1. Record Nr. UNINA9910466092903321

Autore	Ankam Venkat
Titolo	Big data analytics : a handy reference guide for data analysts and data scientists to help obtain value from big data analytics using Spark on Hadoop clusters / / Venkat Ankam
Pubbl/distr/stampa	Birmingham, England : , : Packt Publishing, , 2016 ©2016
Edizione	[1st edition]
Descrizione fisica	1 online resource (326 pages) : illustrations
Disciplina	005.8
Soggetti	Big data - Security measures Electronic books.
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Sommario/riassunto	A handy reference guide for data analysts and data scientists to help to obtain value from big data analytics using Spark on Hadoop clusters About This Book This book is based on the latest 2.0 version of Apache Spark and 2.7 version of Hadoop integrated with most commonly used tools. Learn all Spark stack components including latest topics such as DataFrames, DataSets, GraphFrames, Structured Streaming, DataFrame based ML Pipelines and SparkR. Integrations with frameworks such as HDFS, YARN and tools such as Jupyter, Zeppelin, NiFi, Mahout, HBase Spark Connector, GraphFrames, H2O and Hivemall. Who This Book Is For Though this book is primarily aimed at data analysts and data scientists, it will also help architects, programmers, and practitioners. Knowledge of either Spark or Hadoop would be beneficial. It is assumed that you have basic programming background in Scala, Python, SQL, or R programming with basic Linux experience. Working experience within big data environments is not mandatory. What You Will Learn Find out and implement the tools and techniques of big data analytics using Spark on Hadoop clusters with wide variety of tools used with Spark and Hadoop Understand all the Hadoop and Spark ecosystem components Get to know all the Spark components: Spark Core, Spark

SQL, DataFrames, DataSets, Conventional and Structured Streaming, MLLib, ML Pipelines and Graphx See batch and real-time data analytics using Spark Core, Spark SQL, and Conventional and Structured Streaming Get to grips with data science and machine learning using MLLib, ML Pipelines, H2O, Hivemall, Graphx, SparkR and Hivemall. In Detail Big Data Analytics book aims at providing the fundamentals of Apache Spark and Hadoop. All Spark components ? Spark Core, Spark SQL, DataFrames, Data sets, Conventional Streaming, Structured Streaming, MLlib, Graphx and Hadoop core components ? HDFS, MapReduce and Yarn are explored in greater depth with implementation examples on Spark + Hadoop clusters. It is moving away from MapReduce to Spark. So, advantages of Spark over MapReduce are explained at great depth to reap benefits of in-memory speeds. DataFrames API, Data Sources API and new Data set API are explained for building Big Data analytical applications. Real-time data analytics using Spark Streaming with Apache Kafka and HBase is covered to help building streaming applications. New Structured streaming concept is explained with an IOT (Internet of Things) use case. Machine learni...

2. Record Nr. UNISA996465425703316

Titolo	Advances in Databases and Information Systems [[electronic resource]] : 8th East European Conference, ADBIS 2004, Budapest, Hungary, September 22-25, 2004, Proceedings / / edited by Georg Gottlob, Andras Benczur, Janos Demetrovics
Pubbl/distr/stampa	Berlin, Heidelberg : , : Springer Berlin Heidelberg : , : Imprint : Springer, , 2004
ISBN	3-540-30204-2
Edizione	[1st ed. 2004.]
Descrizione fisica	1 online resource (XI, 426 p.)
Collana	Lecture Notes in Computer Science, , 0302-9743 ; ; 3255
Disciplina	005.74
Soggetti	Data structures (Computer science) Database management Information storage and retrieval Application software Multimedia information systems User interfaces (Computer systems) Data Structures and Information Theory Database Management Information Storage and Retrieval Information Systems Applications (incl. Internet) Multimedia Information Systems User Interfaces and Human Computer Interaction
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Bibliographic Level Mode of Issuance: Monograph
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Constraint Databases -- Quantifier-Elimination for the First-Order Theory of Boolean Algebras with Linear Cardinality Constraints -- Deductive Databases -- Update Propagation in Deductive Databases Using Soft Stratification -- Heterogenous and Web Information Systems -- Query Rewriting Using Views in a Typed Mediator Environment -- Reasoning About Web Information Systems Using Story Algebras -- Cross Enterprise Information Systems -- Component Framework for Strategic Supply Network Development -- Knowledge Discovery -- An

Abstract Algebra for Knowledge Discovery in Databases -- Database Modelling -- Beyond Databases: An Asset Language for Conceptual Content Management -- Component-Based Modeling of Huge Databases -- Cognitive Load Effects on End User Understanding of Conceptual Models: An Experimental Analysis -- Template Based, Designer Driven Design Pattern Instantiation Support -- XML and Semistructured Databases -- A High-Level Language for Specifying XML Data Transformations -- Implementing a Query Language for Context-Dependent Semistructured Data -- Static Analysis of Structural Recursion in Semistructured Databases and Its Consequences -- Physical Database Design and Query Evaluation -- Catalogues from a New Perspective: A Data Structure for Physical Organisation -- Database Caching – Towards a Cost Model for Populating Cache Groups -- Towards Quadtree-Based Moving Objects Databases -- A Content-Based Music Retrieval System Using Multidimensional Index of Time-Sequenced Representative Melodies from Music Database -- Solving Stochastic Optimization in Distributed Databases Using Genetic Algorithms -- Transaction Management and Workflow Systems -- ML-1-2PC: An Adaptive Multi-level Atomic Commit Protocol -- Making More Out of an Inconsistent Database -- Process Query Language: A Way to Make Workflow Processes More Flexible -- Triggering Replanning in an Integrated Workflow Planning and Enactment System -- Query Processing and Data Streams -- Grouped Processing of Relational Algebra Expressions over Data Streams -- Processing Sliding Window Join Aggregate in Continuous Queries over Data Streams -- Spatial Databases -- How to Integrate Heterogeneous Spatial Databases in a Consistent Way? -- Vague Spatial Data Types, Set Operations, and Predicates -- Agents and Mobile Systems -- Intelligent Multi-agent Based Database Hybrid Intrusion Prevention System -- Energy Efficient Transaction Processing in Mobile Broadcast Environments.

Sommario/riassunto

This book constitutes the refereed proceedings of the 8th East European Conference on Advances in Databases and Information Systems, ADBIS 2004, held in Budapest, Hungary, in September 2004. The 27 revised full papers presented together with an invited paper were carefully reviewed and selected from 130 submissions. The papers are organized in topical sections on constraint databases, deductive databases, heterogenous and Web information systems, cross enterprise information systems, knowledge discovery, database modeling, XML and semistructured databases, physical database design and query evaluation, transaction management and workflow systems, query processing and data streams, spatial databases, and agents and mobile systems.