据里保存一个锁 ，当没有线程访问时，锁是空的。当有第一个线程访问时，就 在锁里保存这个线程的标识 并允许这个线程访问共享数据。在当前线程释放共享数据之前，如果再有其他线程想要访问共享数据，就要 等待锁释放 。

jvm中有以下三种锁(由上到下越来越“重量级”)：偏向锁,轻量级锁,重量级锁

重量级锁

**Synchronized 原理**

每个对象有一个监视器锁（monitor）。

**轻量级锁**

轻量级锁的核心思想就是“被加锁的代码不会发生并发，如果发生并发，那就膨胀成重量级锁(膨胀指的锁的重量级上升，一旦升级，就不会降级了)”。轻量级锁依赖了一种叫做CAS(compare and swap)的操作。

偏向锁

根据轻量级锁的实现，我们知道虽然轻量级锁不支持“并发”，遇到“并发”就要膨胀为重量级锁，但是轻量级锁可以支持多个线程以串行的方式访问同一个加锁对象。比如A线程可以先获取对象o的轻量锁，然后A释放了轻量锁，这个时候B线程来获取o的轻量锁，是可以成功获取得，以这种方式可以一直串行下去。之所以能实现这种串行，是因为有一个释放锁的动作。那么假设有一个加锁的java方法，这个方法在运行的时候其实从始至终只有一个线程在调用，但是每次调用完却也要释放锁，下次调用还要重新获得锁。那么我们能不能做一个假设：“假设加锁的代码从始至终就只有一个线程在调用，如果发现有多于一个线程调用，再膨胀成轻量级锁也不迟”。这个假设，就是偏向锁的核心思想。偏向锁依赖了一种叫做CAS(compare and swap)的操作。

Synchronized的语义底层是通过一个monitor的对象来完成，其实**wait/notify等方法也依赖于monitor对象，这就是为什么只有在同步的块或者方法中才能调用wait/notify等方法**，否则会抛出java.lang.IllegalMonitorStateException的异常的原因。

Lock VS Synchronized

AbstractQueuedSynchronizer通过构造一个基于阻塞的CLH队列容纳所有的阻塞线程，而对该队列的操作均通过Lock-Free（CAS）操作，但对已经获得锁的线程而言，ReentrantLock实现了偏向锁的功能。

synchronized的底层也是一个基于CAS操作的等待队列，但JVM实现的更精细，把等待队列分为ContentionList和EntryList，目的是为了降低线程的出列速度；当然也实现了偏向锁，从数据结构来说二者设计没有本质区别。但synchronized还实现了自旋锁，并针对不同的系统和硬件体系进行了优化，而Lock则完全依靠系统阻塞挂起等待线程。

当然Lock比synchronized更适合在应用层扩展，可以继承AbstractQueuedSynchronizer定义各种实现，比如实现读写锁（ReadWriteLock），公平或不公平锁；同时，Lock对应的Condition也比wait/notify要方便的多、灵活的多。

|  |  |  |  |
| --- | --- | --- | --- |
| 锁 | 优点 | 缺点 | 适用场景 |
| 偏向锁 | 加锁和解锁不需要额外的消耗，和执行非同步方法比仅存在纳秒级的差距。 | 如果线程间存在锁竞争，会带来**额外的锁撤销的消耗**。 | 适用于**只有一个线程访问**同步块场景。 |
| 轻量级锁 | 竞争的线程不会阻塞，提高了程序的响应速度。 | 如果始终得不到锁竞争的线程使用**自旋会消耗CPU**。 | 追求**响应时间**。  同步块执行速度非常快。 |
| 重量级锁 | 线程竞争不使用自旋，不会消耗CPU。 | **线程阻塞**，响应时间缓慢。 | 追求**吞吐量**。  同步块执行速度较长。 |

## Lock与synchronized的区别

          1. Lock的加锁和解锁都是由java代码配合native方法（调用操作系统的相关方法）实现的，而synchronize的加锁和解锁的过程是由JVM管理的

          2. 当一个线程使用synchronize获取锁时，若锁被其他线程占用着，那么当前只能被阻塞，直到成功获取锁。而Lock则提供超时锁和可中断等更加灵活的方式，在未能获取锁的     条件下提供一种退出的机制。

          3. 一个锁内部可以有多个Condition实例，即有多路条件队列，而synchronize只有一路条件队列；同样Condition也提供灵活的阻塞方式，在未获得通知之前可以通过中断线程以    及设置等待时限等方式退出条件队列。

         4. synchronize对线程的同步仅提供独占模式，而Lock即可以提供独占模式，也可以提供共享模式

## 如何避免死锁

程序中必须同时满足以下四个条件才会引发死锁：

1. 互斥（Mutual exclusion）：线程所使用的资源中至少有一个是不能共享的，它在同一时刻只能由一个线程使用。

2. 持有与等待（Hold and wait）：至少有一个线程已经持有了资源，并且正在等待获取其他的线程所持有的资源。

3. 非抢占式（No pre-emption）：如果一个线程已经持有了某个资源，那么在这个线程释放这个资源之前，别的线程不能把它抢夺过去使用。

4. 循环等待（Circular wait）：假设有 N 个线程在运行，第一个线程持有了一个资源，并且正在等待获取第二个线程持有的资源，而第二个线程正在等待获取第三个线程持有的资源，依此类推……第N 个线程正在等待获取第一个线程持有的资源，由此形成一个循环等待。

避免死锁最简单的方法就是阻止循环等待条件，将系统中所有的资源设置标志位、排序，规定所有的进程申请资源必须以一定的顺序（升序或降序）做操作来避免死锁。

# 线程和线程池

## 什么是线程

线程（英语：thread）是操作系统能够进行运算调度的最小单位。它被包含在进程之中，是进程中的实际运作单位。一条线程指的是进程中一个单一顺序的控制流，一个进程中可以并发多个线程，每条线程并行执行不同的任务。

