1.	Record Nr.	UNINA9910819973803321
	Autore	Mehrotra Shrey
	Titolo	Apache Spark Quick Start guide : quickly learn the art of writing efficient big data applications with Apache Spark / / Shrey Mehrotra, Akash Grade
	Pubbl/distr/stampa	Birmingham : , : Packt, , 2019
	Edizione	[1st edition]
	Descrizione fisica	1 online resource (154 pages)
	Soggetti	Big data Data mining - Computer programs
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Note generali	Includes index.
	Sommario/riassunto	A practical guide for solving complex data processing challenges by applying the best optimizations techniques in Apache Spark. Key Features Learn about the core concepts and the latest developments in Apache Spark Master writing efficient big data applications with Spark's built-in modules for SQL, Streaming, Machine Learning and Graph analysis Get introduced to a variety of optimizations based on the actual experience Book Description Apache Spark is a flexible framework that allows processing of batch and real-time data. Its unified engine has made it quite popular for big data use cases. This book will help you to get started with Apache Spark 2.0 and write big data applications for a variety of use cases. It will also introduce you to Apache Spark ? one of the most popular Big Data processing frameworks. Although this book is intended to help you get started with Apache Spark, but it also focuses on explaining the core concepts. This practical guide provides a quick start to the Spark 2.0 architecture and its components. It teaches you how to set up Spark on your local machine. As we move ahead, you will be introduced to resilient distributed datasets (RDDs) and DataFrame APIs, and their corresponding transformations and actions. Then, we move on to the life cycle of a Spark application and learn about the techniques used to

debug slow-running applications. You will also go through Spark's built-in modules for SQL, streaming, machine learning, and graph analysis. Finally, the book will lay out the best practices and optimization techniques that are key for writing efficient Spark applications. By the end of this book, you will have a sound fundamental understanding of the Apache Spark framework and you will be able to write and optimize Spark applications. What you will learn Learn core concepts such as RDDs, DataFrames, transformations, and more Set up a Spark development environment Choose the right APIs for your applications Understand Spark's architecture and the execution flow of a Spark application Explore built-in modules for SQL, streaming, ML, and graph analysis Optimize your Spark job for better performance Who this book is for If you are a big data enthusiast and love processing huge amount of data, this book is for you. If you are data engineer and looking for the best optimization techniques for your Spark applications, then you will find this book helpful. This book also helps data scientists who want to implement their mach...