

Apache Spark In 24 Hours Sams Teach Yourself Sams Teach Yourself In 24 Hours

Yeah, reviewing a ebook **Apache Spark In 24 Hours Sams Teach Yourself Sams Teach Yourself In 24 Hours** could mount up your near friends listings. This is just one of the solutions for you to be successful. As understood, realization does not recommend that you have astonishing points.

Comprehending as with ease as understanding even more than new will manage to pay for each success. bordering to, the broadcast as with ease as perspicacity of this Apache Spark In 24 Hours Sams Teach Yourself Sams Teach Yourself In 24 Hours can be taken as with ease as picked to act.

Beginning Rust - Carlo Milanese 2018-03-22

Learn to program with Rust in an easy, step-by-step manner on Unix, Linux shell, macOS and the Windows command line. As you read this book, you'll build on the knowledge you gained in previous chapters and see what Rust has to offer. Beginning Rust starts with the basics of Rust, including how to name objects, control execution flow, and handle primitive types. You'll see how to do arithmetic, allocate memory, use iterators, and handle input/output. Once you have mastered these core skills, you'll work on handling errors and using the object-oriented features of Rust to build robust Rust applications in no time. Only a basic knowledge of programming is required, preferably in C or C++. To understand this book, it's enough to know what integers and floating-point numbers are, and to distinguish identifiers from string literals. After reading this book, you'll be ready to build Rust applications. What You'll Learn Get started programming with Rust Understand heterogeneous data structures and data sequences Define functions, generic functions, structs, and more Work with closures, changeable strings, ranges and slices Use traits and learn about lifetimes Who This Book Is For Those who are new to Rust and who have at least some prior experience with programming in general: some C/C++ is recommended particularly.

Armored - Mark Greaney 2022-07-05

A novel inspired by #1 New York Times best-selling author Mark Greaney's Audible Original drama, Armored. Joshua Duffy is a Close Protection Agent—a professional bodyguard—and he's one of the world's elite operatives. That is, he was until his last mission in Lebanon. Against all odds, Josh got his primary out alive, but the cost was high. Josh lost his lower left leg. There's not much call for an elite bodyguard with such an injury. So, Josh has to support his family working as a mall cop in New Jersey. For a man like Josh, this is purgatory on earth, but miracles can occur even in Paramus. A lucky run-in with an old comrade promises to get Josh back in the field for one last job. The UN is sending a peace mission into the Sierra Madre mountains in Mexico, an area so dangerous it's known as Espinazo del Diablo—the Devil's Spine. Only a fool would think they could broker peace between the homicidal drug cartels in the region, and only a madman would sign on to keep those fools alive.

Sams Teach Yourself the Twitter API in 24 Hours -

Christopher Peri 2011-06-20

In just 24 sessions of one hour or less, you'll learn how to build great new social applications with the latest versions of the Twitter API. Using this book's straightforward, step-by-step approach, you'll discover all you can do with the Twitter API, and master everything from the absolute basics to the newest innovations. One step at a time, you'll learn how to build Twitter clients and extend them with more power... use advanced features like streaming and geotagging... even build mobile Twitter apps for iPhone and Android! Each lesson builds on what you've already learned resulting in a fully functional Twitter application, giving you a strong real-world foundation for success, even if you're completely new to Twitter development! Step-by-step instructions carefully walk you through the most common Twitter API programming tasks. Quizzes and Exercises at the end of each chapter help you test your knowledge. By the Way notes present interesting information related to the discussion. Did You Know? tips offer advice or show you easier ways to perform tasks. Watch Out! cautions alert you to possible problems and give you advice

on how to avoid them. Learn how to... Make the right upfront decisions in planning your application Integrate Twitter support into existing sites, services, and applications Set up your Twitter development environment Make Twitter API calls, handle responses, and develop readers Construct dynamic frameworks to efficiently generate and manage multiple API calls Build basic clients and extend them to handle more powerful tasks Securely access server resources with OAuth Use Direct Messages, Lists, Search, and other Twitter API features Enable users to control their accounts, establish favorites and friendships, send notifications, and block individuals Build location-based, geotagged applications with Twitter's GEO API Give users up-to-the-minute information about the hottest Twitter topics Get started with Twitter development for iPhone and Android Understand the future of Twitter API development

