
Hadoop In Action Manning Publications

Hadoop in Practice
Grokking Algorithms
Spark in Action
Tika in Action
Think Like a Data Scientist
Big Data
Spark GraphX in Action
Kafka Streams in Action
Rust in Action
Spark: The Definitive Guide
Hadoop in Action
Mongodb in Action
Storm Applied
Data Algorithms
R in Action
Machine Learning in Action
Mastering Large Datasets
Solr in Action
Elasticsearch in Action
Data Pipelines with Apache Airflow
Hadoop in Action
Making Sense of NoSQL
Scala in Action
Mahout in Action
Spark in Action
HBase in Action
Flink in Action
Designing Cloud Data Platforms
Collective Intelligence in Action
Data-Intensive Text Processing with MapReduce
Hadoop in Action
Kafka: The Definitive Guide
HADOOP IN ACTION
Mining of Massive Datasets
Big Data Processing with Apache Spark
OAuth 2 in Action
Kafka in Action
Lucene in Action

Data Analysis with Python and PySpark
Clojure in Action

Hadoop In Action Manning Publications

Downloaded from intra.itu.edu by guest

NOELLE MARSHALL

Hadoop in Practice Manning Publications

Summary Hadoop in Practice, Second Edition provides over 100 tested, instantly useful techniques that will help you conquer big data, using Hadoop. This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2. Brand new chapters cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop. You'll also get new and updated techniques for Flume, Sqoop, and Mahout, all of which have seen major new versions recently. In short, this is the most practical, up-to-date coverage of Hadoop available anywhere. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book It's always a good time to upgrade your Hadoop skills! Hadoop in Practice, Second Edition provides a collection of 104 tested, instantly useful techniques for analyzing real-time streams, moving data securely, machine learning, managing large-scale clusters, and taming big data using Hadoop. This completely revised edition covers changes and new features in Hadoop core, including MapReduce 2 and YARN. You'll pick up hands-on best practices for integrating Spark, Kafka, and Impala with Hadoop, and get new and updated techniques for the latest versions of Flume, Sqoop, and Mahout. In short, this is the most practical, up-to-date coverage of Hadoop available. Readers need to know a programming language like Java and have basic familiarity with Hadoop. What's Inside Thoroughly updated for Hadoop 2 How to write YARN applications Integrate real-time technologies like Storm, Impala, and Spark Predictive analytics using Mahout and RR Readers need to know a programming language like Java and have basic familiarity with Hadoop. About the Author Alex Holmes works on tough big-data problems. He is a software engineer, author, speaker, and blogger specializing in large-scale Hadoop projects. Table of Contents PART 1 BACKGROUND AND FUNDAMENTALS Hadoop in a heartbeat Introduction to YARN PART 2 DATA LOGISTICS Data serialization—working with text and beyond Organizing and optimizing data in HDFS Moving data into and out of Hadoop PART 3 BIG DATA PATTERNS Applying MapReduce patterns to big data Utilizing data structures and algorithms at scale Tuning, debugging, and testing PART 4 BEYOND MAPREDUCE SQL on Hadoop Writing a YARN application

Grokking Algorithms Lulu.com

"Provides pragmatic guidance on what to do ... and what not to do." - From the Foreword by Ian Glazer, Salesforce OAuth 2 in Action teaches you the practical use and deployment of this HTTP-based protocol from the perspectives of a client, authorization server, and resource server. You'll learn how to confidently and securely build and deploy OAuth on both the client and server sides. Foreword by Ian Glazer. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Think of OAuth 2 as the web version of a valet key. It is an HTTP-based security protocol that allows users of a service to enable applications to use that service on their behalf without handing over full control. And OAuth is used everywhere,

