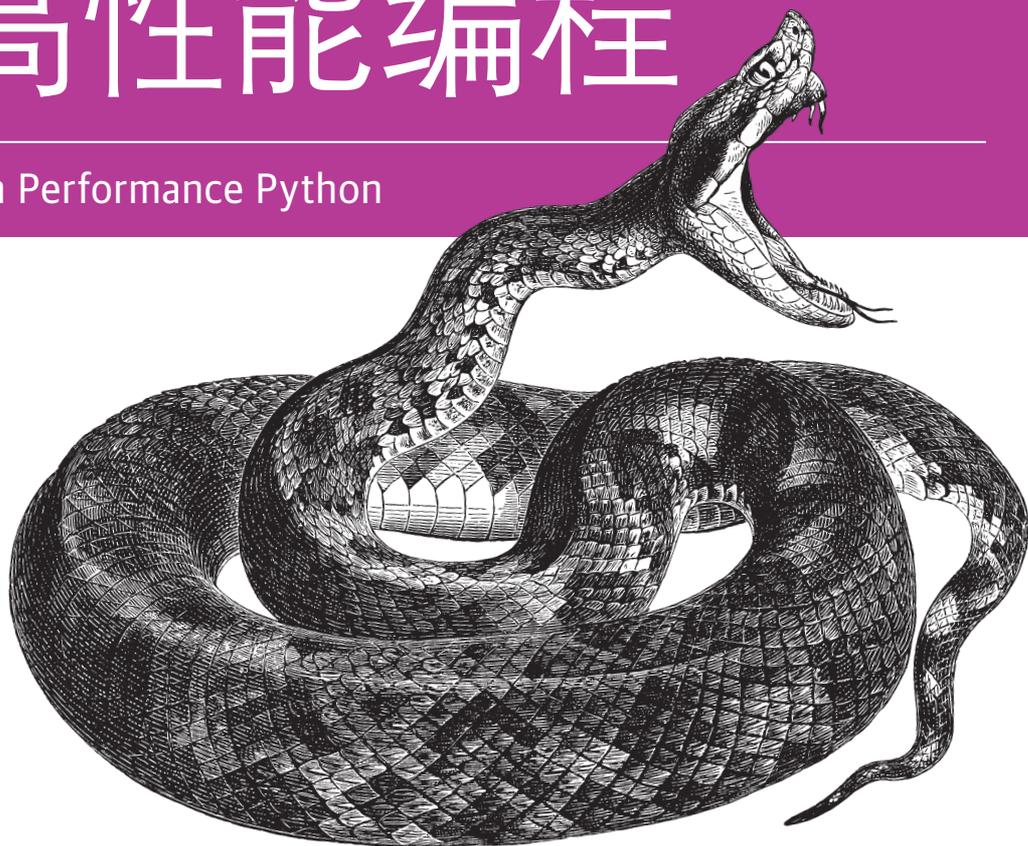


O'REILLY®

异步图书  
www.epubit.com.cn

# Python 高性能编程

High Performance Python



[美] Micha Gorelick Ian Ozsvald 著  
胡世杰 徐旭彬 译

 中国工信出版集团

 人民邮电出版社  
POSTS & TELECOM PRESS

# Python 高性能编程

[美] Micha Gorelick  
Ian Ozsvald 著  
胡世杰 徐旭彬 译

人民邮电出版社

北京

## 图书在版编目 (C I P) 数据

Python高性能编程 / (美) 戈雷利克  
(Micha Gorelick), (美) 欧日沃尔德 (Ian Ozsvald)  
著; 胡世杰, 徐旭彬译. — 北京: 人民邮电出版社,  
2017.7

ISBN 978-7-115-45489-8

I. ①P… II. ①戈… ②欧… ③胡… ④徐… III. ①  
软件工具—程序设计 IV. ①TP311.561

中国版本图书馆CIP数据核字(2017)第114298号

## 版权声明

Copyright © 2014 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2017.  
Authorized translation of the English edition, 2015 O'Reilly Media, Inc., the owner of all rights to publish  
and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

本书中文简体版由 **O'Reilly Media, Inc.** 授权人民邮电出版社出版。未经出版者书面许可, 对本书的  
任何部分不得以任何方式复制或抄袭。

