

目录

📄 Python教程

Python简介

安装Python

第一个Python程序

Python基础

函数

高级特性

函数式编程

模块

面向对象编程

面向对象高级编程

错误、调试和测试

IO编程

进程和线程

多线程

ThreadLocal

进程 vs 线程

分布式进程

正则表达式

常用内建模块

常用第三方模块

virtualenv

图形界面

网络编程

电子邮件

访问数据库

Web开发

异步IO

实战

FAQ

期末总结

关于作者

 廖雪峰

北京 朝阳区

+ 加关注

廖雪峰

自己的Python课程

Python商业爬虫全解码

让天下没有爬不到的数据!

Python爬虫

+

数据分析

Python机器学习

+

深度学习

.....

找廖雪峰老师

廖雪峰老师

自己的Java课程

Java高级架构师

更专业

更权威

源码分析专题

+

微服务架构专题

高并发分布式专题

+

性能优化专题

.....

找廖雪峰老师

unicef

联合国儿童基金会



疾病对贫困家庭来说

是雪上加霜

让他不再无助 >>

unicef

联合国儿童基金会



疾病对贫困家庭来说

是雪上加霜

让他不再无助 >>

## 多线程

阅读: 382577

多任务可以由多进程完成，也可以由一个进程内的多线程完成。

我们前面提到了进程是由若干线程组成的，一个进程至少有一个线程。

由于线程是操作系统直接支持的执行单元，因此，高级语言通常都内置多线程的支持，Python也不例外，并且，Python的线程是真正的Posix Thread，而不是模拟出来的线程。

Python的标准库提供了两个模块：`_thread` 和 `threading`，`_thread` 是低级模块，`threading` 是高级模块，对 `_thread` 进行了封装。绝大多数情况下，我们只需要使用 `threading` 这个高级模块。

启动一个线程就是把一个函数传入并创建 `Thread` 实例，然后调用 `start()` 开始执行：

```
import time, threading

# 新线程执行的代码:
def loop():
    print('thread %s is running...' % threading.current_thread().name)
    n = 0
    while n < 5:
        n = n + 1
        print('thread %s >>> %s' % (threading.current_thread().name, n))
        time.sleep(1)
    print('thread %s ended.' % threading.current_thread().name)

print('thread %s is running...' % threading.current_thread().name)
t = threading.Thread(target=loop, name='LoopThread')
t.start()
t.join()
print('thread %s ended.' % threading.current_thread().name)
```

执行结果如下：

```
thread MainThread is running...
thread LoopThread is running...
thread LoopThread >>> 1
thread LoopThread >>> 2
thread LoopThread >>> 3
thread LoopThread >>> 4
thread LoopThread >>> 5
thread LoopThread ended.
thread MainThread ended.
```

由于任何进程默认就会启动一个线程，我们把该线程称为主线程，主线程又可以启动新的线程，Python的 `threading` 模块有个 `current_thread()` 函数，它永远返回当前线程的实例。主线程实例的名字叫 `MainThread`，子线程的名字在创建时指定，我们用 `LoopThread` 命名子线程。名字仅仅在打印时用来显示，完全没有其他意义，如果不起名字Python就自动给线程命名为 `Thread-1`，`Thread-2`.....

## Lock

多线程和多进程最大的不同在于，多进程中，同一个变量，各自有一份拷贝存在于每个进程中，互不影响，而多线程中，所有变量都由所有线程共享，所以，任何一个变量都可以被任何一个线程修改，因此，线程之间共享数据最大的危险在于多个线程同时改一个变量，把内容给改乱了。

来看看多个线程同时操作一个变量怎么把内容给改乱了：

```
import time, threading

# 假定这是你的银行存款:
balance = 0

def change_it(n):
    # 先存后取，结果应该为0:
    global balance
    balance = balance + n
    balance = balance - n

def run_thread(n):
    for i in range(100000):
        change_it(n)

t1 = threading.Thread(target=run_thread, args=(5,))
t2 = threading.Thread(target=run_thread, args=(8,))
t1.start()
t2.start()
t1.join()
t2.join()
print(balance)
```