from Facebook and Google, to startups and cloud services. About the Book OAuth 2 in Action teaches you practical use and deployment of OAuth 2 from the perspectives of a client, an authorization server, and a resource server. You'll begin with an overview of OAuth and its components and interactions. Next, you'll get hands-on and build an OAuth client, an authorization server, and a protected resource. Then you'll dig into tokens, dynamic client registration, and more advanced topics. By the end, you'll be able to confidently and securely build and deploy OAuth on both the client and server sides. What's Inside Covers OAuth 2 protocol and design Authorization with OAuth 2 OpenID Connect and User-Managed Access Implementation risks JOSE, introspection, revocation, and registration Protecting and accessing REST APIs About the Reader Readers need basic programming skills and knowledge of HTTP and JSON. About the Author Justin Richer is a systems architect and software engineer. Antonio Sanso is a security software engineer and a security researcher. Both authors contribute to open standards and open source. Table of Contents What is OAuth 2.0 and why should you care? The OAuth dance Building a simple OAuth client Building a simple OAuth protected resource Building a simple OAuth authorization server OAuth 2.0 in the real world Common client vulnerabilities Common protected resources vulnerabilities Common authorization server vulnerabilities Common OAuth token vulnerabilities OAuth tokens Dynamic client registration User authentication with OAuth 2.0 Protocols and profiles using OAuth 2.0 Beyond bearer tokens Summary and conclusions Part 1 - First steps Part 2 - Building an OAuth 2 environment Part 3 - OAuth 2 implementation and vulnerabilities Part 4 - Taking OAuth further *Spark in Action* Manning Publications

Summary Elasticsearch in Action teaches you how to build scalable search applications using Elasticsearch. You'll ramp up fast, with an informative overview and an engaging introductory example. Within the first few chapters, you'll pick up the core concepts you need to implement basic searches and efficient indexing. With the fundamentals well in hand, you'll go on to gain an organized view of how to optimize your design. Perfect for developers and administrators building and managing search-oriented applications. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Modern search seems like magic—you type a few words and the search engine appears to know what you want. With the Elasticsearch real-time search and analytics engine, you can give your users this magical experience without having to do complex low-level programming or understand advanced data science algorithms. You just install it, tweak it, and get on with your work. About the Book Elasticsearch in Action teaches you how to write applications that deliver professional quality search. As you read, you'll learn to add basic search features to any application, enhance search results with predictive analysis and relevancy ranking, and use saved data from prior searches to give users a custom experience. This practical book focuses on Elasticsearch's REST API via HTTP. Code snippets are written mostly in bash using cURL, so they're easily translatable to other languages. What's Inside What is a great search application? Building scalable search solutions Using Elasticsearch with any language Configuration and tuning About the Reader For developers and administrators building and

managing search-oriented applications. About the Authors Radu Gheorghe is a search consultant and software engineer. Matthew Lee Hinman develops highly available, cloud-based systems. Roy Russo is a specialist in predictive analytics. Table of Contents PART 1 CORE ELASTICSEARCH FUNCTIONALITY Introducing Elasticsearch Diving into the functionality Indexing, updating, and deleting data Searching your data Analyzing your data Searching with relevancy Exploring your data with aggregations Relations among documents PART 2 ADVANCED ELASTICSEARCH FUNCTIONALITY Scaling out Improving performance Administering your cluster

Tika in Action Simon and Schuster

Summary A fully revised edition that covers the new features available in Clojure 1.6. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Clojure is a modern Lisp for the JVM. It has the strengths you expect: first-class functions, macros, and Lisp's clean programming style. It supports functional programming, making it ideal for concurrent programming and for creating domain-specific languages. Clojure lets you solve harder problems, make faster changes, and end up with a smaller code base. It's no wonder that there are so many Clojure success stories. About the Book Clojure in Action, Second Edition is an expanded and improved version that's been updated to cover the new features of Clojure 1.6. The book gives you a rapid introduction to the Clojure language, moving from abstract theory to practical examples. You'll start by learning how to use Clojure as a general-purpose language. Next, you'll explore Clojure's efficient concurrency model, based on the database concept of Software Transactional Memory (STM). You'll gain a new level of productivity through Clojure DSLs that can run on the JVM. Along the way, you'll learn countless tips, tricks, and techniques for writing smaller, safer, and faster code. What's Inside Functional programming basics Metaprogramming with Clojure's macros Interoperating with Java Covers Clojure 1.6 About the Reader Assumes readers are familiar with a programming language like C, Java, Ruby, or Python. Table of Contents Introducing Clojure Clojure elements: Data structures and functions Building blocks of Clojure Multimethod polymorphism Exploring Clojure and Java interop State and the concurrent world Evolving Clojure through macros More on functional programming Protocols, records, and types Test-driven development and more More macros and DSL