Apache Spark in 24 Hours, Sams Teach Yourself - Jeffrey Aven 2016-08-31

Apache Spark is a fast, scalable, and flexible open source distributed processing engine for big data systems and is one of the most active open source big data projects to date. In just 24 lessons of one hour or less, Sams Teach Yourself Apache Spark in 24 Hours helps you build practical Big Data solutions that leverage Spark's amazing speed, scalability, simplicity, and versatility. This book's straightforward, step-by-step approach shows you how to deploy, program, optimize, manage, integrate, and extend Spark—now, and for years to come. You'll discover how to create powerful solutions encompassing cloud computing, real-time stream processing, machine learning, and more. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success. Whether you are a data analyst, data engineer, data scientist, or data steward, learning Spark will help you to advance your career or embark on a new career in the booming area of Big Data. Learn how to • Discover what Apache Spark does and how it fits into the Big Data landscape • Deploy and run Spark locally or in the cloud • Interact with Spark from the shell • Make the most of the Spark Cluster Architecture • Develop Spark applications with Scala and functional Python • Program with the Spark API, including transformations and actions • Apply practical data engineering/analysis approaches designed for Spark • Use Resilient Distributed Datasets (RDDs) for caching, persistence, and output • Optimize Spark solution performance • Use Spark with SQL (via Spark SQL) and with NoSQL (via Cassandra) • Leverage cutting-edge functional programming techniques • Extend Spark with streaming, R, and Sparkling Water • Start building Spark-based machine learning and graph-processing applications • Explore advanced messaging technologies, including Kafka • Preview and prepare for Spark's next generation of innovations Instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished, you'll be comfortable using Apache Spark to solve a wide spectrum of Big Data problems.

Practical Apache Spark - Subhashini Chellappan 2018-12-12 Work with Apache Spark using Scala to deploy and set up single-node, multi-node, and high-availability clusters. This book discusses various components of Spark such as Spark Core, DataFrames, Datasets and SQL, Spark Streaming, Spark MLlib, and R on Spark with the help of practical code snippets for each topic.

Practical Apache Spark also covers the integration of Apache Spark with Kafka with examples. You'll follow a learn-to-do-by-yourself approach to learning – learn the concepts, practice the code snippets in Scala, and complete the assignments given to get an overall exposure. On completion, you'll have knowledge of the functional programming aspects of Scala, and hands-on expertise in various Spark components. You'll also become familiar with machine learning algorithms with real-time usage. What You Will Learn Discover the functional programming features of Scala Understand the complete architecture of Spark and its components Integrate Apache Spark with Hive and Kafka Use Spark SQL, DataFrames, and Datasets to process data using traditional SQL queries Work with different machine learning concepts and libraries using Spark's MLib packages Who This Book Is For Developers and professionals who deal with batch and stream data processing.

Programming Hive - Edward Capriolo 2012-09-26

Describes the features and functions of Apache Hive, the data infrastructure for Hadoop.

Apache Spark in 24 Hours - Andrea Avena 2017

Advanced Analytics with Spark - Sandy Ryza 2015-04-02

In this practical book, four Cloudera data scientists present a set of self-contained patterns for performing large-scale data analysis with Spark. The authors bring Spark, statistical methods, and real-world data sets together to teach you how to approach analytics problems by example. You'll start with an introduction to Spark and its ecosystem, and then dive into patterns that apply common techniques—classification, collaborative filtering, and anomaly detection among others—to fields such as genomics, security, and finance. If you have an entry-level understanding of machine learning and statistics, and you program in Java, Python, or Scala, you'll find these patterns useful for working on your own data applications. Patterns include: Recommending music and the Audioscrobbler data set Predicting forest cover with decision trees Anomaly detection in network traffic with K-means clustering Understanding Wikipedia with Latent Semantic Analysis Analyzing co-occurrence networks with GraphX Geospatial and temporal data analysis on the New York City Taxi Trips data Estimating financial risk through Monte Carlo simulation Analyzing genomics data and the BDG project Analyzing neuroimaging data with PySpark and Thunder

Sams Teach Yourself Java 2 in 24 Hours - Rogers Cadenhead 1999

Here is the quickest way for readers to learn everything they really need to know about Java 2. "Teach Yourself Java 2 in 24 Hours" is a step-by-step tutorial broken up into 24 short, easy one-hour chapters.

Learning Spark - Holden Karau 2015-01-28

Data in all domains is getting bigger. How can you work with it efficiently? Recently updated for Spark 1.3, this book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java, and Scala. This edition includes new information on Spark SQL, Spark Streaming, setup, and Maven coordinates. Written by the developers of Spark, this book will have data scientists and engineers up and running in no time. You'll learn how to express parallel jobs with just a few lines of code, and cover applications from simple batch jobs to stream processing and machine learning. Quickly dive into Spark capabilities such as distributed datasets, in-memory caching, and the interactive shell Leverage Spark's powerful built-in libraries, including Spark SQL, Spark Streaming, and MLib Use one programming paradigm instead of mixing and matching tools like Hive, Hadoop, Mahout, and Storm Learn how to deploy interactive, batch, and streaming applications Connect to data sources including HDFS, Hive, JSON, and S3 Master advanced topics like data partitioning and shared variables

Data Analytics with Hadoop - Benjamin Bengfort 2016-06

Ready to use statistical and machine-learning techniques across large data sets? This practical guide shows you why the Hadoop ecosystem is perfect for the job. Instead of deployment, operations, or software development usually associated with distributed computing, you'll focus on particular analyses you can build, the data warehousing techniques that Hadoop provides, and

higher order data workflows this framework can produce. Data scientists and analysts will learn how to perform a wide range of techniques, from writing MapReduce and Spark applications with Python to using advanced modeling and data management with Spark MLib, Hive, and HBase. You'll also learn about the analytical processes and data systems available to build and empower data products that can handle—and actually require—huge amounts of data. Understand core concepts behind Hadoop and cluster computing Use design patterns and parallel analytical algorithms to create distributed data analysis jobs Learn about data management, mining, and warehousing in a distributed context using Apache Hive and HBase Use Sqoop and Apache Flume to ingest data from relational databases Program complex Hadoop and Spark applications with Apache Pig and Spark DataFrames Perform machine learning techniques such as classification, clustering, and collaborative filtering with Spark's MLib