我们定义了一个共享变量 `balance`，初始值为 `0`，并且启动两个线程，先存后取，理论上结果应该为 `0`，但是，由于线程的调度是由操作系统决定的，当 `t1`、`t2` 交替执行时，只要循环次数足够多，`balance` 的结果就不一定是 `0` 了。

原因是因为高级语言的一条语句在CPU执行时是若干条语句，即使一个简单的计算：

```
balance = balance + n
```

也分两步：

1. 计算 `balance + n`，存入临时变量中；
2. 将临时变量的值赋给 `balance`。

也就是可以看成：

```
x = balance + n
balance = x
```

由于 `x` 是局部变量，两个线程各自都有自己的 `x`，当代码正常执行时：

```
初始值 balance = 0

t1: x1 = balance + 5 # x1 = 0 + 5 = 5
t1: balance = x1 # balance = 5
t1: x1 = balance - 5 # x1 = 5 - 5 = 0
t1: balance = x1 # balance = 0

t2: x2 = balance + 8 # x2 = 0 + 8 = 8
t2: balance = x2 # balance = 8
t2: x2 = balance - 8 # x2 = 8 - 8 = 0
t2: balance = x2 # balance = 0

结果 balance = 0
```

但是 `t1` 和 `t2` 是交替运行的，如果操作系统以下的顺序执行 `t1`、`t2`：

```
初始值 balance = 0

t1: x1 = balance + 5 # x1 = 0 + 5 = 5
t2: x2 = balance + 8 # x2 = 0 + 8 = 8
t2: balance = x2 # balance = 8

t1: balance = x1 # balance = 5
t1: x1 = balance + 5 # x1 = 5 + 5 = 10
t1: balance = x1 # balance = 10

t2: x2 = balance - 8 # x2 = 10 - 8 = 2
t2: balance = x2 # balance = 2

结果 balance = 2
```

究其原因，是因为修改 `balance` 需要多条语句，而执行这几条语句时，线程可能中断，从而导致多个线程把同一个对象的内容改乱了。

两个线程同时一存一取，就可能造成余额不对，你肯定不希望你的银行存款莫名其妙地变成了负数，所以，我们必须确保一个线程在修改 `balance` 的时候，别的线程一定不能改。

如果我们要确保 `balance` 计算正确，就要给 `change_it()` 上一把锁，当某个线程开始执行 `change_it()` 时，我们说，该线程因为获得了锁，因此其他线程不能同时执行 `change_it()`，只能等待，直到锁被释放后，获得该锁以后才能改。由于锁只有一个，无论多少线程，同一时刻最多只有一个线程持有该锁，所以，不会造成修改的冲突。创建一个锁就是通过 `threading.Lock()` 来实现：

```
balance = 0
lock = threading.Lock()

def run_thread(n):
    for i in range(100000):
        # 先获取锁:
        lock.acquire()
        try:
            # 放心地改吧:
            change_it(n)
        finally:
            # 改完了一定要释放锁:
            lock.release()
```

当多个线程同时执行 `lock.acquire()` 时，只有一个线程能成功地获取锁，然后继续执行代码，其他线程就继续等待直到获得锁为止。

获得锁的线程用完后一定要释放锁，否则那些苦苦等待的线程将永远等待下去，成为死线程。所以我们用 `try...finally` 来确保锁一定会被释放。

锁的好处就是确保了某段关键代码只能由一个线程从头到尾完整地执行，坏处当然也很多，首先是阻止了多线程并发执行，包含锁的某段代码实际上只能以单线程模式执行，效率就大大地下降了。其次，由于可以存在多个锁，不同的线程持有不同的锁，并试图获取对方持有的锁时，可能造成死锁，导致多个线程全部挂起，既不能执行，也无法结束，只能靠操作系统强制终止。