Think Like a Data Scientist Simon and Schuster

Summary Tika in Action is a hands-on guide to content mining with Apache Tika. The book's many examples and case studies offer real-world experience from domains ranging from search engines to digital asset management and scientific data processing. About the Technology Tika is an Apache toolkit that has built into it everything you and your app need to know about file formats. Using Tika, your applications can discover and extract content from digital documents in almost any format, including exotic ones. About this Book Tika in Action is the ultimate guide to content mining using Apache Tika. You'll learn how to pull usable information from otherwise inaccessible sources, including internet media and file archives. This example-rich book teaches you to build and extend applications based on real-world experience with search engines, digital asset management, and scientific data processing. In addition to architectural overviews, you'll find detailed chapters on features like metadata extraction, automatic language detection, and custom parser development. This book is written for developers who are new to both Scala and Lift and covers just enough Scala

to get you started. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. What's Inside Crack MS Word, PDF, HTML, and ZIP Integrate with search engines, CMS, and other data sources Learn through experimentation Many examples This book requires no previous knowledge of Tika or text mining techniques. It assumes a working knowledge of Java.

=====
Table of Contents PART 1 GETTING STARTED The case for the digital Babel fish Getting started with Tika The information landscape PART 2 TIKA IN DETAIL Document type detection Content extraction Understanding metadata Language detection What's in a file? PART 3 INTEGRATION AND ADVANCED USE The big picture Tika and the Lucene search stack Extending Tika PART 4 CASE STUDIES Powering NASA science data systems Content management with Apache Jackrabbit Curating cancer research data with Tika The classic search engine example

Big Data Simon and Schuster

There's a great deal of wisdom in a crowd, but how do you listen to a thousand people talking at once? Identifying the wants, needs, and knowledge of internet users can be like listening to a mob. In the Web 2.0 era, leveraging the collective power of user contributions, interactions, and feedback is the key to market dominance. A new category of powerful programming techniques lets you discover the patterns, inter-relationships, and individual profiles--the collective intelligence--locked in the data people leave behind as they surf websites, post blogs, and interact with other users. Collective Intelligence in Action is a hands-on guidebook for implementing collective intelligence concepts using Java. It is the first Java-based book to emphasize the underlying algorithms and technical implementation of vital data gathering and mining techniques like analyzing trends, discovering relationships, and making predictions. It provides a pragmatic approach to personalization by combining content-based analysis with collaborative approaches. This book is for Java developers implementing Collective Intelligence in real, high-use applications. Following a running example in which you harvest and use information from blogs, you learn to develop software that you can embed in your own applications. The code examples are immediately reusable and give the Java developer a working collective intelligence toolkit. Along the way, you work with, a number of APIs and open-source toolkits including text analysis and search using Lucene, web-crawling using Nutch, and applying machine learning algorithms using WEKA and the Java Data Mining (JDM) standard. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book.

Spark GraphX in Action Simon and Schuster

The massive datasets required for most modern businesses are too large to safely store and efficiently process on a single server. Hadoop is an open source data processing framework that provides a distributed file system that can manage data stored across clusters of servers and implements the MapReduce data processing model so that users can effectively query and utilize big data. The new Hadoop 2.0 is a stable, enterprise-ready platform supported by a rich ecosystem of tools and related technologies such as Pig, Hive, YARN, Spark, Tez, and many more. Hadoop in Action, Second Edition, provides a comprehensive introduction to Hadoop and shows how to write programs in the MapReduce style. It starts with a few easy examples and then moves quickly to

show how Hadoop can be used in more complex data analysis tasks. It covers how YARN, new in Hadoop 2, simplifies and supercharges resource management to make streaming and real-time applications more feasible. Included are best practices and design patterns of MapReduce programming. The book expands on the first edition by enhancing coverage of important Hadoop 2 concepts and systems, and by providing new chapters on data management and data science that reinforce a practical understanding of Hadoop. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

Kafka Streams in Action Simon and Schuster

Now in its second edition, this book focuses on practical algorithms for mining data from even the largest datasets.

