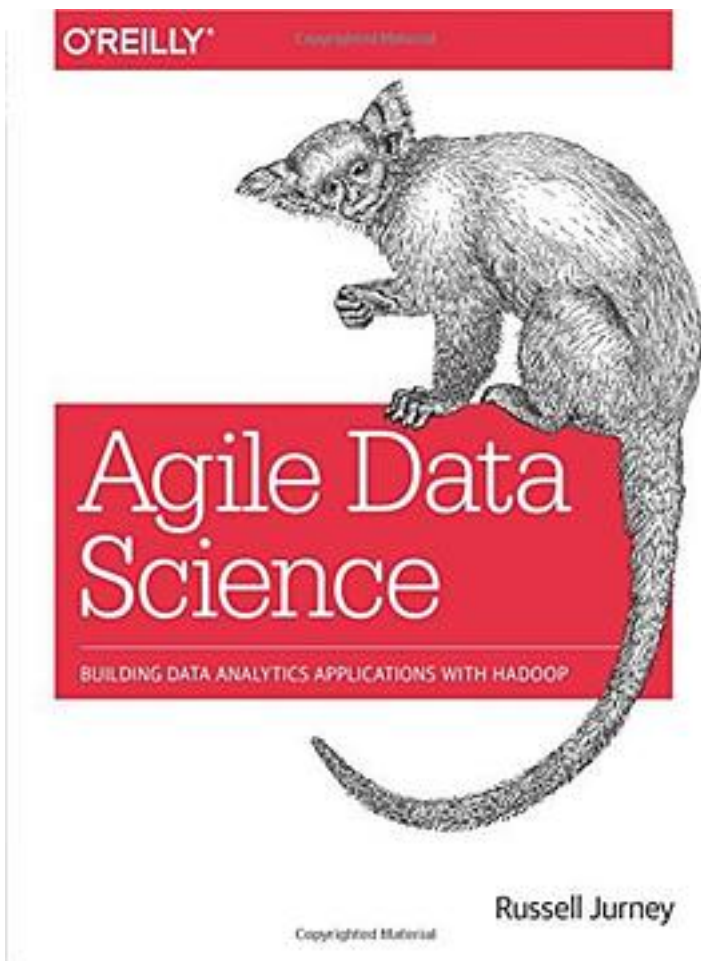


Agile Data Science



[Agile Data Science 下载链接1](#)

著者:Russell Journey

出版者:O'Reilly Media

出版时间:2013-10-25

装帧:Paperback

isbn:9781449326265

Mining data requires a deep investment in people and time. How can you be sure you're building the right models? What tools help you connect with the customer's needs? With this hands-on book, you'll learn a flexible toolset and methodology for

building effective analytics applications. Agile Data shows you how to create an environment for exploring data, using lightweight tools such as Ruby, Python, Apache Pig, and the D3.js (Data-Driven Documents) JavaScript library. You'll learn an iterative approach that allows you to quickly change the kind of analysis you're doing, as you discover what the data is telling you. All the example code in this book is available as working Heroku apps. Build an application to mine your own email inbox Use several data structures to extract multiple features from a single dataset, and learn how different perspectives can yield insight Rapidly boot your applications as simple front-ends to key/value stores Add features driven by descriptive and inferential statistics, machine learning, and data visualization Gather usage data and talk to real users to help guide your data-driven exploration You can provide constructive comments on the manuscript through O'Reilly's Open Feedback Publishing System (OFPS). Learn more at <http://labs.oreilly.com/ofps.html>.

作者介绍:

About the Author

Russell Journey cut his data teeth in casino gaming, building web apps to analyze the performance of slot machines in the US and Mexico. After dabbling in entrepreneurship, interactive media and journalism, he moved to silicon valley to build analytics applications at scale at Ning and LinkedIn. He lives on the ocean in Pacifica, California with his wife Kate and two fuzzy dogs.

目录:

[Agile Data Science 下载链接1](#)

标签

数据挖掘

数据分析

DataScience

计算机科学

机器学习

数据科学

计算机

大数据

评论

介绍了很多工具，怎么组合使用还是要看场景。

hadoop, pig, 入门书，讲解了一个data science 团队的分工和流程， example 是：利用 gmail 数据做一个智能邮箱（推荐发件人等）

入门教程

以发送邮件为例，比较形象的展现了
数据分析的五个级别，记录—图表—报表—预测—驱动及敏捷数据分析用到的一些工具

入门教程，不错

[Agile Data Science 下载链接1](#)

书评

这本书的第二版已经于2018年出版了。这第一版面世于2014年，第二版在此基础之上
有非常大幅度的修改。但最最基本的思路没有变化：端到端，全栈，敏捷，技术为具体
业务服务。第二版的链接在下面：[Spark全栈数据分析]
对比两个版本，除了内容扩充了不少，处理的问题更加充实， ...

[Agile Data Science 下载链接1](#)