版权所有, 侵权必究。

- 
- ◆ 著 [美] Micha Gorelick Ian Ozsvald  
译 胡世杰 徐旭彬  
责任编辑 陈冀康  
责任印制 焦志炜
  - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号  
邮编 100164 电子邮件 315@ptpress.com.cn  
网址 <http://www.ptpress.com.cn>  
三河市海波印务有限公司印刷
  - ◆ 开本: 800×1000 1/16  
印张: 22  
字数: 415千字 2017年7月第1版  
印数: 1-3000册 2017年7月河北第1次印刷  
著作权合同登记号 图字: 01-2013-9303号

---

定价: 79.00元

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315

---

# 内容提要

Python 语言是一种脚本语言，其应用领域非常广泛，包括数据分析、自然语言处理、机器学习、科学计算、推荐系统构建等。

本书共有 12 章，围绕如何进行代码优化和加快实际应用的运行速度进行详细讲解。本书主要包含以下主题：计算机内部结构的背景知识、列表和元组、字典和集合、迭代器和生成器、矩阵和矢量计算、并发、集群和工作队列等。最后，通过一系列真实案例展现了在应用场景中需要注意的问题。

本书适合初级和中级 Python 程序员、有一定 Python 语言基础想要得到进阶和提高的读者阅读。

---

# 前言

Python 很容易学。你之所以阅读本书可能是因为你的代码现在能够正确运行，而你希望它能跑得更快。你可以很轻松地修改代码，反复地实现你的想法，你对这一点很满意。但*能够轻松实现*和*代码跑得够快*之间的取舍却是一个世人皆知且令人惋惜的现象。而这个问题其实是可以解决的。

有些人想要让顺序执行的过程跑得更快。有些人需要利用多核架构、集群，或者图形处理单元的优势来解决他们的问题。有些人需要可伸缩系统在保证可靠性的前提下酌情或根据资金多少处理更多或更少的工作。有些人意识到他们的编程技巧，通常是来自其他语言，可能不如别人的自然。

我们会在本书中覆盖所有这些主题，给出明智的指导去了解瓶颈并提出效率更高、伸缩性更好的解决方案。我们也会在本书中包含那些来自前人的战场故事，让你可以避免重蹈覆辙。

Python 很适合快速开发、生产环境部署，以及可伸缩系统。Python 的生态系统里到处都是帮你解决伸缩性的人，让你有更多时间处理那些更有挑战性的工作。

## 本书适合哪些人

你使用 Python 的时间已经足够长，了解为什么某些代码会跑得慢，也见过以本书讨论的 Cython、numpy 以及 PyPy 等技术作为解决方案。你可能还有其他语言的编程经验，因此知道解决性能问题的路不止一条。

本书主要的目标读者是那些需要解决 CPU 密集型问题的人，同时我们也会关注数据传输以及内存密集型问题。科学家、工程师、数据分析专家、学者通常会面临这些问题。

我们还会关注网页开发者可能面临的问题，包括数据的移动以及为了快速提升性能而使用 PyPy 这样的即时（JIT）编译器。

如果你有一个 C（或 C++，或 Java）的背景可能会有帮助，但这不是必须的。Python 最常用的解释器（CPython——你在命令行输入 `python` 时启动的标准解释器）是用 C 写的，所以各种钩子和库全都血淋淋地暴露了内部的 C 机制。但我们也会谈到对 C 一无所知的人也能使用的许多其他技术。

你可能还具有 CPU、内存架构和数据总线的底层知识，还是那句话，这些也不完全是必须的。

## 本书不适合哪些人

本书适用于中高级 Python 程序员。积极的 Python 初学者可能可以跟上，但我们建议要具有坚实的 Python 基础。

我们不会探讨存储系统优化。如果你有一个 SQL 或 NoSQL 瓶颈，本书可能帮不了你。

## 你会学到什么

我们两位作者在业界和学术界工作了很多年，专门应对大数据应用、处理*我需要更快得到答案!*之类的请求、可伸缩架构等需求。我们会将自己经历千辛万苦获得的经验传授于你，让你免于重蹈覆辙。