Hadoop: The Definitive Guide - Tom White 2012-05-10

Ready to unlock the power of your data? With this comprehensive guide, you'll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters. You'll find illuminating case studies that demonstrate how Hadoop is used to solve specific problems. This third edition covers recent changes to Hadoop, including material on the new MapReduce API, as well as MapReduce 2 and its more flexible execution model (YARN). Store large datasets with the Hadoop Distributed File System (HDFS) Run distributed computations with MapReduce Use Hadoop's data and I/O building blocks for compression, data integrity, serialization (including Avro), and persistence Discover common pitfalls and advanced features for writing real-world MapReduce programs Design, build, and administer a dedicated Hadoop cluster—or run Hadoop in the cloud Load data from relational databases into HDFS, using Sqoop Perform large-scale data processing with the Pig query language Analyze datasets with Hive, Hadoop's data warehousing system Take advantage of HBase for structured and semi-structured data, and ZooKeeper for building distributed systems

PHP and MySQL Web Development - Luke Welling 2008-10-01

PHP and MySQL Web Development, Fourth Edition The definitive guide to building database-driven Web applications with PHP and MySQL and MySQL are popular open-source technologies that are ideal for quickly developing database-driven Web applications. PHP is a powerful scripting language designed to enable developers to create highly featured Web applications quickly, and MySQL is a fast, reliable database that integrates well with PHP and is suited for dynamic Internet-based applications. PHP and MySQL Web Development shows how to use these tools together to produce effective, interactive Web applications. It clearly describes the basics of the PHP language, explains how to set up and work with a MySQL database, and then shows how to use PHP to interact with the database and the server. The fourth edition of PHP and MySQL Web Development has been thoroughly updated, revised, and expanded to cover developments in PHP 5 through version 5.3, such as namespaces and closures, as well as features introduced in MySQL 5.1. This is the eBook version of the title. To gain access to the contents on the CD bundled with the printed book, please register your product at informit.com/register

Big Data Analytics with Microsoft HDInsight in 24 Hours, Sams Teach Yourself - Manpreet Singh 2015-11-12

Sams Teach Yourself Big Data Analytics with Microsoft HDInsight in 24 Hours In just 24 lessons of one hour or less, Sams Teach Yourself Big Data Analytics with Microsoft HDInsight in 24 Hours helps you leverage Hadoop's power on a flexible, scalable cloud platform using Microsoft's newest business intelligence, visualization, and productivity tools. This book's straightforward, step-by-step approach shows you how to provision, configure, monitor, and troubleshoot HDInsight and use Hadoop cloud services to solve real analytics problems. You'll gain more of Hadoop's benefits, with less complexity—even if you're completely new to Big Data analytics. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success. Practical, hands-on examples show you how to apply what you learn Quizzes and exercises help you test your knowledge and stretch your skills Notes and tips point out

shortcuts and solutions Learn how to... · Master core Big Data and NoSQL concepts, value propositions, and use cases · Work with key Hadoop features, such as HDFS2 and YARN · Quickly install, configure, and monitor Hadoop (HDInsight) clusters in the cloud · Automate provisioning, customize clusters, install additional Hadoop projects, and administer clusters · Integrate, analyze, and report with Microsoft BI and Power BI · Automate workflows for data transformation, integration, and other tasks · Use Apache HBase on HDInsight · Use Sqoop or SSIS to move data to or from HDInsight · Perform R-based statistical computing on HDInsight datasets · Accelerate analytics with Apache Spark · Run real-time analytics on high-velocity data streams · Write MapReduce, Hive, and Pig programs Register your book at informit.com/register for convenient access to downloads, updates, and corrections as they become available.

Hadoop in 24 Hours, Sams Teach Yourself - Jeffrey Aven 2017-04-07

Apache Hadoop is the technology at the heart of the Big Data revolution, and Hadoop skills are in enormous demand. Now, in just 24 lessons of one hour or less, you can learn all the skills and techniques you'll need to deploy each key component of a Hadoop platform in your local environment or in the cloud, building a fully functional Hadoop cluster and using it with real programs and datasets. Each short, easy lesson builds on all that's come before, helping you master all of Hadoop's essentials, and extend it to meet your unique challenges. Apache Hadoop in 24 Hours, Sams Teach Yourself covers all this, and much more: Understanding Hadoop and the Hadoop Distributed File System (HDFS) Importing data into Hadoop, and process it there Mastering basic MapReduce Java programming, and using advanced MapReduce API concepts Making the most of Apache Pig and Apache Hive Implementing and administering YARN Taking advantage of the full Hadoop ecosystem Managing Hadoop clusters with Apache Ambari Working with the Hadoop User Environment (HUE) Scaling, securing, and troubleshooting Hadoop environments Integrating Hadoop into the enterprise Deploying Hadoop in the cloud Getting started with Apache Spark Step-by-step instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished, you'll be comfortable using Apache Hadoop to solve a wide spectrum of Big Data problems.

