1. Record Nr. UNINA9910787510003321

Autore	Lublinsky Boris
Titolo	Professional hadoop solutions / / Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich
Pubbl/distr/stampa	Indianapolis, IN : , : John Wiley and Sons, , [2013] ©2013
ISBN	1-118-82418-0 1-118-61254-X
Edizione	[1st edition]
Descrizione fisica	1 online resource (506 p.)
Collana	Wrox Programmer to programmer
Altri autori (Persone)	SmithKevin T YakubovichAlexey
Disciplina	005.74
Soggetti	Electronic data processing - Distributed processing File organization (Computer science) Cloud computing
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Professional Hadoop® Solutions; Copyright; Credits; About the Authors; About the Technical Editors; Acnowledgments; Contents; Introduction; Who This Book Is For; What This Book Covers; How This Book Is Structured; What You Need to Use This Book; Conventions; Source Code; Errata; P2P.Wrox.Com; Chapter 1: Big Data and the Hadoop Ecosystem; Big Data Meets Hadoop; Hadoop: Meeting the Big Data Challenge; Data Science in the Business World; The Hadoop Ecosystem; Hadoop Core Components; Hadoop Distributions; Developing Enterprise Applications with Hadoop; Summary; Chapter 2: Storing Data in Hadoop HDFSHDFS Architecture; Using HDFS Files; Hadoop-Specific File Types; HDFS Federation and High Availability; HBase; HBase Architecture; HBase Schema Design; Programming for HBase; New HBase Features; Combining HDFS and HBase for Effective Data Storage; Using Apache Avro; Managing Metadata with HCatalog; Choosing an Appropriate Hadoop Data Organization for Your Applications; Summary; Chapter 3: Processing Your Data with MapReduce; Getting to Know MapReduce; MapReduce Execution Pipeline; Runtime Coordination and Task

Management in MapReduce; Your First MapReduce Application
Building and Executing MapReduce ProgramsDesigning MapReduce Implementations; Using MapReduce as a Framework for Parallel Processing; Simple Data Processing with MapReduce; Building Joins with MapReduce; Building Iterative MapReduce Applications; To MapReduce or Not to MapReduce?; Common MapReduce Design Gotchas; Summary; Chapter 4: Customizing MapReduce Execution; Controlling MapReduce Execution with InputFormat; Implementing InputFormat for Compute-Intensive Applications; Implementing InputFormat to Control the Number of Maps; Implementing InputFormat for Multiple HBase Tables Reading Data Your Way with Custom RecordReadersImplementing a Queue-Based RecordReader; Implementing RecordReader for XML Data; Organizing Output Data with Custom Output Formats; Implementing OutputFormat for Splitting MapReduce Job's Output into Multiple Directories; Writing Data Your Way with Custom RecordWriters; Implementing a RecordWriter to Produce Output tar Files; Optimizing Your MapReduce Execution with a Combiner; Controlling Reducer Execution with Partitioners; Implementing a Custom Partitioner for One-to-Many Joins; Using Non-Java Code with Hadoop; Pipes; Hadoop Streaming
Using JNISummary; Chapter 5: Building Reliable MapReduce Apps; Unit Testing MapReduce Applications; Testing Mappers; Testing Reducers; Integration Testing; Local Application Testing with Eclipse; Using Logging for Hadoop Testing; Processing Applications Logs; Reporting Metrics with Job Counters; Defensive Programming in MapReduce; Summary; Chapter 6: Automating Data Processing with Oozie; Getting to Know Oozie; Oozie Workflow; Executing Asynchronous Activities in Oozie Workflow; Oozie Recovery Capabilities; Oozie Workflow Job Life Cycle; Oozie Coordinator; Oozie Bundle
Oozie Parameterization with Expression Language

Sommario/riassunto

The go-to guidebook for deploying Big Data solutions with Hadoop Today's enterprise architects need to understand how the Hadoop frameworks and APIs fit together, and how they can be integrated to deliver real-world solutions. This book is a practical, detailed guide to building and implementing those solutions, with code-level instruction in the popular Wrox tradition. It covers storing data with HDFS and Hbase, processing data with MapReduce, and automating data processing with Oozie. Hadoop security, running Hadoop with Amazon Web Services, best practices, and automating Hadoop