在每一章开头，我们会列出问题，并在后续的文字中回答（如果没有回答，告诉我们，我们会在下一个版本中修正!）。

我们会覆盖下面这些主题：

- 计算机内部结构的背景知识，让你知道在底层发生了什么。
- 列表和元组——在这些基本数据结构中细微的语义和速度区别。
- 字典和集合——在这些重要数据结构中的内存分配策略和访问算法。
- 迭代器——Python 风格的代码应该怎样写，用迭代打开无限数据流的大门。
- 纯 Python 方法——如何高效使用 Python 及其模块。
- 使用 `numpy` 的矩阵——像一头野兽一样使用心爱的 `numpy` 库。
- 编译和即时计算——编译成机器码可以跑得更快，让性能分析的结果指引你。
- 并发——高效移动数据的方法。
- `multiprocessing`——使用内建 `multiprocessing` 库进行并行计算的各种方式，高效共享 `numpy` 矩阵、进程间通信（IPC）的代价和收益。
- 集群计算——将你的 `multiprocessing` 代码转换成在研究系统以及生产系统的本地集群或远程集群上运行的代码。
- 使用更少的 RAM——不需要购买大型机就能解决大型问题的方法。
- 现场教训——来自前人的战场故事，让你可以避免重蹈覆辙。

## Python 2.7

Python 2.7 在科学和工程计算中是占主导地位的 Python 版本。在 \*nix 环境(通常是 Linux 或 Mac) 下, 64 位的版本占了主导地位。64 位让你能够拥有更宽广的 RAM 寻址范围。\*nix 让你构建出的应用程序的行为、部署和配置方法都可以很容易地被别人所理解。

如果你是一个 Windows 用户, 那么你就需要系好安全带了。我们展示的大多数代码都可以正常工作, 但有些东西是针对特定操作系统的, 你将不得不研究 Windows 下的解决方案。Windows 用户可能面临的最大的困难是模块的安装: 搜索 StackOverflow 等站点应该可以帮助你找到你需要的答案。如果你正在使用 Windows, 那么使用一台安装了 Linux 的虚拟机(比如 VirtualBox) 可能可以帮助你更自由地进行实验。

Windows 用户绝对应该看看那些通过 Anaconda、Canopy、Python(x,y) 或 Sage 等 Python 发行版提供的打包的解决方案。这些发行版也会让 Linux 和 Mac 用户的生活简单许多。

## 迁移至 Python 3

Python 3 是 Python 的未来, 每一个人都在迁移过去。尽管 Python 2.7 还将在接下来的很多年里面继续跟我们做伴(有些安装版仍在 使用 2004 年的 Python 2.4), 它的退役日期已经被定在 2020 年了。

升级到 Python 3.3+ 让 Python 库的开发者伤透了脑筋, 人们移植代码的速度一直都很慢(这是有原因的), 所以人们转用 Python 3 的速度也很慢。这主要是因为要把一个混用了 Python 2 的 string 和 Unicode 数据类型的应用程序切换到 Python 3 的 Unicode 和 byte 实在太过复杂了。

通常来说, 当你需要重现基于一批值得信任的库的结果时, 你不会想要站在危险的技术前沿。高性能 Python 的开发者更有可能在接下来的几年里使用和信任 Python 2.7。

本书的大多数代码只需要稍做修改就能运行于 Python 3.3+ (最明显的修改是 print 从一个语句变成了一个函数)。在一些地方, 我们将特地关注 Python 3.3+ 带来的性能提升。一个可能需要你关注的地方是在 Python 2.7 中 / 表示 *integer* 的除法, 而在 Python 3 中它变成了 *float* 的除法。当然, 作为一个好的开发者, 你精心编写的单元测试应该已经在测试你的关键代码路径了, 所以如果你的代码需要关注这点, 那么你应该已经收到来自你单元测试的警告了。

scipy 和 numpy 从 2010 年开始就已经兼容 Python 3 了。matplotlib 从 2012 年开始兼容, scikit-learn 是 2013, NLTK 是 2014, Django 是 2013。这些库的迁移备忘录可以在它们各自的代码库和新闻组里查看。如果你也有旧代码需要移植到 Python 3, 那么就值得回顾一下这些库移植的过程。

我们鼓励你用 Python 3.3+ 进行新项目的开发，但你要当心那些最近刚刚移植还没有多少用户的库——追踪 bug 将会更困难。比较明智的做法是让你的代码可以兼容 Python 3.3+（学习一下 `__future__` 模块的导入），这样未来的升级就会更简单。

有两本参考手册不错：《把 Python 2 的代码移植到 Python 3》和《移植到 Python 3：深度指南》。Anaconda 或 Canopy 这样的发行版让你可以同时运行 Python 2 和 Python 3——这会让你的移植变得简单一些。

