

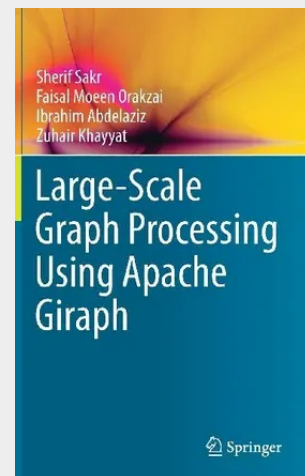
Large-Scale Graph Processing Using Apache Giraph

This book takes its reader on a journey through Apache Giraph, a popular distributed graph processing platform designed to bring the power of big data processing to graph data. Designed as a step-by-step self-study guide for everyone interested in large-scale graph processing, it describes the fundamental abstractions of the system, its programming models and various techniques for using the system to process graph data at scale, including the implementation of several popular and advanced graph analytics algorithms.

The book is organized as follows: Chapter 1 starts by providing a general background of the big data phenomenon and a general introduction to the Apache Giraph system, its abstraction, programming model and design architecture. Next, chapter 2 focuses on Giraph as a platform and how to use it. Based on a sample job, even more advanced topics like monitoring the Giraph application lifecycle and different methods for monitoring Giraph jobs are explained. Chapter 3 then provides an introduction to Giraph programming, introduces the basic Giraph graph model and explains how to write Giraph programs. In turn, Chapter 4 discusses in detail the implementation of some popular graph algorithms including PageRank, connected components, shortest paths and triangle closing. Chapter 5 focuses on advanced Giraph programming, discussing common Giraph algorithmic optimizations, tunable Giraph configurations that determine the system's utilization of the underlying resources, and how to write a custom graph input and output format. Lastly, chapter 6 highlights two systems that have been introduced to tackle the challenge of large scale graph processing, GraphX and GraphLab, and explains the main commonalities and differences between these systems and Apache Giraph.

This book serves as an essential reference guide for students, researchers and practitioners in the domain of large scale graph processing. It offers step-by-step guidance, with several code examples and the complete source code available in the related github repository. Students will find a comprehensive introduction to and hands-on practice with tackling large scale graph processing problems using the Apache Giraph system, while researchers will discover thorough coverage of the emerging and ongoing advancements in big graph processing systems.

This book takes its reader on a journey through Apache Giraph, a popular distributed graph processing platform designed to bring the power of big data processing to graph data. Designed as a step-by-step self-study guide for everyone interested in large-scale graph processing, it describes the fundamental abstractions of the system, its programming models and various techniques for using the system to process graph data at scale, including the implementation of several popular and advanced graph analytics algorithms. The book is organized as follows: Chapter 1 starts by providing a general background of the big data phenomenon and a general introduction to the Apache Giraph system, its abstraction, programming model and design architecture. Next, chapter 2 focuses on Giraph as a platform and how to use it. Based on a sample job, even more advanced topics like monitoring the Giraph application lifecycle and different methods for monitoring Giraph jobs are explained. Chapter 3 then provides an introduction to Giraph programming, introduces the basic Giraph graph model and explains how to write Giraph programs. In turn, Chapter 4 discusses in detail the implementation of some popular graph algorithms including PageRank, connected components, shortest paths and triangle closing. Chapter 5 focuses on advanced Giraph programming, discussing common Giraph algorithmic optimizations, tunable Giraph configurations that determine the system's utilization of the underlying resources, and how to write a custom graph input and output format. Lastly, chapter 6 highlights two systems that have been introduced to tackle the challenge of large scale graph processing, GraphX and GraphLab, and explains the main commonalities and differences



58,84 €

54,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9783319474304

Medium: Buch

ISBN: 978-3-319-47430-4

Verlag: Springer

Erscheinungstermin: 12.01.2017

Sprache(n): Englisch

Auflage: 1. Auflage 2016

Produktform: Gebunden

Gewicht: 4616 g

Seiten: 197

Format (B x H): 160 x 241 mm

