1. Record Nr. UNINA9910828209303321

Autore	DeRoos Dirk
Titolo	Hadoop for dummies / / by Dirk deRoos [and four others]
Pubbl/distr/stampa	Hoboken, New Jersey : , : John Wiley & Sons, , 2014 ©2014
ISBN	1-118-70503-3
Descrizione fisica	1 online resource (411 p.)
Collana	For Dummies
Disciplina	005.74
Soggetti	File organization (Computer science) - Computer programs Electronic data processing - Distributed processing
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	"Making everything easier"--Cover. Includes index.
Nota di contenuto	Contents at a Glance; Table of Contents; Introduction; About this Book; Foolish Assumptions; How This Book Is Organized; Icons Used in This Book; Beyond the Book; Where to Go from Here; Part I: Getting Started with Hadoop; Chapter 1: Introducing Hadoop and Seeing What It's Good For; Big Data and the Need for Hadoop; The Origin and Design of Hadoop; Examining the Various Hadoop Offerings; Chapter 2: Common Use Cases for Big Data in Hadoop; The Keys to Successfully Adopting Hadoop (Or, "Please, Can We Keep Him?"); Log Data Analysis; Data Warehouse Modernization; Fraud Detection; Risk Modeling Social Sentiment AnalysisImage Classification; Graph Analysis; To Infinity and Beyond; Chapter 3: Setting Up Your Hadoop Environment; Choosing a Hadoop Distribution; Choosing a Hadoop Cluster Architecture; The Hadoop For Dummies Environment; Your First Hadoop Program: Hello Hadoop!; Part II: How Hadoop Works; Chapter 4: Storing Data in Hadoop: The Hadoop Distributed File System; Data Storage in HDFS; Sketching Out the HDFS Architecture; HDFS Federation; HDFS High Availability; Chapter 5: Reading and Writing Data; Compressing Data; Managing Files with the Hadoop File System Commands Ingesting Log Data with FlumeChapter 6: MapReduce Programming; Thinking in Parallel; Seeing the Importance of MapReduce; Doing

Things in Parallel: Breaking Big Problems into Many Bite-Size Pieces; Writing MapReduce Applications; Getting Your Feet Wet: Writing a Simple MapReduce Application; Chapter 7: Frameworks for Processing Data in Hadoop: YARN and MapReduce; Running Applications Before Hadoop 2; Seeing a World beyond MapReduce; Real-Time and Streaming Applications; Chapter 8: Pig: Hadoop Programming Made Easier; Admiring the Pig Architecture; Going with the Pig Latin Application Flow
Working through the ABCs of Pig LatinEvaluating Local and Distributed Modes of Running Pig scripts; Checking Out the Pig Script Interfaces; Scripting with Pig Latin; Chapter 9: Statistical Analysis in Hadoop; Pumping Up Your Statistical Analysis; Machine Learning with Mahout; R on Hadoop; Chapter 10: Developing and Scheduling Application Workflows with Oozie; Getting Oozie in Place; Developing and Running an Oozie Workflow; Scheduling and Coordinating Oozie Workflows; Part III: Hadoop and Structured Data; Chapter 11: Hadoop and the Data Warehouse: Friends or Foes?
Comparing and Contrasting Hadoop with Relational DatabasesModernizing the Warehouse with Hadoop; Chapter 12: Extremely Big Tables: Storing Data in HBase; Say Hello to HBase; Understanding the HBase Data Model; Understanding the HBase Architecture; Taking HBase for a Test Run; Getting Things Done with HBase; HBase and the RDBMS world; Deploying and Tuning HBase; Chapter 13: Applying Structure to Hadoop Data with Hive; Saying Hello to Hive; Seeing How the Hive is Put Together; Getting Started with Apache Hive; Examining the Hive Clients; Working with Hive Data Types Creating and Managing Databases and Tables

Sommario/riassunto

Let Hadoop For Dummies help harness the power of your data and rein in the information overloadBig data has become big business, and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed. Enter Hadoop and this easy-to-understand For Dummies guide. Hadoop For Dummies helps readers understand the value of big data, make a business case for using Hadoop, navigate the Hadoop ecosystem, and build and manage Hadoop applications and clusters.</p