## 版权声明

本书版权符合知识共享协议“署名-非商业性使用-禁止演绎 3.0”。

欢迎以非商业性目的使用本书，包括非商业性教育。本书许可完整转载，如果你需要部分转载，请联系 O'Reilly（见后面“联系我们”部分）。请根据下面的提示进行署名。

我们经过协商认为本书应该使用知识共享许可证，让其内容在世界上更广泛传播。如果这个决定帮到了你，我们将十分高兴收到你的啤酒。我们估计 O'Reilly 的员工对啤酒的看法跟我们相同。

## 如何引用

如果你需要使用本书，知识共享许可证要求你署名。署名意味着你需要写一些东西让其他人能够找到本书。下面是一个不错的例子：“High Performance Python by Micha Gorelick and Ian Ozsvald (O'Reilly). Copyright 2014 Micha Gorelick and Ian Ozsvald, 978-1-449-36159-4.”

## 勘误和反馈

我们鼓励你在 Amazon 这样的公开网站上评论本书——请帮助其他人了解他们是否能从本书中受益！你也可以发 E-mail 给我们：[feedback@highperformancepython.com](mailto:feedback@highperformancepython.com)。

我们特别希望听到您指出本书的错误，本书帮到你的成功案例，以及我们应该在下一版本加上的高性能技术。你可以通过 O'Reilly 官网给我们留言。

至于抱怨，欢迎你使用即时抱怨传输服务 `> /dev/null`。

## 排版约定

本书采用下列排版约定：

*斜体*

表示新词、E-mail 地址、文件名，以及文件扩展名。

等宽

用于程序列印，以及在文字中表示命令、模块和程序元素，如变量或函数名、数据库、数据类型、环境变量、语句和关键字。

**等宽加粗**

表示命令或其他需要用户原封不动输入的文字。

*等宽斜体*

表示需要被替换成用户指定的值或根据上下文决定的值。



**问题**

这个记号表示一个问题或练习。



**备忘**

这个记号表示一个备忘。



**警告**

这个记号表示一个警告或注意。

## 使用示例代码

补充材料（示例代码、练习等）可以通过 [GitHub](#) 下载。

本书是为了帮你搞定你的问题。通常来说，只要是本书提供的示例代码，你就可以在你的程序和文档中使用。你不需要联系我们获得许可，除非你需要对很大一部分代码进行转载。比如，写一个使用了好几段本书代码的程序不需要许可。以 CD-ROM 的形式销售或分发 O'Reilly 图书中的示例需要许可。引用本书文字和示例代码回答问题不需要许可。在你的产品文档中合并大量本书示例代码需要许可。

如果你觉得你对示例代码的使用超出了上述的许可范围，请通过 [permissions@oreilly.com](mailto:permissions@oreilly.com) 联系我们。

## Safari® 在线图书



*Safari 在线图书*是一个按需数字图书馆，它以书和视频的形式提供来自全球顶尖作者的技术和商业内容。

技术专家、软件开发人员、网页设计者，以及商业和创新人员将 Safari 在线图书当成他们研究、解决问题、学习和资格认证训练的主要资源。

Safari 在线图书为企业，政府，教育和个人提供了各种收费标准。

其成员可以通过全文搜索数据库访问成千上万的图书，训练视频，以及还未正式出版的手稿。它们来自 O'Reilly Media、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Focal Press、Cisco Press、John Wiley & Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、McGraw-Hill、Jones & Bartlett、Course Technology，以及其他几百个出版社。更多信息请在线访问 <https://www.safaribooksonline.com/>。

## 联系我们

请将关于本书的评论和问题发给本书出版社：

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

800-998-9938 (in the United States or Canada)

707-829-0515 (international or local)

707-829-0104 (fax)

你也可以发送 E-mail 到 [bookquestions@oreilly.com](mailto:bookquestions@oreilly.com) 对本书进行评论或询问技术问题。

关于我们的图书、课程、会议和新闻等更多信息请访问我们的网站。

我们的 Facebook: <http://facebook.com/oreilly>

Twitter: <http://twitter.com/oreillymedia>

YouTube: <http://www.youtube.com/oreillymedia>

## 致谢

感谢来自 Jake Vanderplas、Brian Granger、Dan Foreman-Mackey、Kyran Dale、John Montgomery、Jamie Matthews、Calvin Giles、William Winter、Christian Schou Oxvig、

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：《Python高性能编程》【美】戈雷利克 (Micha Gorelick) 著.pdf

请登录 <https://shgis.com/post/4066.html> 下载完整文档。

手机端请扫码查看：