一个进程至少得有一个线程，单线程只能干一件事情，无法并发和并行。进程之间是隔离的，进程是拥有资源的基本单位， 线程是CPU调度的基本单位，

JVM是一个进程，所以只能多线程编程。

## Java中线程具有五种状态：

初始化；可运行；运行中；阻塞；销毁



但是，线程不同状态之间的转化是谁来实现的呢？是JVM吗？

并不是。JVM需要通过操作系统内核中的TCB（Thread Control Block）模块来改变线程的状态，这一过程需要耗费一定的CPU资源。

## 什么是协程

协程，英文Coroutines，是一种比线程更加轻量级的存在。正如一个进程可以拥有多个线程一样，一个线程也可以拥有多个协程

最重要的是，协程不是被操作系统内核所管理，而完全是由程序所控制（也就是在用户态执行）。

这样带来的好处就是性能得到了很大的提升，不会像线程切换那样消耗资源。

由于Java的原生语法中并没有实现协程（某些开源框架实现了协程，但是很少被使用）

## 线程池

单个线程: Executors.newSingleThreadExecutor();

缓存线程: Executors.newCachedThreadPool();

固定线程Executors.newFixedThreadPool(2);

定时线程: Executors.newScheduledThreadPool(3);

## ThreadPoolExecutor

ThreadPoolExecutor 继承 AbstractExecutorService；AbstractExecutorService 实现 ExecutorService， ExecutorService 继承 Executor

public ThreadPoolExecutor(int corePoolSize,

int maximumPoolSize,

long keepAliveTime,

TimeUnit unit,

BlockingQueue<Runnable> workQueue,

ThreadFactory threadFactory,

RejectedExecutionHandler handler) {...}

corePoolSize : 核心线程数的大小

maximumPoolSize : 线程池中允许的最大线程数

keepAliveTime : 空闲线程允许的最大的存活时间

unit : 存活时间的单位

workQueue : 阻塞任务队列

threadFactory : 线程工厂用来创建线程

handler : 拒绝策略，针对当队列满了时新来任务的处理方式

**线程池的执行过程：**

在刚刚创建线程池的时候，内部线程的数量是 0，当首个任务进行添加的时候，会根据参数的配置进行线程的创建，并随着任务数的增加，会逐渐创建新的线程直到线程数达到核心线程的大小。这时，如果再添加任务会将任务放置到阻塞队列当中,如果队列装不下了这时会判断当前线程的数量是否达到了最大线程数，如果未达到，新添加的任务会由最大重新创建线程并立马执行。如果队列满了，而且已经达到最大线程数了，这时再添加任何就会由拒绝策略来处理，默认的拒绝策略是抛出异常。

## 为什么要使用线程池

虽然线程是个轻量级的东西， 但是对于互联网应用来说，如果每个用户的请求都创建一个线程，那会非常得多，服务器也是难于承受， 再说了，众多的线程去竞争CPU，不断切换，也会让CPU调度不堪重负，很多线程将不得不等待。

线程可以预先创建，当线程池的线程刚创建时，让他们进入阻塞状态：等待某个任务的到来。 如果任务来了，那就好办，唤醒其中一个线程，让它拿到任务去执行即可。

## 线程池原理

原文：<https://mp.weixin.qq.com/s/LExpStGw7XXp0M1E_T0yKA>

https://mp.weixin.qq.com/s/DJPuYVhRGRXkkUGMuDdrSw

根据源码可以发现整个线程池大致分为 3 个部分，1. 是创建 worker 线程，2. 添加任务到 workQueue; 3.worker 线程执行具体任务

线程池每个线程会包装成worker，用来循环执行take任务。

第一步是生产者，也就是任务提供者他执行了一个 execute() 方法，本质上就是往这个内部队列里放了一个任务。

之前已经创建好了的 Worker 线程会执行一个 while 循环 ---> 不停的从这个 内部队列里获取任务。(这一步是竞争的关系，都会抢着从队列里获取任务，由这个队列内部实现了线程安全。)

获取得到一个任务后，其实也就是拿到了一个 Runnable 对象(也就是 execute(Runnabletask) 这里所提交的任务)，接着执行这个 Runnable 的 run() 方法，而不是 start()，为什么线程池在调度的时候执行的是 Runnable 的 run() 方法，而不是 start() 方法呢？线程池这个场景中却恰好就是要利用它只是一个普通方法调用。如果我们往一个核心、最大线程数为 2 的线程池里丢了 1000 个任务，那么它会额外的创建 1000 个线程，同时每个任务都是异步执行的，一下子就执行完毕了。

所谓线程池本质是一个hashSet。多余的任务会放在阻塞队列中。

只有当阻塞队列满了后，才会触发非核心线程的创建。所以非核心线程只是临时过来打杂的。直到空闲了，然后自己关闭了。

线程池提供了两个钩子（beforeExecute，afterExecute）给我们，我们继承线程池，在执行任务前后做一些事情。

线程池原理关键技术：锁（lock,cas）、阻塞队列、hashSet（资源池）

当线程池在执行一个任务时抛出未捕获异常，这个任务还会执行吗？



右图放大后的那一块，也就是内部队列并没有其他线程往里边丢任务执行 execute() 方法。

而一旦发生未捕获的异常后， Worker1 被回收，顺带的它所调度的线程 task1（这个task1 也就是在执行一个 while 循环消费左图中的那个队列） 也会被回收掉。

新创建的 Worker2 会取代 Worker1 继续执行 while 循环从内部队列里获取任务，但此时这个队列就一直会是空的，所以也就是处于 Waiting 状态。

原文：<https://mp.weixin.qq.com/s/SpxYhfExN-GwTYflkhvKRg>

## executor与submit的区别

1、接收的参数不一样

2、submit有返回值，而execute没有

3、submit方便Exception处理：意思就是如果你在你的task里会抛出checked或者unchecked exception，而你又希望外面的调用者能够感知这些exception并做出及时的处理，那么就需要用到submit，通过捕获Future.get抛出的异常。