Rust in Action Simon and Schuster

Summary Storm Applied is a practical guide to using Apache Storm for the real-world tasks associated with processing and analyzing real-time data streams. This immediately useful book starts by building a solid foundation of Storm essentials so that you learn how to think about designing Storm solutions the right way from day one. But it quickly dives into real-world case studies that will bring the novice up to speed with productionizing Storm. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. Summary Storm Applied is a practical guide to using Apache Storm for the real-world tasks associated with processing and analyzing real-time data streams. This immediately useful book starts by building a solid foundation of Storm essentials so that you learn how to think about designing Storm solutions the right way from day one. But it quickly dives into real-world case studies that will bring the novice up to speed with productionizing Storm. About the Technology It's hard to make sense out of data when it's coming at you fast. Like Hadoop, Storm processes large amounts of data but it does it reliably and in real time, guaranteeing that every message will be processed. Storm allows you to scale with your data as it grows, making it an excellent platform to solve your big data problems. About the Book Storm Applied is an example-driven guide to processing and analyzing real-time data streams. This immediately useful book starts by teaching you how to design Storm solutions the right way. Then, it quickly dives into real-world case studies that show you how to scale a high-throughput stream processor, ensure smooth operation within a production cluster, and more. Along the way, you'll learn to use Trident for stateful stream processing, along with other tools from the Storm ecosystem. This book moves through the basics quickly. While prior experience with Storm is not assumed, some experience with big data and real-time systems is helpful. What's Inside Mapping real problems to Storm components Performance tuning and scaling Practical troubleshooting and debugging Exactly-once processing with Trident About the Authors Sean Allen, Matthew Jankowski, and Peter Pathirana lead the development team for a high-volume, search-intensive commercial web application at TheLadders. Table of Contents Introducing Storm Core Storm concepts Topology design Creating robust topologies Moving from local to remote topologies Tuning in Storm Resource contention Storm internals Trident

Spark: The Definitive Guide "O'Reilly Media, Inc."

R is a powerful language for statistical computing and graphics that can handle virtually any data-crunching task. It runs on all important platforms and provides thousands of useful specialized

modules and utilities. This makes R a great way to get meaningful information from mountains of raw data. R in Action, Second Edition is a language tutorial focused on practical problems. Written by a research methodologist, it takes a direct and modular approach to quickly give readers the information they need to produce useful results. Focusing on realistic data analyses and a comprehensive integration of graphics, it follows the steps that real data analysts use to acquire their data, get it into shape, analyze it, and produce meaningful results that they can provide to clients. Purchase of the print book comes with an offer of a free PDF eBook from Manning. Also available is all code from the book.

Hadoop in Action Simon and Schuster

"This well-written book will help you make the most of what Rust has to offer." - Ramnivas Laddad, author of AspectJ in Action Rust in Action is a hands-on guide to systems programming with Rust. Written for inquisitive programmers, it presents real-world use cases that go far beyond syntax and structure. Summary Rust in Action introduces the Rust programming language by exploring numerous systems programming concepts and techniques. You'll be learning Rust by delving into how computers work under the hood. You'll find yourself playing with persistent storage, memory, networking and even tinkering with CPU instructions. The book takes you through using Rust to extend other applications and teaches you tricks to write blindingly fast code. You'll also discover parallel and concurrent programming. Filled to the brim with real-life use cases and scenarios, you'll go beyond the Rust syntax and see what Rust has to offer in real-world use cases. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Rust is the perfect language for systems programming. It delivers the low-level power of C along with rock-solid safety features that let you code fearlessly. Ideal for applications requiring concurrency, Rust programs are compact, readable, and blazingly fast. Best of all, Rust's famously smart compiler helps you avoid even subtle coding errors. About the book Rust in Action is a hands-on guide to systems programming with Rust. Written for inquisitive programmers, it presents real-world use cases that go far beyond syntax and structure. You'll explore Rust implementations for file manipulation, networking, and kernel-level programming and discover awesome techniques for parallelism and concurrency. Along the way, you'll master Rust's unique borrow checker model for memory management without a garbage collector. What's inside Elementary to advanced Rust programming Practical examples from systems programming Command-line, graphical and networked applications About the reader For intermediate programmers. No previous experience with Rust required. About the author Tim McNamara uses Rust to build data processing pipelines and generative art. He is an expert in natural language processing and data engineering. Table of Contents 1 Introducing Rust PART 1 RUST LANGUAGE DISTINCTIVES 2 Language foundations 3 Compound data types 4 Lifetimes, ownership, and borrowing PART 2 DEMYSTIFYING SYSTEMS PROGRAMMING 5 Data in depth 6 Memory 7 Files and storage 8 Networking 9 Time and timekeeping 10 Processes, threads, and containers 11 Kernel 12 Signals, interrupts, and exceptions