Unreal Engine 4 Game Development in 24 Hours, Sams Teach Yourself - Aram Cookson 2016

"In just 24 lessons of one hour or less, learn how to start using Unreal Engine 4 to build amazing games for Windows, Mac, PS4, Xbox One, iOS, Android, the web, Linux -- all of them! This book's straightforward, step-by-step approach shows you how to work with Unreal Engine 4's interface, its workflows, and its most powerful editors and tools. In just hours you'll be creating effects, scripting warfare, implementing physics--even developing for mobile devices and HUDs. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success." --

SQL in 10 Minutes, Sams Teach Yourself - Ben Forta 2012-10-25
Sams Teach Yourself SQL in 10 Minutes, Fourth Edition New full-color code examples help you see how SQL statements are structured Whether you're an application developer, database administrator, web application designer, mobile app developer, or Microsoft Office users, a good working knowledge of SQL is an important part of interacting with databases. And Sams Teach Yourself SQL in 10 Minutes offers the straightforward, practical answers you need to help you do your job. Expert trainer and popular author Ben Forta teaches you just the parts of SQL you need to know--starting with simple data retrieval and quickly going on to more complex topics including the use of joins, subqueries, stored procedures, cursors, triggers, and table constraints. You'll learn methodically, systematically, and simply--in 22 short, quick lessons that will each take only 10 minutes or less to complete. With the Fourth Edition of this worldwide bestseller, the book has been thoroughly updated, expanded, and improved. Lessons now cover the latest versions of IBM DB2, Microsoft Access, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, SQLite, MariaDB, and Apache Open Office Base. And new full-color SQL code listings help

the beginner clearly see the elements and structure of the language. 10 minutes is all you need to learn how to... Use the major SQL statements Construct complex SQL statements using multiple clauses and operators Retrieve, sort, and format database contents Pinpoint the data you need using a variety of filtering techniques Use aggregate functions to summarize data Join two or more related tables Insert, update, and delete data Create and alter database tables Work with views, stored procedures, and more Table of Contents 1 Understanding SQL 2 Retrieving Data 3 Sorting Retrieved Data 4 Filtering Data 5 Advanced Data Filtering 6 Using Wildcard Filtering 7 Creating Calculated Fields 8 Using Data Manipulation Functions 9 Summarizing Data 10 Grouping Data 11 Working with Subqueries 12 Joining Tables 13 Creating Advanced Joins 14 Combining Queries 15 Inserting Data 16 Updating and Deleting Data 17 Creating and Manipulating Tables 18 Using Views 19 Working with Stored Procedures 20 Managing Transaction Processing 21 Using Cursors 22 Understanding Advanced SQL Features Appendix A: Sample Table Scripts Appendix B: Working in Popular Applications Appendix C : SQL Statement Syntax Appendix D: Using SQL Datatypes Appendix E: SQL Reserved Words

The Art of R Programming - Norman Matloff 2011-10-11

R is the world's most popular language for developing statistical software: Archaeologists use it to track the spread of ancient civilizations, drug companies use it to discover which medications are safe and effective, and actuaries use it to assess financial risks and keep economies running smoothly. The Art of R Programming takes you on a guided tour of software development with R, from basic types and data structures to advanced topics like closures, recursion, and anonymous functions. No statistical knowledge is required, and your programming skills can range from hobbyist to pro. Along the way, you'll learn about functional and object-oriented programming, running mathematical simulations, and rearranging complex data into simpler, more useful formats. You'll also learn to: -Create artful graphs to visualize complex data sets and functions -Write more efficient code using parallel R and vectorization -Interface R with C/C++ and Python for increased speed or functionality -Find new R packages for text analysis, image manipulation, and more -Squash annoying bugs with advanced debugging techniques Whether you're designing aircraft, forecasting the weather, or you just need to tame your data, The Art of R Programming is your guide to harnessing the power of statistical computing.

Learning Spark - Jules S. Damji 2020-07-16

Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walkthroughs, code snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

High Performance Spark - Holden Karau 2017-05-25

Apache Spark is amazing when everything clicks. But if you haven't seen the performance improvements you expected, or still don't feel confident enough to use Spark in production, this practical book is for you. Authors Holden Karau and Rachel Warren demonstrate performance optimizations to help your Spark queries run faster and handle larger data sizes, while using fewer resources. Ideal for software engineers, data engineers, developers, and system administrators working with large-scale data applications, this book describes techniques that can reduce data infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you'll also learn how to make it sing. With this book, you'll explore: How Spark SQL's new interfaces improve performance over SQL's RDD

data structure The choice between data joins in Core Spark and Spark SQL Techniques for getting the most out of standard RDD transformations How to work around performance issues in Spark's key/value pair paradigm Writing high-performance Spark code without Scala or the JVM How to test for functionality and performance when applying suggested improvements Using Spark MLib and Spark ML machine learning libraries Spark's Streaming components and external community packages

