1. Record Nr. UNINA9910300752003321

Autore	Luu Hien
Titolo	Beginning Apache Spark 2 [[electronic resource]] : With Resilient Distributed Datasets, Spark SQL, Structured Streaming and Spark Machine Learning library / / by Hien Luu
Pubbl/distr/stampa	Berkeley, CA : , : Apress : , : Imprint : Apress, , 2018
ISBN	1-4842-3579-7
Edizione	[1st ed. 2018.]
Descrizione fisica	1 online resource (XI, 393 p. 86 illus.)
Disciplina	005.7
Soggetti	Big data Java (Computer program language) Data mining Open source software Computer programming Big Data Java Data Mining and Knowledge Discovery Open Source
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	1. Introduction to Apache Spark -- 2. Working with Apache Spark -- 3. Resilient Distributed Dataset -- 4. Spark SQL - Foundation -- 5. Spark SQL - Advanced -- 6. Spark Streaming -- 7. Spark Streaming Advanced -- 8. Machine Learning with Spark.
Sommario/riassunto	Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud technologies. Beginning Apache Spark 2 gives you an introduction to Apache Spark and shows you how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applications with Spark Structured Streaming. Furthermore, you'll learn the fundamentals of Spark ML for machine learning and much more.

After you read this book, you will have the fundamentals to become proficient in using Apache Spark and know when and how to apply it to your big data applications. You will: Understand Spark unified data processing platform Use and manipulate RDDs Deal with structured data using Spark SQL Build real-time applications using Spark Structured Streaming Develop intelligent applications with the Spark Machine Learning library.