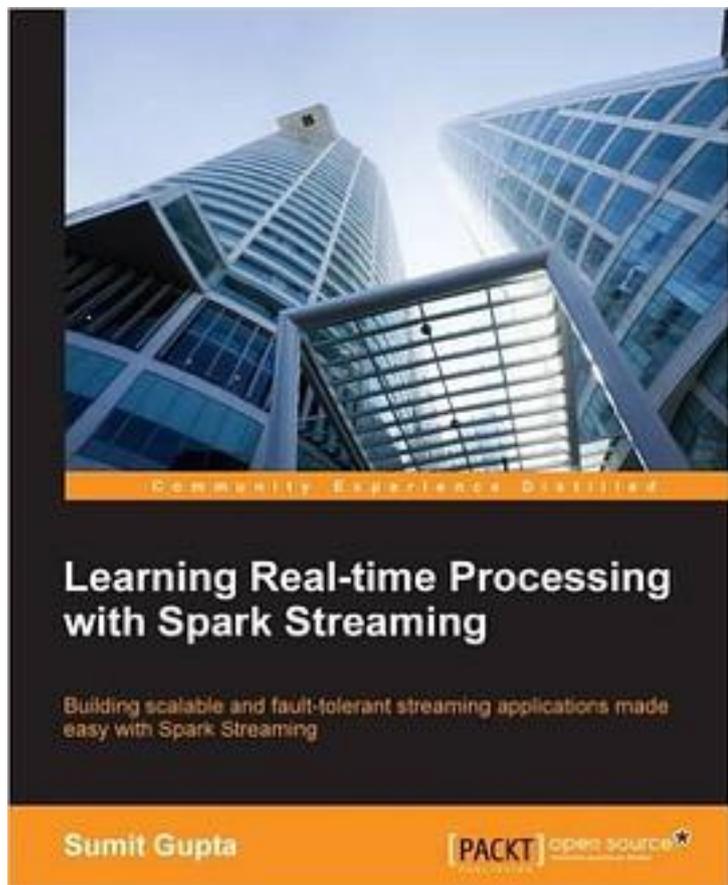


Learning Real Time processing with Spark Streaming



[Learning Real Time processing with Spark Streaming_下载链接1](#)

著者:Sumit Gupta

出版者:Packt Publishing

出版时间:2015-10-1

装帧:Paperback

isbn:9781783987665

Building scalable and fault-tolerant streaming applications made easy with Spark streaming

About This Book

Process live data streams more efficiently with better fault recovery using Spark StreamingImplement and deploy real-time log file analysisLearn about integration with Advance Spark Libraries – GraphX, Spark SQL, and MLib.

Who This Book Is For

This book is intended for big data developers with basic knowledge of Scala but no knowledge of Spark. It will help you grasp the basics of developing real-time applications with Spark and understand efficient programming of core elements and applications.

What You Will Learn

Install and configure Spark and Spark Streaming to execute applicationsExplore the architecture and components of Spark and Spark Streaming to use it as a base for other librariesProcess distributed log files in real-time to load data from distributed sourcesApply transformations on streaming data to use its functionsIntegrate Apache Spark with the various advance libraries like MLib and GraphXApply production deployment scenarios to deploy your application

In Detail

Using practical examples with easy-to-follow steps, this book will teach you how to build real-time applications with Spark Streaming.

Starting with installing and setting the required environment, you will write and execute your first program for Spark Streaming. This will be followed by exploring the architecture and components of Spark Streaming along with an overview of libraries/functions exposed by Spark. Next you will be taught about various client APIs for coding in Spark by using the use-case of distributed log file processing. You will then apply various functions to transform and enrich streaming data. Next you will learn how to cache and persist datasets. Moving on you will integrate Apache Spark with various other libraries/components of Spark like Mlib, GraphX, and Spark SQL. Finally, you will learn about deploying your application and cover the different scenarios ranging from standalone mode to distributed mode using Mesos, Yarn, and private data centers or on cloud infrastructure.

Style and approach

A Step-by-Step approach to learn Spark Streaming in a structured manner, with detailed explanation of basic and advance features in an easy-to-follow Style. Each topic is explained sequentially and supported with real world examples and executable code snippets that appeal to the needs of readers with the wide range of experiences.

作者介绍:

About the Author

Sumit Gupta

Sumit Gupta is a seasoned professional, innovator, and technology evangelist with over 100 months of experience in architecting, managing, and delivering enterprise solutions revolving around a variety of business domains such as hospitality, healthcare, risk management, insurance, and so on. He is passionate about technology with an overall 14 years of hands-on experience in the software industry and has been using big data and cloud technologies over the past 4 to 5 years to solve complex business problems. He is also the author of Neo4j Essentials and Building Web Applications with Python and Neo4j both by Packt Publishing.

目录:

[Learning Real Time processing with Spark Streaming_下载链接1](#)

标签

Spark

大数据

DataScience

评论

入门级别

书中很多实用的例子可以学习。

[Learning Real Time processing with Spark Streaming_下载链接1](#)

书评

[Learning Real Time processing with Spark Streaming_下载链接1_](#)