Mongodb in Action Simon and Schuster

Our world is being revolutionized by data-driven methods: access to large amounts of data has generated new insights and opened exciting new opportunities in commerce, science, and

computing applications. Processing the enormous quantities of data necessary for these advances requires large clusters, making distributed computing paradigms more crucial than ever. MapReduce is a programming model for expressing distributed computations on massive datasets and an execution framework for large-scale data processing on clusters of commodity servers. The programming model provides an easy-to-understand abstraction for designing scalable algorithms, while the execution framework transparently handles many system-level details, ranging from scheduling to synchronization to fault tolerance. This book focuses on MapReduce algorithm design, with an emphasis on text processing algorithms common in natural language processing, information retrieval, and machine learning. We introduce the notion of MapReduce design patterns, which represent general reusable solutions to commonly occurring problems across a variety of problem domains. This book not only intends to help the reader "think in MapReduce", but also discusses limitations of the programming model as well. Table of Contents: Introduction / MapReduce Basics / MapReduce Algorithm Design / Inverted Indexing for Text Retrieval / Graph Algorithms / EM Algorithms for Text Processing / Closing Remarks

Storm Applied Springer Nature

This book teaches you how to build and maintain effective data pipelines. You'll explore the most common usage patterns, including aggregating multiple data sources, connecting to and from data lakes, and cloud deployment. --

Data Algorithms Simon and Schuster

Summary Machine Learning in Action is unique book that blends the foundational theories of machine learning with the practical realities of building tools for everyday data analysis. You'll use the flexible Python programming language to build programs that implement algorithms for data classification, forecasting, recommendations, and higher-level features like summarization and simplification. About the Book A machine is said to learn when its performance improves with experience. Learning requires algorithms and programs that capture data and ferret out the interesting or useful patterns. Once the specialized domain of analysts and mathematicians, machine learning is becoming a skill needed by many. Machine Learning in Action is a clearly written tutorial for developers. It avoids academic language and takes you straight to the techniques you'll use in your day-to-day work. Many (Python) examples present the core algorithms of statistical data processing, data analysis, and data visualization in code you can reuse. You'll understand the concepts and how they fit in with tactical tasks like classification, forecasting, recommendations, and higher-level features like summarization and simplification. Readers need no prior experience with machine learning or statistical processing. Familiarity with Python is helpful. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. What's Inside A no-nonsense introduction Examples showing common ML tasks Everyday data analysis Implementing classic algorithms like Apriori and Adaboos Table of Contents PART 1 CLASSIFICATION Machine learning basics Classifying with k-Nearest Neighbors Splitting datasets one feature at a time: decision trees Classifying with probability theory: naïve Bayes Logistic regression Support vector machines Improving classification with the AdaBoost meta algorithm PART 2 FORECASTING NUMERIC VALUES WITH REGRESSION Predicting numeric values: regression Tree-based regression PART 3 UNSUPERVISED LEARNING Grouping unlabeled items using

k-means clustering Association analysis with the Apriori algorithm Efficiently finding frequent itemsets with FP-growth PART 4 ADDITIONAL TOOLS Using principal component analysis to simplify data Simplifying data with the singular value decomposition Big data and MapReduce *R in Action* "O'Reilly Media, Inc."