## 多核CPU

如果你不幸拥有一个多核CPU，你肯定在想，多核应该可以同时执行多个线程。

如果写一个死循环的话，会出现什么情况呢？

打开Mac OS X的Activity Monitor，或者Windows的Task Manager，都可以监控某个进程的CPU使用率。

我们可以监控到一个死循环线程会100%占用一个CPU。

如果有两个死循环线程，在多核CPU中，可以监控到会占用200%的CPU，也就是占用两个CPU核心。

要想把 `N` 核CPU的核心全部跑满，就必须启动 `N` 个死循环线程。

试试用Python写个死循环：

```
import threading, multiprocessing

def loop():
    x = 0
    while True:
        x = x + 1

for i in range(multiprocessing.cpu_count()):
    t = threading.Thread(target=loop)
    t.start()
```

启动与CPU核心数量相同的 `N` 个线程，在4核CPU上可以监控到CPU占用率仅有102%，也就是仅使用了一核。

但是用C、C++或Java来改写相同的死循环，直接可以吧全部核心跑满，4核就跑到400%，8核就跑到800%，为什么Python不行呢？

因为Python的线程虽然是真正的线程，但解释器执行代码时，有一个GIL锁：Global Interpreter Lock，任何Python线程执行前，必须先获得GIL锁，然后，每执行100条字节码，解释器就自动释放GIL锁，让别的线程有机会执行。这个GIL全局锁实际上把所有线程的执行代码都给上了锁，所以，多线程在Python中只能交替执行，即使100个线程跑在100核CPU上，也只能用到1个核。

GIL是Python解释器设计的历史遗留问题，通常我们用的解释器是官方实现的CPython，要真正利用多核，除非重写一个不带GIL的解释器。

所以，在Python中，可以使用多线程，但不要指望能有效利用多核。如果一定要通过多线程利用多核，那只能通过C扩展来实现，不过这样就失去了Python简单易用的特点。

不过，也不用于担心，Python虽然不能利用多线程实现多核任务，但可以通过多进程实现多核任务。多个Python进程有各自独立的GIL锁，互不影响。

## 小结

多线程编程，模型复杂，容易发生冲突，必须用锁加以隔离，同时，又要小心死锁的发生。

Python解释器由于设计时有GIL全局锁，导致了多线程无法利用多核。多线程的开发在Python中就是一个美丽的梦。

## 参考源码

[multi\\_threading.py](#)

[do\\_lock.py](#)

读后有收获可以请作者喝咖啡，读后有疑问请加群讨论：



还可以分享给朋友：

分享到微博

← 上一页

下一页 →

廖雪峰

官方 | 独家

Python

商业爬虫全解码

找廖雪峰老师

ACM金牌得主

全球顶尖名企一线数据科学家倾力指导

人工智能与自然语言/计算机视觉课程培训

Artificial Intelligence For NLP/CV Courses

无offer退全款

廖雪峰推荐

JAVA进阶教程

原价1599元

0元领取

阿里云

高性能云服务器首台5折

企业上云零门槛，最大20Gbps内网带宽，450万PPS

立即购买

5折

python免费公开课

编程学习网

授课模式：在线直播+课后视频，从零基础到中级开发工程师

查看详情

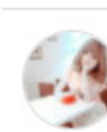
python免费公开课

编程学习网

授课模式：在线直播+课后视频，从零基础到中级开发工程师

查看详情

## 评论

 import time, threading # 假定这是你的银行存款: balance = 0 def change\_it(n): # 先存后取，结果应该为0: global b


执行时间: created at April 8, 2019 9:15 PM, Last updated at May 24, 2019 6:17 PM

为什么我执行lock那边的那段代码，输出的是0

 木小由

Created at May 4, 2019 5:51 PM

你需要多次尝试，会出现几次不是0的情况

 IceWeet

Created at May 12, 2019 12:15 AM