

O'REILLY®



Learning Spark

LIGHTNING-FAST DATA ANALYSIS

Holden Karau, Andy Konwinski,
Patrick Wendell & Matei Zaharia

Table of Contents

1. [introduction](#)

Learning Spark: Lightning-Fast Big Data Analysis

Chinese translation

Translation the book of Learning Spark: Lightning-Fast Big Data Analysis is only for spark developer educational purposes. If I violated your copyright, please let me know.

《Learning Spark: Lightning-Fast Big Data Analysis》的中文翻译纯属个人对于Spark的兴趣，仅供学习。

如果我的翻译行为侵犯您的版权，请您告知，我将停止对此书的开源翻译。

About the Author

Holden Karau is a software development engineer at Databricks and is active in open source. She is the author of an earlier Spark book. Prior to Databricks she worked on a variety of search and classification problems at Google, Foursquare, and Amazon. She graduated from the University of Waterloo with a Bachelors of Mathematics in Computer Science. Outside of software she enjoys playing with fire, welding, and hula hooping.

Most recently, Andy Konwinski co-founded Databricks. Before that he was a PhD student and then postdoc in the AMPLab at UC Berkeley, focused on large scale distributed computing and cluster scheduling. He co-created and is a committer on the Apache Mesos project. He also worked with systems engineers and researchers at Google on the design of Omega, their next generation cluster scheduling system. More recently, he developed and led the AMP Camp Big Data Bootcamps and first Spark Summit, and has been contributing to the Spark project.

Patrick Wendell is an engineer at Databricks as well as a Spark Committer and PMC member. In the Spark project, Patrick has acted as release manager for several Spark releases, including Spark 1.0. Patrick also maintains several subsystems of Spark's core engine. Before helping start Databricks, Patrick obtained an M.S. in Computer Science at UC Berkeley. His research focused on low latency scheduling for large scale analytics workloads. He holds a B.S.E in Computer Science from Princeton University

Matei Zaharia is the creator of Apache Spark and CTO at Databricks. He holds a PhD from UC Berkeley, where he started Spark as a research project. He now serves as its Vice President at Apache. Apart from Spark, he has made research and open source contributions to other projects in the cluster computing area, including Apache Hadoop (where he is a committer) and Apache Mesos (which he also helped start at Berkeley).

Examples for Learning Spark

codes <https://github.com/gaoxuesong/learning-spark/> forked from <https://github.com/databricks/learning-spark>