# 多线程

## 线程间通信

等待通知模式是 Java 中比较经典的线程通信方式，常用的等待通知模式有：

两个线程通过对同一对象调用等待 wait() 和通知 notify() 方法来进行通讯。

join() 方法，在 join 线程完成后会调用 notifyAll() 方法，是在 JVM 实现中调用。

采用共享内存的方式进行线程通信的类型有：

volatile 共享内存

CountDownLatch 并发工具：CountDownLatch 也是基于 AQS(AbstractQueuedSynchronizer) 实现的，更多实现参考 ReentrantLock 实现原理

线程池 awaitTermination() 方法：如果是用线程池来管理线程，可以使用以下方式来让主线程等待线程池中所有任务执行完毕:

管道通信 PipedWriter，PipedReader，writer.connect(reader)

## 一个线程OOM，进程里其他线程还能运行么？

答案是还能运行

在多线程环境下，每个线程拥有一个栈和一个程序计数器。栈和程序计数器用来保存线程的执行历史和线程的执行状态，是线程私有的资源。其他的资源（比如堆、地址空间、全局变量）是由同一个进程内的多个线程共享。

当一个线程抛出OOM异常后，即使是堆内存溢出，它所占据的内存资源会全部被释放掉，从而不会影响其他线程的运行！如果是栈溢出，结论也是一样。

JvisualVM看堆的变化

## 多线程有什么用？

（1）发挥多核CPU的优势

（2）防止阻塞

（3）便于建模：假设有一个大的任务A，单线程编程，那么就要考虑很多，建立整个程序模型比较麻烦。但是如果把这个大的任务A分解成几个小任务，任务B、任务C、任务D，分别建立程序模型，并通过多线程分别运行这几个任务，那就简单很多了。

## start()方法和run()方法的区别？

只有调用了start()方法，才会表现出多线程的特性，不同线程的run()方法里面的代码交替执行。如果只是调用run()方法，那么代码还是同步执行的，必须等待一个线程的run()方法里面的代码全部执行完毕之后，另外一个线程才可以执行其run()方法里面的代码。

## Runnable接口和Callable接口的区别？

Runnable接口中的run()方法的返回值是void，它做的事情只是纯粹地去执行run()方法中的代码而已；Callable接口中的call()方法是有返回值的，是一个泛型，和Future、FutureTask配合可以用来获取异步执行的结果。

## CyclicBarrier和CountDownLatch的区别？

两个看上去有点像的类，都在java.util.concurrent下，都可以用来表示代码运行到某个点上，二者的区别在于：

（1）CyclicBarrier的某个线程运行到某个点上之后，该线程即停止运行，直到所有的线程都到达了这个点，所有线程才重新运行；CountDownLatch则不是，某线程运行到某个点上之后，只是给某个数值-1而已，该线程继续运行

（2）CyclicBarrier只能唤起一个任务，CountDownLatch可以唤起多个任务

（3）CyclicBarrier可重用，CountDownLatch不可重用，计数值为0该CountDownLatch就不可再用了

## volatile/'vɒlətaɪl/关键字的作用？

1. **可见性**
2. **防止指令重排**

这里的可见性是什么意思呢？当一个线程修改了变量的值，新的值会立刻同步到主内存当中。而其他线程读取这个变量的时候，也会从主内存中拉取最新的变量值。

为什么volatile关键字可以有这样的特性？这得益于java语言的先行发生原则（happens-before）。

count++这一行代码本身并不是原子性操作

什么时候适合用volatile呢？

1.运行结果并不依赖变量的当前值，或者能够确保只有单一的线程修改变量的值。

2.变量不需要与其他的状态变量共同参与不变约束。

第一条很好理解，就是上面的代码例子。第二条是什么意思呢？可以看看下面这个场景：

volatile static int start = 3;

volatile static int end = 6;

线程A执行如下代码：

while (start < end){

//do something

}

线程B执行如下代码：

start+=3

end+=3;

这种情况下，一旦在线程A的循环中执行了线程B，start有可能先更新成6，造成了一瞬间 start == end，从而跳出while循环的可能性。从实践角度而言，volatile的一个重要作用就是和CAS结合，保证了原子性，详细的可以参见 java.util.concurrent.atomic 包下的类，比如 AtomicInteger。

## 什么是线程安全

如果你的代码在多线程下执行和在单线程下执行永远都能获得一样的结果，那么你的代码就是线程安全的。

这个问题有值得一提的地方，就是线程安全也是有几个级别的：

（1）不可变

像String、Integer、Long这些，都是final类型的类，任何一个线程都改变不了它们的值，要改变除非新创建一个，因此这些不可变对象不需要任何同步手段就可以直接在多线程环境下使用

（2）绝对线程安全

不管运行时环境如何，调用者都不需要额外的同步措施。要做到这一点通常需要付出许多额外的代价，Java中标注自己是线程安全的类，实际上绝大多数都不是线程安全的，不过绝对线程安全的类，Java中也有，比方说CopyOnWriteArrayList、CopyOnWriteArraySet

（3）相对线程安全

相对线程安全也就是我们通常意义上所说的线程安全，像Vector这种，add、remove方法都是原子操作，不会被打断，但也仅限于此，如果有个线程在遍历某个Vector、有个线程同时在add这个Vector，99%的情况下都会出现ConcurrentModificationException，也就是fail-fast机制。

（4）线程非安全

这个就没什么好说的了，ArrayList、LinkedList、HashMap等都是线程非安全的类

## Java中如何获取到线程dump文件

死循环、死锁、阻塞、页面打开慢等问题，打线程dump是最好的解决问题的途径。所谓线程dump也就是线程堆栈，获取到线程堆栈有两步：

（1）获取到线程的pid，可以通过使用jps命令，在Linux环境下还可以使用ps -ef | grep java

（2）打印线程堆栈，可以通过使用jstack pid命令，在Linux环境下还可以使用kill -3 pid

另外提一点，Thread类提供了一个getStackTrace()方法也可以用于获取线程堆栈。这是一个实例方法，因此此方法是和具体线程实例绑定的，每次获取获取到的是具体某个线程当前运行的堆栈，

