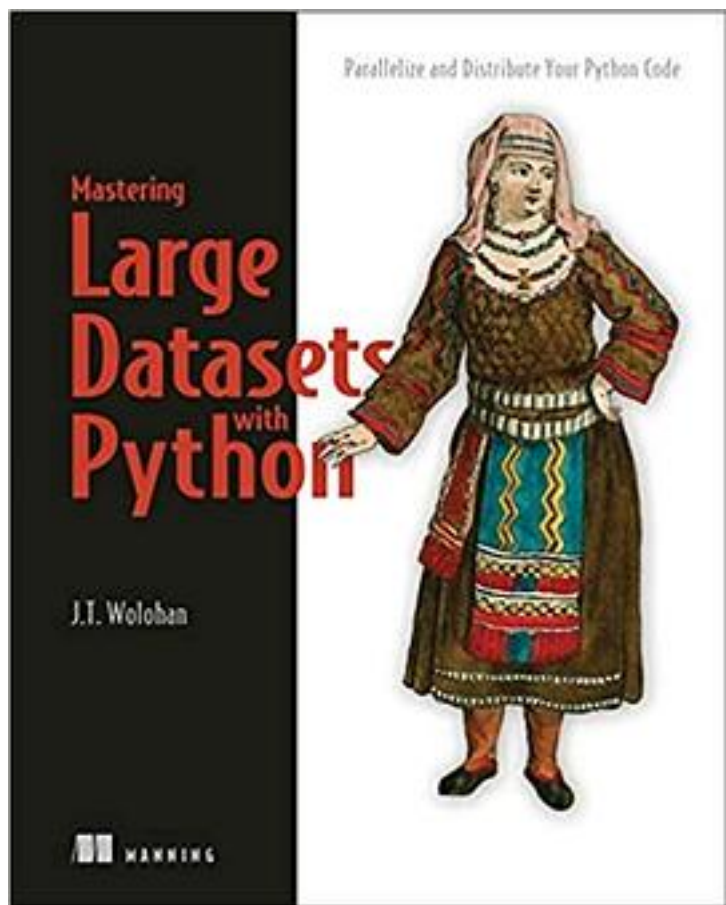


# Mastering Large Datasets



[Mastering Large Datasets 下载链接1](#)

著者:J. T. Wolohan

出版者:Manning Publications

出版时间:2020-1-6

装帧:Paperback

isbn:9781617296239

With an emphasis on clarity, style, and performance, author J.T. Wolohan expertly guides you through implementing a functionally-influenced approach to Python coding. You'll get familiar with Python's functional built-ins like the `functools` operator and `itertools` modules, as well as the `toolz` library.

Mastering Large Datasets teaches you to write easily readable, easily scalable Python code that can efficiently process large volumes of structured and unstructured data. By the end of this comprehensive guide, you'll have a solid grasp on the tools and methods that will take your code beyond the laptop and your data science career to the next level!

作者介绍:

J.T. Wolohan is a lead data scientist at Booz Allen Hamilton and a PhD researcher at Indiana University, Bloomington, affiliated with the Department of Information and Library Science and the School of Informatics and Computing. His professional work focuses on rapid prototyping and scalable AI. His research focuses on computational analysis of social uses of language online.

目录:

[Mastering Large Datasets\\_ 下载链接1](#)

标签

分布式

Python

2020

软件工程

Methodology

评论

part1讲parallel and laziness in python讲得非常深入浅出，但后面讲hadoop和spark就有点虎头蛇尾的感觉了，假如是为了看一看具体的mapreduce架构的实现最好不要看这本，但是假如是想具体了解大数据的思路和一些最基本的trade off还是值得读的

-----  
就还是有点浅，两小时就刷完了。

-----  
[Mastering Large Datasets\\_ 下载链接1](#)

书评

-----  
[Mastering Large Datasets\\_ 下载链接1](#)