Business Analytics - Thomas W. Jackson 2018-09-21

This innovative new textbook, co-authored by an established academic and a leading practitioner, is the first to bring together issues of cloud computing, business intelligence and big data analytics in order to explore how organisations use cloud technology to analyse data and make decisions. In addition to offering an up-to-date exploration of key issues relating to data privacy and ethics, information governance, and the future of analytics, the text describes the options available in deploying analytic solutions to the cloud and draws on real-world, international examples from companies such as Rolls Royce, Lego, Volkswagen and Samsung. Combining academic and practitioner perspectives that are crucial to the understanding of this growing field, Business Analytics acts an ideal core text for undergraduate, postgraduate and MBA modules on Big Data, Business and Data Analytics, and Business Intelligence, as well as functioning as a supplementary text for modules in Marketing Analytics. The book is also an invaluable resource for practitioners and will quickly enable the next generation of 'Information Builders' within organisations to understand innovative cloud based-analytic solutions.

SQL in 10 Minutes a Day, Sams Teach Yourself - Ben Forta 2019-10-29

Sams Teach Yourself SQL in 10 Minutes offers straightforward, practical answers when you need fast results. By working through the book's 22 lessons of 10 minutes or less, you'll learn what you need to know to take advantage of the SQL language. Lessons cover IBM DB2, Microsoft SQL Server and SQL Server Express, MariaDB, MySQL, Oracle and Oracle express, PostgreSQL, and SQLite. Full-color code examples help you understand how SQL statements are structured Tips point out shortcuts and solutions Cautions help you avoid common pitfalls Notes explain additional concepts, and provide additional information 10 minutes is all you need to learn how to... Use the major SQL statements Construct complex SQL statements using multiple clauses and operators Retrieve, sort, and format database contents Pinpoint the data you need using a variety of filtering techniques Use aggregate functions to summarize data Join two or more related tables Insert, update, and delete data Create and alter database tables Work with views, stored procedures, and more

Learning Spark SQL - Aurobindo Sarkar 2017-09-07

Design, implement, and deliver successful streaming applications, machine learning pipelines and graph applications using Spark SQL API About This Book Learn about the design and implementation of streaming applications, machine learning pipelines, deep learning, and large-scale graph processing applications using Spark SQL APIs and Scala. Learn data exploration, data munging, and how to process structured and semi-structured data using real-world datasets and gain hands-on exposure to the issues and challenges of working with noisy and "dirty" real-world data. Understand design considerations for scalability and performance in web-scale Spark application architectures. Who This Book Is For If you are a developer, engineer, or an architect and want to learn how to use Apache Spark in a web-scale project, then this is the book for you. It is assumed that you have prior knowledge of SQL querying. A basic programming knowledge with Scala, Java, R, or Python is all you need to get started with this book. What You Will Learn Familiarize yourself with Spark SQL programming, including working with DataFrame/Dataset API and SQL Perform a series of hands-on exercises with different types of data sources, including CSV, JSON, Avro, MySQL, and MongoDB Perform data quality checks, data visualization, and basic statistical analysis tasks Perform data munging tasks on publically available datasets Learn how to use Spark SQL and Apache Kafka to build streaming applications Learn key performance-tuning tips and tricks in Spark SQL applications Learn key architectural components and patterns in large-scale

Spark SQL applications In Detail In the past year, Apache Spark has been increasingly adopted for the development of distributed applications. Spark SQL APIs provide an optimized interface that helps developers build such applications quickly and easily. However, designing web-scale production applications using Spark SQL APIs can be a complex task. Hence, understanding the design and implementation best practices before you start your project will help you avoid these problems. This book gives an insight into the engineering practices used to design and build real-world, Spark-based applications. The book's hands-on examples will give you the required confidence to work on any future projects you encounter in Spark SQL. It starts by familiarizing you with data exploration and data munging tasks using Spark SQL and Scala. Extensive code examples will help you understand the methods used to implement typical use-cases for various types of applications. You will get a walkthrough of the key concepts and terms that are common to streaming, machine learning, and graph applications. You will also learn key performance-tuning details including Cost Based Optimization (Spark 2.2) in Spark SQL applications. Finally, you will move on to learning how such systems are architected and deployed for a successful delivery of your project. Style and approach This book is a hands-on guide to designing, building, and deploying Spark SQL-centric production applications at scale.

NoSQL Distilled - Pramod J. Sadalage 2013

'NoSQL Distilled' is designed to provide you with enough background on how NoSQL databases work, so that you can choose the right data store without having to trawl the whole web to do it. It won't answer your questions definitively, but it should narrow down the range of options you have to consider.