## 一个线程如果出现了运行时异常会怎么样

如果这个异常没有被捕获的话，这个线程就停止执行了。另外重要的一点是：如果这个线程持有某个某个对象的监视器，那么这个对象监视器会被立即释放

## 如何在两个线程之间共享数据

通过在线程之间共享对象就可以了，然后通过wait/notify/notifyAll、await/signal/signalAll进行唤起和等待，比方说阻塞队列BlockingQueue就是为线程之间共享数据而设计的。

## sleep方法和wait方法有什么区别

这个问题常问，sleep方法和wait方法都可以用来放弃CPU一定的时间，不同点在于如果线程持有某个对象的监视器，sleep方法不会放弃这个对象的监视器，wait方法会放弃这个对象的监视器

## 生产者消费者模型的作用是什么

（1）通过平衡生产者的生产能力和消费者的消费能力来提升整个系统的运行效率，这是生产者消费者模型最重要的作用

（2）解耦，这是生产者消费者模型附带的作用，解耦意味着生产者和消费者之间的联系少，联系越少越可以独自发展而不需要收到相互的制约

## ThreadLocal有什么用

简单说ThreadLocal就是一种以空间换时间的做法，在每个Thread里面维护了一个以开地址法实现的ThreadLocal.ThreadLocalMap，把数据进行隔离，数据不共享，自然就没有线程安全方面的问题了

## 为什么wait()方法和notify()/notifyAll()方法要在同步块中被调用

这是JDK强制的，wait()方法和notify()/notifyAll()方法在调用前都必须先获得对象的锁

wait()方法和notify()/notifyAll()方法在放弃对象监视器时有什么区别？

wait()方法和notify()/notifyAll()方法在放弃对象监视器的时候的区别在于：wait()方法立即释放对象监视器，notify()/notifyAll()方法则会等待线程剩余代码执行完毕才会放弃对象监视器。

## 为什么要使用线程池

避免频繁地创建和销毁线程，达到线程对象的重用。另外，使用线程池还可以根据项目灵活地控制并发的数目。

## 怎么检测一个线程是否持有对象监视器

Thread类提供了一个holdsLock(Object obj)方法，当且仅当对象obj的监视器被某条线程持有的时候才会返回true，注意这是一个static方法，这意味着"某条线程"指的是当前线程。

## synchronized和ReentrantLock的区别

synchronized是和if、else、for、while一样的关键字，ReentrantLock是类，这是二者的本质区别。

ReentrantLock比synchronized的扩展性体现在几点上：

（1）ReentrantLock可以对获取锁的等待时间进行设置，这样就避免了死锁

（2）ReentrantLock可以获取各种锁的信息

（3）ReentrantLock可以灵活地实现多路通知

另外，二者的锁机制其实也是不一样的。ReentrantLock底层调用的是Unsafe的park方法加锁，synchronized操作的应该是对象头中mark word，这点我不能确定。

## ConcurrentHashMap的并发度是什么

ConcurrentHashMap的并发度就是segment的大小，默认为16，这意味着最多同时可以有16条线程操作ConcurrentHashMap

首先明确一下，不是说ReentrantLock不好，只是ReentrantLock某些时候有局限。如果使用ReentrantLock，可能本身是为了防止线程A在写数据、线程B在读数据造成的数据不一致，但这样，如果线程C在读数据、线程D也在读数据，读数据是不会改变数据的，没有必要加锁，但是还是加锁了，降低了程序的性能。

因为这个，才诞生了读写锁ReadWriteLock。ReadWriteLock是一个读写锁接口，ReentrantReadWriteLock是ReadWriteLock接口的一个具体实现，实现了读写的分离，读锁是共享的，写锁是独占的，读和读之间不会互斥，读和写、写和读、写和写之间才会互斥，提升了读写的性能。

## FutureTask是什么

FutureTask表示一个异步运算的任务。FutureTask里面可以传入一个Callable的具体实现类，可以对这个异步运算的任务的结果进行等待获取、判断是否已经完成、取消任务等操作。当然，由于FutureTask也是Runnable接口的实现类，所以FutureTask也可以放入线程池中。

## Linux环境下如何查找哪个线程使用CPU最长

（1）获取项目的pid，jps或者ps -ef | grep java，这个前面有讲过

（2）top -H -p pid，顺序不能改变

这样就可以打印出当前的项目，每条线程占用CPU时间的百分比。注意这里打出的是LWP，也就是操作系统原生线程的线程号

使用"top -H -p pid"+"jps pid"可以很容易地找到某条占用CPU高的线程的线程堆栈，从而定位占用CPU高的原因，一般是因为不当的代码操作导致了死循环。

最后提一点，"top -H -p pid"打出来的LWP是十进制的，"jps pid"打出来的本地线程号是十六进制的，转换一下，就能定位到占用CPU高的线程的当前线程堆栈了。

## 怎么唤醒一个阻塞的线程

如果线程是因为调用了wait()、sleep()或者join()方法而导致的阻塞，可以中断线程，并且通过抛出InterruptedException来唤醒它；如果线程遇到了IO阻塞，无能为力，因为IO是操作系统实现的，Java代码并没有办法直接接触到操作系统。

## 不可变对象对多线程有什么帮助

前面有提到过的一个问题，不可变对象保证了对象的内存可见性，对不可变对象的读取不需要进行额外的同步手段，提升了代码执行效率。

## 什么是多线程的上下文切换

多线程的上下文切换是指CPU控制权由一个已经正在运行的线程切换到另外一个就绪并等待获取CPU执行权的线程的过程。

## 如果你提交任务时，线程池队列已满，这时会发生什么

如果使用的是无界队列LinkedBlockingQueue，也就是无界队列的话，没关系，继续添加任务到阻塞队列中等待执行，因为LinkedBlockingQueue可以近乎认为是一个无穷大的队列，可以无限存放任务

如果使用的是有界队列比如ArrayBlockingQueue，任务首先会被添加到ArrayBlockingQueue中，ArrayBlockingQueue满了，会根据maximumPoolSize的值增加线程数量，如果增加了线程数量还是处理不过来，ArrayBlockingQueue继续满，那么则会使用拒绝策略RejectedExecutionHandler处理满了的任务，默认是AbortPolicy