Summary The Spark distributed data processing platform provides an easy-to-implement tool for ingesting, streaming, and processing data from any source. In Spark in Action, Second Edition, you'll learn to take advantage of Spark's core features and incredible processing speed, with applications including real-time computation, delayed evaluation, and machine learning. Spark skills are a hot commodity in enterprises worldwide, and with Spark's powerful and flexible Java APIs, you can reap all the benefits without first learning Scala or Hadoop. Foreword by Rob Thomas. About the technology Analyzing enterprise data starts by reading, filtering, and merging files and streams from many sources. The Spark data processing engine handles this varied volume like a champ, delivering speeds 100 times faster than Hadoop systems. Thanks to SQL support, an intuitive interface, and a straightforward multilanguage API, you can use Spark without learning a complex new ecosystem. About the book Spark in Action, Second Edition, teaches you to create end-to-end analytics applications. In this entirely new book, you'll learn from interesting Java-based examples, including a complete data pipeline for processing NASA satellite data. And you'll discover Java, Python, and Scala code samples hosted on GitHub that you can explore and adapt, plus appendixes that give you a cheat sheet for installing tools and understanding Spark-specific terms. What's inside Writing Spark applications in Java Spark application architecture Ingestion through files, databases, streaming, and Elasticsearch Querying distributed datasets with Spark SQL About the reader This book does not assume previous experience with Spark, Scala, or Hadoop. About the author Jean-Georges Perrin is an experienced data and software architect. He is France's first IBM Champion and has been honored for 12 consecutive years. Table of Contents PART 1 - THE THEORY CRIPPLED BY AWESOME EXAMPLES 1 So, what is Spark, anyway? 2 Architecture and flow 3 The majestic role of the dataframe 4 Fundamentally lazy 5 Building a simple app for deployment 6 Deploying your simple app PART 2 - INGESTION 7 Ingestion from files 8 Ingestion from databases 9 Advanced ingestion: finding data sources and building your own 10 Ingestion through structured streaming PART 3 - TRANSFORMING YOUR DATA 11 Working with SQL 12 Transforming your data 13 Transforming entire documents 14 Extending transformations with user-defined functions 15 Aggregating your data PART 4 - GOING FURTHER 16 Cache and checkpoint: Enhancing Spark's performances 17 Exporting data and building full data pipelines 18 Exploring deployment

Machine Learning in Action Simon and Schuster

Summary Spark GraphX in Action starts out with an overview of Apache Spark and the GraphX graph processing API. This example-based tutorial then teaches you how to configure GraphX and how to use it interactively. Along the way, you'll collect practical techniques for enhancing applications and applying machine learning algorithms to graph data. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology GraphX is a powerful graph processing API for the Apache Spark analytics engine that lets you draw insights from large datasets. GraphX gives you unprecedented speed and capacity for running massively parallel and machine learning algorithms. About the Book Spark GraphX in Action begins with the big

picture of what graphs can be used for. This example-based tutorial teaches you how to use GraphX interactively. You'll start with a crystal-clear introduction to building big data graphs from regular data, and then explore the problems and possibilities of implementing graph algorithms and architecting graph processing pipelines. Along the way, you'll collect practical techniques for enhancing applications and applying machine learning algorithms to graph data. What's Inside Understanding graph technology Using the GraphX API Developing algorithms for big graphs Machine learning with graphs Graph visualization About the Reader Readers should be comfortable writing code. Experience with Apache Spark and Scala is not required. About the Authors Michael Malak has worked on Spark applications for Fortune 500 companies since early 2013. Robin East has worked as a consultant to large organizations for over 15 years and is a data scientist at Worldpay. Table of Contents PART 1 SPARK AND GRAPHS Two important technologies: Spark and graphs GraphX quick start Some fundamentals PART 2 CONNECTING VERTICES GraphX Basics Built-in algorithms Other useful graph algorithms Machine learning PART 3 OVER THE ARC The missing algorithms Performance and monitoring Other languages and tools