Beginning Apache Spark 2 - Hien Luu 2018-08-16

Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud technologies. Beginning Apache Spark 2 gives you an introduction to Apache Spark and shows you how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applications with Spark Structured Streaming. Furthermore, you'll learn the fundamentals of Spark ML for machine learning and much more. After you read this book, you will have the fundamentals to become proficient in using Apache Spark and know when and how to apply it to your big data applications. What You Will Learn Understand Spark unified data processing platform How to run Spark in Spark Shell or Databricks Use and manipulate RDDs Deal with structured data using Spark SQL through its operations and advanced functions Build real-time applications using Spark Structured Streaming Develop intelligent applications with the Spark Machine Learning library Who This Book Is For Programmers and developers active in big data, Hadoop, and Java but who are new to the Apache Spark platform. **PySpark Recipes** - Raju Kumar Mishra 2017-12-09

Quickly find solutions to common programming problems encountered while processing big data. Content is presented in the popular problem-solution format. Look up the programming problem that you want to solve. Read the solution. Apply the solution directly in your own code. Problem solved! PySpark Recipes covers Hadoop and its shortcomings. The architecture of Spark, PySpark, and RDD are presented. You will learn to apply RDD to solve day-to-day big data problems. Python and NumPy are included and make it easy for new learners of PySpark to understand and adopt the model. What You Will Learn Understand the advanced features of PySpark2 and SparkSQL Optimize your code Program SparkSQL with Python Use Spark Streaming and Spark MLib with Python Perform graph analysis with GraphFrames Who This Book Is For Data analysts, Python programmers, big data enthusiasts

Mastering Regular Expressions - Jeffrey E. F. Friedl 2002

Introduces regular expressions and how they are used, discussing topics including metacharacters, nomenclature, matching and modifying text, expression processing, benchmarking, optimizations, and loops.

Learn Azure in a Month of Lunches, Second Edition - Iain Foulds 2020-10-06

Learn Azure in a Month of Lunches, Second Edition, is a tutorial on

writing, deploying, and running applications in Azure. In it, you'll work through 21 short lessons that give you real-world experience. Each lesson includes a hands-on lab so you can try out and lock in your new skills. Summary You can be incredibly productive with Azure without mastering every feature, function, and service. Learn Azure in a Month of Lunches, Second Edition gets you up and running quickly, teaching you the most important concepts and tasks in 21 practical bite-sized lessons. As you explore the examples, exercises, and labs, you'll pick up valuable skills immediately and take your first steps to Azure mastery! This fully revised new edition covers core changes to the Azure UI, new Azure features, Azure containers, and the upgraded Azure Kubernetes Service. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Microsoft Azure is vast and powerful, offering virtual servers, application templates, and prebuilt services for everything from data storage to AI. To navigate it all, you need a trustworthy guide. In this book, Microsoft engineer and Azure trainer Iain Foulds focuses on core skills for creating cloud-based applications. About the book Learn Azure in a Month of Lunches, Second Edition, is a tutorial on writing, deploying, and running applications in Azure. In it, you'll work through 21 short lessons that give you real-world experience. Each lesson includes a hands-on lab so you can try out and lock in your new skills. What's inside Understanding Azure beyond point-and-click Securing applications and data Automating your environment Azure services for machine learning, containers, and more About the reader This book is for readers who can write and deploy simple web or client/server applications. About the author Iain Foulds is an engineer and senior content developer with Microsoft. Table of Contents PART 1 - AZURE CORE SERVICES 1 Before you begin 2 Creating a virtual machine 3 Azure Web Apps 4 Introduction to Azure Storage 5 Azure Networking basics PART 2 - HIGH AVAILABILITY AND SCALE 6 Azure Resource Manager 7 High availability and redundancy 8 Load-balancing applications 9 Applications that scale 10 Global databases with Cosmos DB 11 Managing network traffic and routing 12 Monitoring and troubleshooting PART 3 - SECURE BY DEFAULT 13 Backup, recovery, and replication 14 Data encryption 15 Securing information with Azure Key Vault 16 Azure Security Center and updates PART 4 - THE COOL STUFF 17 Machine learning and artificial intelligence 18 Azure Automation 19 Azure containers 20 Azure and the Internet of Things 21 Serverless computing **AngularJS, JavaScript, and JQuery All in One in 24 Hours** - Brad Dayley 2015

As you complete the lessons in this book, you will gain a practical understanding of how to provide rich user interactions in your web pages. You will learn how to add dynamic code that allows web pages to instantly react to mouse clicks and finger swipes, interact with back-end services to store and retrieve data from the web server, and create robust Internet applications.

Deep Learning - Josh Patterson 2017-07-28

Although interest in machine learning has reached a high point, lofty expectations often scuttle projects before they get very far. How can machine learning—especially deep neural networks—make a real difference in your organization? This hands-on guide not only provides the most practical information available on the subject, but also helps you get started building efficient deep learning networks. Authors Adam Gibson and Josh Patterson provide theory on deep learning before introducing their open-source DeepLearning4j (DL4J) library for developing production-class workflows. Through real-world examples, you'll learn methods and strategies for training deep network architectures and running deep learning workflows on Spark and Hadoop with DL4J. Dive into machine learning concepts in general, as well as deep learning in particular Understand how deep networks evolved from neural network fundamentals Explore the major deep network architectures, including Convolutional and Recurrent Learn how to map specific deep networks to the right problem Walk through the fundamentals of tuning general neural networks and specific deep network architectures Use vectorization techniques for different data types with DataVec, DL4J's workflow tool Learn how to use DL4J natively on Spark and Hadoop

Spark Cookbook - Rishi Yadav 2015-07-27

By introducing in-memory persistent storage, Apache Spark eliminates the need to store intermediate data in filesystems, thereby increasing processing speed by up to 100 times. This book will focus on how to analyze large and complex sets of data. Starting with installing and configuring Apache Spark with various cluster managers, you will cover setting up development environments. You will then cover various recipes to perform interactive queries using Spark SQL and real-time streaming with various sources such as Twitter Stream and Apache Kafka. You will then focus on machine learning, including supervised learning, unsupervised learning, and recommendation engine algorithms. After mastering graph processing using GraphX, you will cover various recipes for cluster optimization and troubleshooting.