## Java中用到的线程调度算法是什么

抢占式。一个线程用完CPU之后，操作系统会根据线程优先级、线程饥饿情况等数据算出一个总的优先级并分配下一个时间片给某个线程执行。

## Thread.sleep(0)的作用是什么

为了让某些优先级比较低的线程也能获取到CPU控制权，可以使用Thread.sleep(0)手动触发一次操作系统分配时间片的操作，这也是平衡CPU控制权的一种操作。

## 什么是自旋

很多synchronized里面的代码只是一些很简单的代码，执行时间非常快，此时等待的线程都加锁可能是一种不太值得的操作，因为线程阻塞涉及到用户态和内核态切换的问题。既然synchronized里面的代码执行得非常快，不妨让等待锁的线程不要被阻塞，而是在synchronized的边界做忙循环，这就是自旋。如果做了多次忙循环发现还没有获得锁，再阻塞，这样可能是

## 什么是Java内存模型

Java内存模型定义了一种多线程访问Java内存的规范。

（1）Java内存模型将内存分为了主内存和工作内存。类的状态，也就是类之间共享的变量，是存储在主内存中的，每次Java线程用到这些主内存中的变量的时候，会读一次主内存中的变量，并让这些内存在自己的工作内存中有一份拷贝，运行自己线程代码的时候，用到这些变量，操作的都是自己工作内存中的那一份。在线程代码执行完毕之后，会将最新的值更新到主内存中去

（2）定义了几个原子操作，用于操作主内存和工作内存中的变量

（3）定义了volatile变量的使用规则

（4）happens-before，即先行发生原则，定义了操作A必然先行发生于操作B的一些规则，比如在同一个线程内控制流前面的代码一定先行发生于控制流后面的代码、一个释放锁unlock的动作一定先行发生于后面对于同一个锁进行锁定lock的动作等等，只要符合这些规则，则不需要额外做同步措施，如果某段代码不符合所有的happens-before规则，则这段代码一定是线程非安全的

## 什么是CAS

CAS，全称为Compare and Swap，即比较-替换。假设有三个操作数：内存值V、旧的预期值A、要修改的值B，当且仅当预期值A和内存值V相同时，才会将内存值修改为B并返回true，否则什么都不做并返回false。当然CAS一定要volatile变量配合，这样才能保证每次拿到的变量是主内存中最新的那个值，否则旧的预期值A对某条线程来说，永远是一个不会变的值A，只要某次CAS操作失败，永远都不可能成功。

## 什么是乐观锁和悲观锁

（1）乐观锁：就像它的名字一样，对于并发间操作产生的线程安全问题持乐观状态，乐观锁认为竞争不总是会发生，因此它不需要持有锁，将比较-替换这两个动作作为一个原子操作尝试去修改内存中的变量，如果失败则表示发生冲突，那么就应该有相应的重试逻辑。

（2）悲观锁：还是像它的名字一样，对于并发间操作产生的线程安全问题持悲观状态，悲观锁认为竞争总是会发生，因此每次对某资源进行操作时，都会持有一个独占的锁，就像synchronized，不管三七二十一，直接上了锁就操作资源了。

## 什么是AQS

简单说一下AQS，AQS全称为AbstractQueuedSychronizer，翻译过来应该是抽象队列同步器。

如果说java.util.concurrent的基础是CAS的话，那么AQS就是整个Java并发包的核心了，ReentrantLock、CountDownLatch、Semaphore等等都用到了它。AQS实际上以双向队列的形式连接所有的Entry，比方说ReentrantLock，所有等待的线程都被放在一个Entry中并连成双向队列，前面一个线程使用ReentrantLock好了，则双向队列实际上的第一个Entry开始运行。

AQS定义了对双向队列所有的操作，而只开放了tryLock和tryRelease方法给开发者使用，开发者可以根据自己的实现重写tryLock和tryRelease方法，以实现自己的并发功能。

## 单例模式的线程安全性

首先要说的是单例模式的线程安全意味着：某个类的实例在多线程环境下只会被创建一次出来。单例模式有很多种的写法，我总结一下：

（1）饿汉式单例模式的写法：线程安全

（2）懒汉式单例模式的写法：非线程安全

（3）双检锁单例模式的写法：线程安全

## Semaphore有什么作用

Semaphore就是一个信号量，它的作用是限制某段代码块的并发数。Semaphore有一个构造函数，可以传入一个int型整数n，表示某段代码最多只有n个线程可以访问，如果超出了n，那么请等待，等到某个线程执行完毕这段代码块，下一个线程再进入。由此可以看出如果Semaphore构造函数中传入的int型整数n=1，相当于变成了一个synchronized了。

## Hashtable的size()方法中明明只有一条语句"return count"，为什么还要做同步

（1）同一时间只能有一条线程执行固定类的同步方法，但是对于类的非同步方法，可以多条线程同时访问。所以，这样就有问题了，可能线程A在执行Hashtable的put方法添加数据，线程B则可以正常调用size()方法读取Hashtable中当前元素的个数，那读取到的值可能不是最新的，可能线程A添加了完了数据，但是没有对size++，线程B就已经读取size了，那么对于线程B来说读取到的size一定是不准确的。而给size()方法加了同步之后，意味着线程B调用size()方法只有在线程A调用put方法完毕之后才可以调用，这样就保证了线程安全性

（2）CPU执行代码，执行的不是Java代码，这点很关键，一定得记住。Java代码最终是被翻译成机器码执行的，机器码才是真正可以和硬件电路交互的代码。即使你看到Java代码只有一行，甚至你看到Java代码编译之后生成的字节码也只有一行，也不意味着对于底层来说这句语句的操作只有一个。一句"return count"假设被翻译成了三句汇编语句执行，一句汇编语句和其机器码做对应，完全可能执行完第一句，线程就切换了。

