1. Record Nr. UNINA9910739483403321

Titolo	Big Data Processing Using Spark in Cloud [[electronic resource] /] / edited by Mamta Mittal, Valentina E. Balas, Lalit Mohan Goyal, Raghvendra Kumar
Pubbl/distr/stampa	Singapore : , : Springer Singapore : , : Imprint : Springer, , 2019
ISBN	981-13-0550-1
Edizione	[1st ed. 2019.]
Descrizione fisica	1 online resource (XIII, 264 p. 89 illus., 62 illus. in color.)
Collana	Studies in Big Data, , 2197-6503 ; ; 43
Disciplina	005.7
Soggetti	Big data Computer security Big Data Systems and Data Security Big Data/Analytics
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Concepts of Big Data and Apache Spark -- Big Data Analysis in Cloud and Machine Learning -- Security Issues and Challenges related to Big Data -- Big Data Security Solutions in Cloud -- Data Science and Analytics -- Big Data Technologies -- Data Analysis with Casandra and Spark -- Spin up the Spark Cluster -- Learn Scala -- IO for Spark -- Processing with Spark -- Spark Data Frames and Spark SQL -- Machine Learning and Advanced Analytics -- Parallel Programming with Spark -- Distributed Graph Processing with Spark -- Real Time Processing with Spark -- Spark in Real World -- Case Studies. .
Sommario/riassunto	The book describes the emergence of big data technologies and the role of Spark in the entire big data stack. It compares Spark and Hadoop and identifies the shortcomings of Hadoop that have been overcome by Spark. The book mainly focuses on the in-depth architecture of Spark and our understanding of Spark RDDs and how RDD complements big data’s immutable nature, and solves it with lazy evaluation, cacheable and type inference. It also addresses advanced topics in Spark, starting with the basics of Scala and the core Spark framework, and exploring Spark data frames, machine learning using

Mllib, graph analytics using Graph X and real-time processing with Apache Kafka, AWS Kenisis, and Azure Event Hub. It then goes on to investigate Spark using PySpark and R. Focusing on the current big data stack, the book examines the interaction with current big data tools, with Spark being the core processing layer for all types of data. The book is intended for data engineers and scientists working on massive datasets and big data technologies in the cloud. In addition to industry professionals, it is helpful for aspiring data processing professionals and students working in big data processing and cloud computing environments.