Sams Teach Yourself Java in 21 Days (Covers Java 11/12) - Rogers Cadenhead 2019-12-09

In just 21 days, you can acquire the knowledge and skills necessary to develop applications on your computer, web servers, and mobile devices. With this complete tutorial you'll quickly master the basics and then move on to more advanced features and concepts. Completely updated for Java 11 and 12, this book teaches you about the Java language and how to use it to create applications for any computing environment. By the time you have finished the book, you'll have well-rounded knowledge of Java and the Java class libraries. No previous programming experience required. By following the 21 carefully organized lessons in this book, anyone can learn the basics of Java programming. Learn at your own pace. You can work through each chapter sequentially to make sure you thoroughly understand all the concepts and methodologies, or you can focus on specific lessons to learn the techniques that interest you most. Test your knowledge. Each chapter ends with a Workshop section filled with questions, answers, and exercises for further study. There are even certification practice questions. Completely revised, updated, and expanded to cover the latest features of Java 11 and 12 Learn to develop Java applications using NetBeans—an excellent programming platform Easy-to-understand, practical examples clearly illustrate the fundamentals of Java programming Discover how to quickly develop programs with a graphical user interface Find out about JDBC programming with the Derby database Learn how to use Inner Classes and Lambda Expressions Learn rapid application development with Apache NetBeans Create a game using Java

[Spark: The Definitive Guide](#) - Bill Chambers 2018-02-08

Learn how to use, deploy, and maintain Apache Spark with this comprehensive guide, written by the creators of the open-source cluster-computing framework. With an emphasis on improvements and new features in Spark 2.0, authors Bill Chambers and Matei Zaharia break down Spark topics into distinct sections, each with unique goals. You'll explore the basic operations and common functions of Spark's structured APIs, as well as Structured Streaming, a new high-level API for building end-to-end streaming applications. Developers and system administrators will learn the fundamentals of monitoring, tuning, and debugging Spark, and explore machine learning techniques and scenarios for employing MLlib, Spark's scalable machine-learning library. Get a gentle overview of big data and Spark Learn about DataFrames, SQL, and Datasets—Spark's core APIs—through worked examples Dive into Spark's low-level APIs, RDDs, and execution of SQL and DataFrames Understand how Spark runs on a cluster Debug, monitor, and tune Spark clusters and applications Learn the power of Structured Streaming, Spark's stream-processing engine Learn how you can apply MLlib to a variety of problems, including classification or recommendation

What Is Marketing? - Harvard Business Review 2006-10-01

Sometimes you need more than a one-sentence answer. While the term marketing generally refers to what a company does to create value for customers, practicing marketers know they have a major role in setting their company's strategic direction. Successful marketing requires a deep knowledge of customers, competitors, and collaborators—and great skill in serving customers profitably. The book provides the foundation for developing those skills and insights. It's organized according to the design of the first-year marketing course in Harvard Business School's MBA program. Each chapter was written by HBS faculty and used by MBA students to analyze marketing opportunities and develop and

execute successful marketing strategies. Areas covered include: Consumer behavior Business-to-business markets The four P's- product, placement, promotion and price Market segmentation, target market selection, and positioning Unique value propositions The design of new products and services Product line extensions and repositioning of exciting businesses Brand valuation and brand equity Fulfillment and after-sale service Direct, retail, and wholesale distribution channels and networks Marketing communications and promotions Advertising, public relations, and choice of media Pricing for profitability Personal selling and sales management Customer relationship management and customer privacy Customer acquisition, retention, and dismissal Basic math for making marketing decisions Timeless yet timely, this book provides valuable background information for understanding and interpreting business and competition from a marketing point of view. That makes it useful in both formal and informal educational settings, including on-the-job training. Simply put, it's required reading for marketing students and a must-have recourse for marketing professionals.

Kafka: The Definitive Guide - Neha Narkhede 2017-08-31
Every enterprise application creates data, whether it's log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you're an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds. Engineers from Confluent and LinkedIn who are responsible for developing Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples, you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer. Understand publish-subscribe messaging and how it fits in the big data ecosystem. Explore Kafka producers and consumers for writing and reading messages Understand Kafka patterns and use-case requirements to ensure reliable data delivery Get best practices for building data pipelines and applications with Kafka Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks Learn the most critical metrics among Kafka's operational measurements Explore how Kafka's stream delivery capabilities make it a perfect source for stream processing systems