## 线程类的构造方法、静态块是被哪个线程调用的

线程类的构造方法、静态块是被new这个线程类所在的线程所调用的，而run方法里面的代码才是被线程自身所调用的。

如果说上面的说法让你感到困惑，那么我举个例子，假设Thread2中new了Thread1，main函数中new了Thread2，那么：

（1）Thread2的构造方法、静态块是main线程调用的，Thread2的run()方法是Thread2自己调用的

（2）Thread1的构造方法、静态块是Thread2调用的，Thread1的run()方法是Thread1自己调用的

## 同步方法和同步块，哪个是更好的选择

同步块，这意味着同步块之外的代码是异步执行的，这比同步整个方法更提升代码的效率。请知道一条原则：同步的范围越小越好。

虽说同步的范围越少越好，但是在Java虚拟机中还是存在着一种叫做锁粗化的优化方法，这种方法就是把同步范围变大。这是有用的，比方说StringBuffer，它是一个线程安全的类，自然最常用的append()方法是一个同步方法，我们写代码的时候会反复append字符串，这意味着要进行反复的加锁->解锁，这对性能不利，因为这意味着Java虚拟机在这条线程上要反复地在内核态和用户态之间进行切换，因此Java虚拟机会将多次append方法调用的代码进行一个锁粗化的操作，将多次的append的操作扩展到append方法的头尾，变成一个大的同步块，这样就减少了加锁-->解锁的次数，有效地提升了代码执行的效率。

高并发、任务执行时间短的业务怎样使用线程池？并发不高、任务执行时间长的业务怎样使用线程池？并发高、业务执行时间长的业务怎样使用线程池？

（1）高并发、任务执行时间短的业务，线程池线程数可以设置为CPU核数+1，减少线程上下文的切换

（2）并发不高、任务执行时间长的业务要区分开看：

a）假如是业务时间长集中在IO操作上，也就是IO密集型的任务，因为IO操作并不占用CPU，所以不要让所有的CPU闲下来，可以加大线程池中的线程数目，让CPU处理更多的业务

b）假如是业务时间长集中在计算操作上，也就是计算密集型任务，这个就没办法了，和（1）一样吧，线程池中的线程数设置得少一些，减少线程上下文的切换

（3）并发高、业务执行时间长，解决这种类型任务的关键不在于线程池而在于整体架构的设计，看看这些业务里面某些数据是否能做缓存是第一步，增加服务器是第二步，至于线程池的设置，设置参考其他有关线程池的文章。最后，业务执行时间长的问题，也可能需要分析一下，看看能不能使用中间件对任务进行拆分和解耦。

## java程序，主进程需要等待多个子进程结束之后再执行后续的代码，有哪些方案可以实现

**1.join方法**

比如在线程B中调用了线程A的join()方法，直到线程A执行完毕后，才会继续执行线程B。

public static void main(String[] args) throws InterruptedException {

Vector<Thread> vector = new Vector<>();

for(int i=0;i<5;i++) {

Thread childThread= new Thread(new Runnable() {

@Override

public void run() {

// TODO Auto-generated method stub

try {

Thread.sleep(1000);

} catch (InterruptedException e) {

// TODO Auto-generated catch block

e.printStackTrace();

}

System.out.println("子线程被执行");

}

});

vector.add(childThread);

childThread.start();

}

for(Thread thread : vector) {

thread.join();

}

System.out.println("主线程被执行");

}

**2.等待多线程完成的CountDownLatch**

public static void main(String[] args) throws InterruptedException {

final CountDownLatch latch = new CountDownLatch(5);

for(int i=0;i<5;i++) {

Thread childThread= new Thread(new Runnable() {

@Override

public void run() {

// TODO Auto-generated method stub

try {

Thread.sleep(1000);

} catch (InterruptedException e) {

// TODO Auto-generated catch block

e.printStackTrace();

}

System.out.println("子线程被执行");

latch.countDown();

}

});

childThread.start();

}

latch.await();//阻塞当前线程直到latch中的值

System.out.println("主线程被执行");

}

**3.同步屏障CyclicBarrier**

public static void main(String[] args) throws InterruptedException, BrokenBarrierException {

final CyclicBarrier barrier = new CyclicBarrier(5);

for(int i=0;i<4;i++) {

Thread childThread= new Thread(new Runnable() {

@Override

public void run() {

// TODO Auto-generated method stub

try {

Thread.sleep(1000);

} catch (InterruptedException e) {

// TODO Auto-generated catch block

e.printStackTrace();

}

System.out.println("子线程被执行");

try {

barrier.await();

} catch (InterruptedException e) {

// TODO Auto-generated catch block

e.printStackTrace();

} catch (BrokenBarrierException e) {

// TODO Auto-generated catch block

e.printStackTrace();

}

}

});

childThread.start();

}

barrier.await();//阻塞当前线程直到latch中的值

System.out.println("主线程被执行");

}

CylicBarrier是控制一组线程的同步，初始化的参数：5的含义是包括主线程在内有5个线程，所以只能有四个子线程，这与CountDownLatch是不一样的。

countDownLatch和cyclicBarrier有什么区别呢，他们的区别：countDownLatch只能使用一次，而CyclicBarrier方法可以使用reset()方法重置，所以CyclicBarrier方法可以能处理更为复杂的业务场景。cyclicBarrier只会在最后一个线程结束时才会给主线程发送信号，而countDownLatch会在每个线程结束时都发送一个信号

5.FutureTast可用于闭锁，类似于CountDownLatch的作用

public class Test5 {

public static void main(String[] args) {

MyThread td = new MyThread();

//1.执行 Callable 方式，需要 FutureTask 实现类的支持，用于接收运算结果。

FutureTask<Integer> result1 = new FutureTask<>(td);

new Thread(result1).start();

FutureTask<Integer> result2 = new FutureTask<>(td);

new Thread(result2).start();

FutureTask<Integer> result3 = new FutureTask<>(td);

new Thread(result3).start();

Integer sum;

try {

sum = result1.get();

sum = result2.get();

sum = result3.get();

//这里获取三个sum值只是为了同步，并没有实际意义

System.out.println(sum);

} catch (InterruptedException e) {

// TODO Auto-generated catch block

e.printStackTrace();

} catch (ExecutionException e) {

// TODO Auto-generated catch block

e.printStackTrace();

