

1. Record Nr.	UNINA9910303442303321
Autore	Chellappan Subhashini
Titolo	Practical Apache Spark : Using the Scala API // by Subhashini Chellappan, Dharanitharan Ganesan
Pubbl/distr/stampa	Berkeley, CA : , : Apress : , : Imprint : Apress, , 2018
ISBN	9781484236529 1484236521
Edizione	[1st ed. 2018.]
Descrizione fisica	1 online resource (288 pages)
Disciplina	005.114
Soggetti	Big data Open source software Computer programming Programming languages (Electronic computers) Big Data Open Source Programming Languages, Compilers, Interpreters
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Sommario/riassunto	Work with Apache Spark using Scala to deploy and set up single-node, multi-node, and high-availability clusters. This book discusses various components of Spark such as Spark Core, DataFrames, Datasets and SQL, Spark Streaming, Spark MLib, and R on Spark with the help of practical code snippets for each topic. Practical Apache Spark also covers the integration of Apache Spark with Kafka with examples. You'll follow a learn-to-do-by-yourself approach to learning – learn the concepts, practice the code snippets in Scala, and complete the assignments given to get an overall exposure. On completion, you'll have knowledge of the functional programming aspects of Scala, and hands-on expertise in various Spark components. You'll also become familiar with machine learning algorithms with real-time usage. You will: Discover the functional programming features of Scala Understand the complete architecture of Spark and its components Integrate

Apache Spark with Hive and Kafka Use Spark SQL, DataFrames, and Datasets to process data using traditional SQL queries Work with different machine learning concepts and libraries using Spark's MLlib packages.

---