Mastering Large Datasets Simon and Schuster

Centralized data warehouses, the long-time defacto standard for housing data for analytics, are rapidly giving way to multi-faceted cloud data platforms. Companies that embrace modern cloud data platforms benefit from an integrated view of their business using all of their data and can take advantage of advanced analytic practices to drive predictions and as yet unimagined data services. Designing Cloud Data Platforms is an hands-on guide to envisioning and designing a modern scalable data platform that takes full advantage of the flexibility of the cloud. As you read, you'll learn the core components of a cloud data platform design, along with the role of key technologies like Spark and Kafka Streams. You'll also explore setting up processes to manage cloud-based data, keep it secure, and using advanced analytic and BI tools to analyse it. about the technology Access to affordable, dependable, serverless cloud services has revolutionized the way organizations can approach data management, and companies both big and small are raring to migrate to the cloud. But without a properly designed data platform, data in the cloud can remain just as siloed and inaccessible as it is today for most organizations. Designing Cloud Data Platforms lays out the principles of a well-designed platform that uses the scalable resources of the public cloud to manage all of an organization's data, and present it as useful business insights. about the book In Designing Cloud Data Platforms, you'll learn how to integrate data from multiple sources into a single, cloud-based, modern data platform. Drawing on their real-world experiences designing cloud data platforms for dozens of organizations, cloud data experts Danil Zburivsky and Lynda Partner take you through a six-layer approach to creating cloud data platforms that maximizes flexibility and manageability and reduces costs. Starting with foundational principles, you'll learn how to get data into your platform from different databases, files, and APIs, the essential practices for organizing and processing that raw data, and how to best take advantage of the services offered by major cloud vendors. As you progress past the basics you'll take a deep dive into advanced topics to get the most out of your data platform, including real-time data management, machine learning analytics, schema management, and more. what's inside The tools of different public cloud for implementing data platforms Best practices for managing structured and unstructured data sets Machine learning

tools that can be used on top of the cloud Cost optimization techniques about the reader For data professionals familiar with the basics of cloud computing and distributed data processing systems like Hadoop and Spark. about the authors Danil Zburivsky has over 10 years experience designing and supporting large-scale data infrastructure for enterprises across the globe. Lynda Partner is the VP of Analytics-as-a-Service at Pythian, and has been on the business side of data for over 20 years.

Solr in Action Manning Publications

"This book does the impossible: it makes math fun and easy!" - Sander Rossel, COAS Software Systems Grokking Algorithms is a fully illustrated, friendly guide that teaches you how to apply common algorithms to the practical problems you face every day as a programmer. You'll start with sorting and searching and, as you build up your skills in thinking algorithmically, you'll tackle more complex concerns such as data compression and artificial intelligence. Each carefully presented example includes helpful diagrams and fully annotated code samples in Python. Learning about algorithms doesn't have to be boring! Get a sneak peek at the fun, illustrated, and friendly examples you'll find in Grokking Algorithms on Manning Publications' YouTube channel. Continue your journey into the world of algorithms with Algorithms in Motion, a practical, hands-on video course available exclusively at Manning.com (www.manning.com/livevideo/algorithms-?in-motion). Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology An algorithm is nothing more than a step-by-step procedure for solving a problem. The algorithms you'll use most often as a programmer have already been discovered, tested, and proven. If you want to understand them but refuse to slog through dense multipage proofs, this is the book for you. This fully illustrated and engaging guide makes it easy to learn how to use the most important algorithms effectively in your own programs. About the Book Grokking Algorithms is a friendly take on this core computer science topic. In it, you'll learn how to apply common algorithms to the practical programming problems you face every day. You'll start with tasks like sorting and searching. As you build up your skills, you'll tackle more complex problems like data compression and artificial intelligence. Each carefully presented example includes helpful diagrams and fully annotated code samples in Python. By the end of this book, you will have mastered widely applicable algorithms as well as how and when to use them. What's Inside Covers search, sort, and graph algorithms Over 400 pictures with detailed walkthroughs Performance trade-offs between algorithms Python-based code samples About the Reader This easy-to-read, picture-heavy introduction is suitable for self-taught programmers, engineers, or anyone who wants to brush up on algorithms. About the Author Aditya Bhargava is a Software Engineer with a dual background in Computer Science and Fine Arts. He blogs on programming at adit.io. Table of Contents Introduction to algorithms Selection sort Recursion Quicksort Hash tables Breadth-first search Dijkstra's algorithm Greedy algorithms Dynamic programming K-nearest neighbors