} //FutureTask 可用于 闭锁 类似于CountDownLatch的作用，在所有的线程没有执行完成之后这里是不会执行的

System.out.println("主线程被执行");

}

}

class MyThread implements Callable<Integer> {

@Override

public Integer call() throws Exception {

int sum = 0;

Thread.sleep(1000);

for (int i = 0; i <= 10; i++) {

sum += i;

}

System.out.println("子线程被执行");

return sum;

}

}

**6.使用callable+future**

Callable+Future最终也是以Callable+FutureTask的形式实现的。 在这种方式中调用了： Future future = executor.submit(task);

public class Test6 {

public static void main(String[] args) throws InterruptedException, ExecutionException {

ExecutorService executor = Executors.newCachedThreadPool();

Task task = new Task();

Future<Integer> future1 = executor.submit(task);

Future<Integer> future2 = executor.submit(task);

//获取线程执行结果，用来同步

Integer result1 = future1.get();

Integer result2 = future2.get();

System.out.println("主线程执行");

executor.shutdown();

}

}

class Task implements Callable<Integer>{

@Override public Integer call() throws Exception {

int sum = 0;

//do something;

System.out.println("子线程被执行");

return sum;

}

}

1）CountDownLatch和CyclicBarrier都能够实现线程之间的等待，只不过它们侧重点不同：CountDownLatch一般用于某个线程A等待若干个其他线程执行完任务之后，它才执行；而CyclicBarrier一般用于一组线程互相等待至某个状态，然后这一组线程再同时执行；另外，CountDownLatch是不能够重用的，而CyclicBarrier是可以重用的。

2）Semaphore其实和锁有点类似，它一般用于控制对某组资源的访问权限。CountDownLatch类实际上是使用计数器的方式去控制的，不难想象当我们初始化CountDownLatch的时候传入了一个int变量这个时候在类的内部初始化一个int的变量，每当我们调用countDownt()方法的时候就使得这个变量的值减1，而对于await()方法则去判断这个int的变量的值是否为0，是则表示所有的操作都已经完成，否则继续等待。

实际上如果了解AQS的话应该很容易想到可以使用AQS的共享式获取同步状态的方式来完成这个功能。而CountDownLatch实际上也就是这么做的。

## 常用的并发编程类

CountDownLatch

CyclicBarrier

阻塞,直到最后一个线程达到屏障时,程序才会继续执行

Semaphore

Semaphore可以维护当前访问自身的线程个数

Exchanger

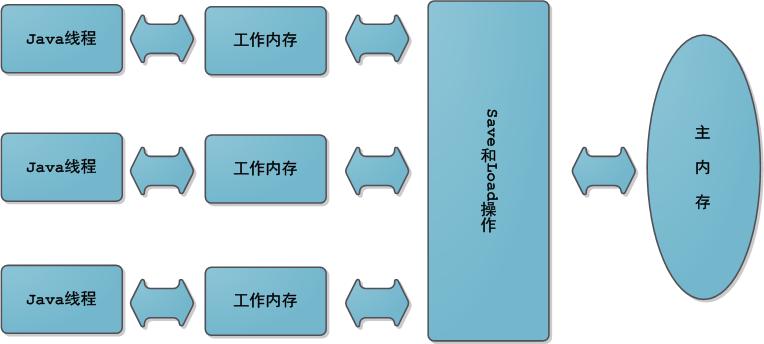
exchanger是一个用于线程间协作的工具类,Exchanger用于进行线程间的数据交换,它提供一个同步点,在这个同步点,二个线程可以交换彼此的数据.在这二个线程通过exchange方法交换数据,如果第一个线程先执行exchange()方法,它会一直等待第二个线程也执行exchange方法,当二个线程都达到同步点时,这二个线程就可以交换数据,将本线程生产出来的数据传递给对方.

Java 5.0及以上版本中，线程的状态被扩充为新建、可运行、阻塞、等待、定时等待、死亡六种。

## java内存模型

Java内存模型的主要目标是定义程序中各个变量的访问规则，即在虚拟机中将变量存储到内存和从内存中取出变量这样底层细节。此处的变量与Java编程时所说的变量不一样，指包括了实例字段、静态字段和构成数组对象的元素，但是不包括局部变量与方法参数，后者是线程私有的，不会被共享。

　　Java内存模型中规定了所有的变量都存储在主内存中，每条线程还有自己的工作内存（可以与前面将的处理器的高速缓存类比），线程的工作内存中保存了该线程使用到的变量到主内存副本拷贝，线程对变量的所有操作（读取、赋值）都必须在工作内存中进行，而不能直接读写主内存中的变量。不同线程之间无法直接访问对方工作内存中的变量，线程间变量值的传递均需要在主内存来完成，线程、主内存和工作内存的交互关系如下图所示，和上图很类似



内存间交互操作

　　关于主内存与工作内存之间的具体交互协议，即一个变量如何从主内存拷贝到工作内存、如何从工作内存同步到主内存之间的实现细节，Java内存模型定义了以下八种操作来完成：

* lock（锁定）：作用于主内存的变量，把一个变量标识为一条线程独占状态。
* unlock（解锁）：作用于主内存变量，把一个处于锁定状态的变量释放出来，释放后的变量才可以被其他线程锁定。
* read（读取）：作用于主内存变量，把一个变量值从主内存传输到线程的工作内存中，以便随后的load动作使用
* load（载入）：作用于工作内存的变量，它把read操作从主内存中得到的变量值放入工作内存的变量副本中。
* use（使用）：作用于工作内存的变量，把工作内存中的一个变量值传递给执行引擎，每当虚拟机遇到一个需要使用变量的值的字节码指令时将会执行这个操作。
* assign（赋值）：作用于工作内存的变量，它把一个从执行引擎接收到的值赋值给工作内存的变量，每当虚拟机遇到一个给变量赋值的字节码指令时执行这个操作。
* store（存储）：作用于工作内存的变量，把工作内存中的一个变量的值传送到主内存中，以便随后的write的操作。
* write（写入）：作用于主内存的变量，它把store操作从工作内存中一个变量的值传送到主内存的变量中。

