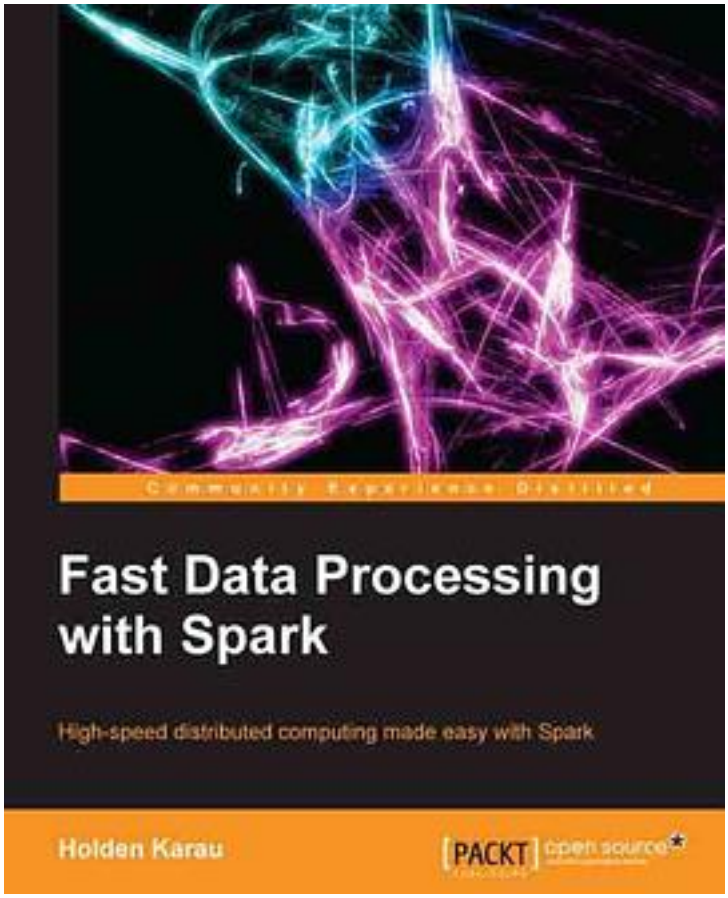
Fast Data Processing with Spark



Fast Data Processing with Spark_下载链接1_

著者:Holden Karau

出版者:Packt Publishing

出版时间:2013-10-23

装帧:Paperback

isbn:9781782167068

Overview

Implement Spark's interactive shell to prototype distributed applications

Deploy Spark jobs to various clusters such as Mesos, EC2, Chef, YARN, EMR, and so on

Use Shark's SQL query-like syntax with Spark

In Detail

Spark is a framework for writing fast, distributed programs. Spark solves similar problems as Hadoop MapReduce does but with a fast in-memory approach and a clean functional style API. With its ability to integrate with Hadoop and inbuilt tools for interactive query analysis (Shark), large-scale graph processing and analysis (Bagel), and real-time analysis (Spark Streaming), it can be interactively used to quickly process and query big data sets.

Fast Data Processing with Spark covers how to write distributed map reduce style programs with Spark. The book will guide you through every step required to write effective distributed programs from setting up your cluster and interactively exploring the API, to deploying your job to the cluster, and tuning it for your purposes.

Fast Data Processing with Spark covers everything from setting up your Spark cluster in a variety of situations (stand-alone, EC2, and so on), to how to use the interactive shell to write distributed code interactively. From there, we move on to cover how to write and deploy distributed jobs in Java, Scala, and Python.

We then examine how to use the interactive shell to quickly prototype distributed programs and explore the Spark API. We also look at how to use Hive with Spark to use a SQL-like query syntax with Shark, as well as manipulating resilient distributed datasets (RDDs).

What you will learn from this book

Prototype distributed applications with Spark's interactive shell

Learn different ways to interact with Spark's distributed representation of data (RDDs)

Load data from the various data sources

Query Spark with a SQL-like query syntax

Integrate Shark queries with Spark programs

Effectively test your distributed software

Tune a Spark installation

Install and set up Spark on your cluster

Work effectively with large data sets

Approach

This book will be a basic, step-by-step tutorial, which will help readers take advantage of all that Spark has to offer.

Who this book is written for

Fast Data Processing with Spark is for software developers who want to learn how to write distributed programs with Spark. It will help developers who have had problems that were too much to be dealt with on a single computer. No previous experience with distributed programming is necessary. This book assumes knowledge of either Java, Scala, or Python.

作者介绍:

Holden Karau

Holden Karau is a transgendered software developer from Canada currently living in San Francisco. Holden graduated from the University of Waterloo in 2009 with a Bachelors of Mathematics in Computer Science. She currently works as a Software Development Engineer at Google. She has worked at Foursquare, where she was introduced to Scala. She worked on search and classification problems at Amazon. Open Source development has been a passion of Holden's from a very young age, and a number of her projects have been covered on Slashdot. Outside of programming, she enjoys playing with fire, welding, and dancing. You can learn more at her website (http://www.holdenkarau.com), blog (http://blog.holdenkarau.com), and github (https://github.com/holdenk).

目录: Table of Contents

Fast Data Processing with Spark_下载链接1_

标签

Spark

数据挖掘

大数据

Data

计算机

美国

科普

数据_处理

评论

这本书拿来做Spark的入门还是不错。只是成书的都过时了，建议还是直接阅读官方docs的好。

这样的小册子都拿出来出书真的大丈夫（而且这种东西需要时时更新的呀喂）。

...

看过中文版的了 ... 没意思

不太详细

可读性比较差 没有提及基本原理和架构 只是大致过了一遍常用的api

只是初步的泛泛讲解，入门可以读读

内容太简单、太少了

充斥的都是Java&Scala
两个版本的Example。。。内容都是简介型，不涉及太多的开发，运维核心内容。价值不大。

内容过于单薄。或者说，Spark的API本身就很简单。感觉如果作者能加一些实际的例子就更好了。

Fast Data Processing with Spark_下载链接1_

书评

饶了我吧，最近太背了，买了这么多垃圾书。 本来以为国外的书，内容会好一些
买来才发现，就是一本骗钱使用手册 薄薄的几页纸，还没doc全。
这样的东西也可以出书。。 实在太无聊了，正在纠结要不要退货呢。

Fast Data Processing with Spark_下载链接1_