Elasticsearch in Action Manning Publications

Master the wicked-fast Apache Kafka streaming platform through hands-on examples and real-world projects. In Kafka in Action you will learn: Understanding Apache Kafka concepts Setting up and executing basic ETL tasks using Kafka Connect Using Kafka as part of a large data project team Performing administrative tasks Producing and consuming event streams Working with Kafka from Java applications Implementing Kafka as a message queue Kafka in Action is a fast-paced

introduction to every aspect of working with Apache Kafka. Starting with an overview of Kafka's core concepts, you'll immediately learn how to set up and execute basic data movement tasks and how to produce and consume streams of events. Advancing quickly, you'll soon be ready to use Kafka in your day-to-day workflow, and start digging into even more advanced Kafka topics. About the technology Think of Apache Kafka as a high performance software bus that facilitates event streaming, logging, analytics, and other data pipeline tasks. With Kafka, you can easily build features like operational data monitoring and large-scale event processing into both large and small-scale applications. About the book Kafka in Action introduces the core features of Kafka, along with relevant examples of how to use it in real applications. In it, you'll explore the most common use cases such as logging and managing streaming data. When you're done, you'll be ready to handle both basic developer- and admin-based tasks in a Kafka-focused team. What's inside Kafka as an event streaming platform Kafka producers and consumers from Java applications Kafka as part of a large data project About the reader For intermediate Java developers or data engineers. No prior knowledge of Kafka required. About the author Dylan Scott is a software developer in the insurance industry. Viktor Gamov is a Kafka-focused developer advocate. At Confluent, Dave Klein helps developers, teams, and enterprises harness the power of event streaming with Apache Kafka. Table of Contents PART 1 GETTING STARTED 1 Introduction to Kafka 2 Getting to know Kafka PART 2 APPLYING KAFKA 3 Designing a Kafka project 4 Producers: Sourcing data 5 Consumers: Unlocking data 6 Brokers 7 Topics and partitions 8 Kafka storage 9 Management: Tools and logging PART 3 GOING FURTHER 10 Protecting Kafka 11 Schema registry 12 Stream processing with Kafka Streams and ksqldb

Data Pipelines with Apache Airflow Simon and Schuster

Best Sellers - Books :

- [My First Library : Boxset Of 10 Board Books For Kids](#)
- [Brown Bear, Brown Bear, What Do You See?](#)
- [How To Catch A Mermaid By Adam Wallace](#)
- [The Democrat Party Hates America By Mark R. Levin](#)
- [Haunting Adeline \(cat And Mouse Duet\) By H. D. Carlton](#)
- [A Letter From Your Teacher: On The First Day Of School](#)
- [Guess How Much I Love You](#)
- [Meditations: A New Translation By Marcus Aurelius](#)
- [Flash Cards: Sight Words](#)
- [Fourth Wing \(the Emyrean, 1\) By Rebecca Yarros](#)

Summary Spark in Action teaches you the theory and skills you need to effectively handle batch and streaming data using Spark. Fully updated for Spark 2.0. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Big data systems distribute datasets across clusters of machines, making it a challenge to efficiently query, stream, and interpret them. Spark can help. It is a processing system designed specifically for distributed data. It provides easy-to-use interfaces, along with the performance you need for production-quality analytics and machine learning. Spark 2 also adds improved programming APIs, better performance, and countless other upgrades. About the Book Spark in Action teaches you the theory and skills you need to effectively handle batch and streaming data using Spark. You'll get comfortable with the Spark CLI as you work through a few introductory examples. Then, you'll start programming Spark using its core APIs. Along the way, you'll work with structured data using Spark SQL, process near-real-time streaming data, apply machine learning algorithms, and munge graph data using Spark GraphX. For a zero-effort startup, you can download the preconfigured virtual machine ready for you to try the book's code. What's Inside Updated for Spark 2.0 Real-life case studies Spark DevOps with Docker Examples in Scala, and online in Java and Python About the Reader Written for experienced programmers with some background in big data or machine learning. About the Authors Petar Zečević and Marko Bonačić are seasoned developers heavily involved in the Spark community. Table of Contents PART 1 - FIRST STEPS Introduction to Apache Spark Spark fundamentals Writing Spark applications The Spark API in depth PART 2 - MEET THE SPARK FAMILY Sparkling queries with Spark SQL Ingesting data with Spark Streaming Getting smart with MLlib ML: classification and clustering Connecting the dots with GraphX PART 3 - SPARK OPS Running Spark Running on a Spark standalone cluster Running on YARN and Mesos PART 4 - BRINGING IT TOGETHER Case study: real-time dashboard Deep learning on Spark with H2O