Java内存模型只要求上述操作必须按顺序执行，而没有保证必须是连续执行。也就是read和load之间，store和write之间是可以插入其他指令的，如对主内存中的变量a、b进行访问时，可能的顺序是read a，read b，load b， load a。Java内存模型还规定了在执行上述八种基本操作时，必须满足如下规则：

* 不允许read和load、store和write操作之一单独出现
* 不允许一个线程丢弃它的最近assign的操作，即变量在工作内存中改变了之后必须同步到主内存中。
* 不允许一个线程无原因地（没有发生过任何assign操作）把数据从工作内存同步回主内存中。
* 一个新的变量只能在主内存中诞生，不允许在工作内存中直接使用一个未被初始化（load或assign）的变量。即就是对一个变量实施use和store操作之前，必须先执行过了assign和load操作。
* 一个变量在同一时刻只允许一条线程对其进行lock操作，lock和unlock必须成对出现
* 如果对一个变量执行lock操作，将会清空工作内存中此变量的值，在执行引擎使用这个变量前需要重新执行load或assign操作初始化变量的值
* 如果一个变量事先没有被lock操作锁定，则不允许对它执行unlock操作；也不允许去unlock一个被其他线程锁定的变量。
* 对一个变量执行unlock操作之前，必须先把此变量同步到主内存中（执行store和write操作）。

## 其它

指令重排

public void run() {

a = 1;

x = b;

}

可能x=b先执行

Java中什么是竞态条件？

多线程对一些资源的竞争的时候就会产生竞态条件，如果首先要执行的程序竞争失败排到后面执行了，那么整个程序就会出现一些不确定的bugs。这种bugs很难发现而且会重复出现，因为线程间的随机竞争。

**一个线程运行时发生异常会怎样？**

简单的说，如果异常没有被捕获该线程将会停止执行

**Java中interrupted 和 isInterruptedd方法的区别？**

interrupted() 和 isInterrupted()的主要区别是前者会将中断状态清除而后者不会。Java多线程的中断机制是用内部标识来实现的，调用Thread.interrupt()来中断一个线程就会设置中断标识为true。**当中断线程调用静态方式Thread.interrupted()来检查中断状态时，中断状态会被清零。而非静态方法isInterrupted()用来查询其它线程的中断状态且不会改变中断状态标识。**简单的说就是任何抛出InterruptedException异常的方法都会将中断状态清零。无论如何，一个线程的中断状态有有可能被其它线程调用中断来改变

**为什么wait和notify方法要在同步块中调用？**

一个原因是为了避免wait和notify之间产生竞态条件。还有一个原因就是wait,notify都要获取对象的锁

**如果你提交任务时，线程池队列已满。会时发会生什么？**

这个问题问得很狡猾，许多程序员会认为该任务会阻塞直到线程池队列有空位。事实上如果一个任务不能被调度执行那么ThreadPoolExecutor’s submit()方法将会抛出一个RejectedExecutionException异常。

**Java线程池中submit() 和 execute()方法有什么区别？**

两个方法都可以向线程池提交任务，execute()方法的返回类型是void，它定义在Executor接口中, 而submit()方法可以返回持有计算结果的Future对象，它定义在ExecutorService接口中，它扩展了Executor接口，其它线程池类像ThreadPoolExecutor和ScheduledThreadPoolExecutor都有这些方法。

当我们考察异常处理的时候，又会发现另外一个不同。当你使用execute提交的任务抛出异常时，此异常将会交由未捕捉异常处理过程来处理(uncaught exception handler)，当你没有显式指定一个异常处理器的话，默认情况下仅仅会通过System.err打印出错误堆栈。当你用submit来提交一个任务的时候，这个任务一旦抛出异常(无论是否是运行时异常)，那这个异常是任务返回对象的一部分。对这样一种情形，当你调用Future.get()方法的时候，这个方法会重新抛出这个异常，并且会使用ExecutionException进行包装。

**CAS缺陷**

1. ABA问题。因为CAS需要在操作值的时候检查下值有没有发生变化，如果没有发生变化则更新，但是如果一个值原来是A，变成了B，又变成了A，那么使用CAS进行检查时会发现它的值没有发生变化，但是实际上却变化了。ABA问题的解决思路就是使用版本号。在变量前面追加上版本号，每次变量更新的时候把版本号加一，那么A－B－A 就会变成1A-2B－3A。

从Java1.5开始JDK的atomic包里提供了一个类AtomicStampedReference来解决ABA问题。这个类的compareAndSet方法作用是首先检查当前引用是否等于预期引用，并且当前标志是否等于预期标志，如果全部相等，则以原子方式将该引用和该标志的值设置为给定的更新值。

2. 循环时间长开销大。自旋CAS如果长时间不成功，会给CPU带来非常大的执行开销。如果JVM能支持处理器提供的pause指令那么效率会有一定的提升，pause指令有两个作用，第一它可以延迟流水线执行指令（de-pipeline）,使CPU不会消耗过多的执行资源，延迟的时间取决于具体实现的版本，在一些处理器上延迟时间是零。第二它可以避免在退出循环的时候因内存顺序冲突（memory order violation）而引起CPU流水线被清空（CPU pipeline flush），从而提高CPU的执行效率。

3. 只能保证一个共享变量的原子操作。当对一个共享变量执行操作时，我们可以使用循环CAS的方式来保证原子操作，但是对多个共享变量操作时，循环CAS就无法保证操作的原子性，这个时候就可以用锁，或者有一个取巧的办法，就是把多个共享变量合并成一个共享变量来操作。比如有两个共享变量i＝2,j=a，合并一下ij=2a，然后用CAS来操作ij。从Java1.5开始JDK提供了AtomicReference类来保证引用对象之间的原子性，你可以把多个变量放在一个对象里来进行CAS操作。

**多线程就一定效率高吗？**

当然不一定，不并行的话效率肯定低

**并发编程需要注意的2个问题**

并发编程尤其需要注意的是共享数据的安全性和锁性能方面的问题

锁竟争带来的程序效率问题