Sams Teach Yourself SQL in 24 Hours - Ryan Stephens 2008-05-30

In just 24 lessons of one hour or less, you will learn professional techniques to design and build efficient databases and query them to extract useful information. Using a straightforward, step-by-step approach, each lesson builds on the previous one, allowing you to learn the essentials of ANSI SQL from the ground up. Example code demonstrates the authors' professional techniques, while exercises written for MySQL offer the reader hands-on learning with an open-source database. Included are advanced techniques for using views, managing transactions, database administration, and extending SQL. Step-by-step instructions carefully walk you through the most common SQL tasks. Q&As, Quizzes, and Exercises at the end of each chapter help you test your knowledge. Notes and Tips point out shortcuts and solutions. New terms are clearly defined and explained. Learn how to... Use SQL-2003, the latest standard for the Structured Query Language Design and deploy efficient, secure databases Build advanced queries for information retrieval Sort, group, and summarize information for best presentation Tune databases and queries for maximum performance Understand database administration and security techniques For more than ten years the authors have studied, applied, and documented the SQL standard and its application to critical database systems. Ryan Stephens and Ron Plew are entrepreneurs, speakers, and cofounders of Perpetual Technologies, Inc. (PTI), a fast-growing IT management and consulting firm which specializes in database technologies. They taught database courses for Indiana University-Purdue University in Indianapolis for five years and have authored more than a dozen books on Oracle, SQL, database design, and the high availability of critical systems. Arie D. Jones is Senior SQL Server database administrator and analyst for PTI. He is a regular speaker

at technical events and has authored several books and articles. Category: Database Covers: ANSI SQL User Level: Beginning-Intermediate Register your book at informit.com/title/9780672330186 for convenient access to updates and corrections as they become available.

Data Algorithms - Mahmoud Parsian 2015-07-13
If you are ready to dive into the MapReduce framework for processing large datasets, this practical book takes you step by step through the algorithms and tools you need to build distributed MapReduce applications with Apache Hadoop or Apache Spark. Each chapter provides a recipe for solving a massive computational problem, such as building a recommendation system. You'll learn how to implement the appropriate MapReduce solution with code that you can use in your projects. Dr. Mahmoud Parsian covers basic design patterns, optimization techniques, and data mining and machine learning solutions for problems in bioinformatics, genomics, statistics, and social network analysis. This book also includes an overview of MapReduce, Hadoop, and Spark. Topics include: Market basket analysis for a large set of transactions Data mining algorithms (K-means, KNN, and Naive Bayes) Using huge genomic data to sequence DNA and RNA Naive Bayes theorem and Markov chains for data and market prediction Recommendation algorithms and pairwise document similarity Linear regression, Cox regression, and Pearson correlation Allelic frequency and mining DNA Social network analysis (recommendation systems, counting triangles, sentiment analysis)

Sams Teach Yourself SQL in 10 Minutes - Ben Forta 2004
With this updated text, readers can learn the fundamentals of SQL quickly through the use of numerous examples depicting all the major components of SQL.

Data Analytics with Spark Using Python - Jeffrey Aven 2018-06-18
Solve Data Analytics Problems with Spark, PySpark, and Related Open Source Tools Spark is at the heart of today's Big Data revolution, helping data professionals supercharge efficiency and performance in a wide range of data processing and analytics tasks. In this guide, Big Data expert Jeffrey Aven covers all you need to know to leverage Spark, together with its extensions, subprojects, and wider ecosystem. Aven combines a language-agnostic introduction to foundational Spark concepts with extensive programming examples utilizing the popular and intuitive PySpark development environment. This guide's focus on Python makes it widely accessible to large audiences of data professionals, analysts, and developers—even those with little Hadoop or Spark experience. Aven's broad coverage ranges from basic to advanced Spark programming, and Spark SQL to machine learning. You'll learn how to efficiently manage all forms of data with Spark: streaming, structured, semi-structured, and unstructured. Throughout, concise topic overviews quickly get you up to speed, and extensive hands-on exercises prepare you to solve real problems. Coverage includes: • Understand Spark's evolving role in the Big Data and Hadoop ecosystems • Create Spark clusters using various deployment modes • Control and optimize the operation of Spark clusters and applications • Master Spark Core RDD API programming techniques • Extend, accelerate, and optimize Spark routines with advanced API platform constructs, including shared variables, RDD storage, and partitioning • Efficiently integrate Spark with both SQL and nonrelational data stores • Perform stream processing and messaging with Spark Streaming and Apache Kafka • Implement predictive modeling with SparkR and Spark MLlib

Sams Teach Yourself Hadoop in 24 Hours - Jeffrey Aven 2017
Apache Hadoop is the technology at the heart of the Big Data revolution, and Hadoop skills are in enormous demand. Now, in just 24 lessons of one hour or less, students can learn all the skills and techniques they'll need to deploy each key component of a Hadoop platform in a local environment or in the cloud, building a fully functional Hadoop cluster and using it with real programs and datasets. Each short, easy lesson builds on all that's come before, helping students master all of Hadoop's essentials, and extend it to meet real-world challenges. Apache Hadoop in 24 Hours, Sams Teach Yourself covers all this, and much more: Understanding Hadoop and the Hadoop Distributed File System (HDFS) Importing data into Hadoop, and process it there Mastering basic MapReduce

Java programming, and using advanced MapReduce API concepts
Making the most of Apache Pig and Apache Hive Implementing
and administering YARN Taking advantage of the full Hadoop
ecosystem Managing Hadoop clusters with Apache Ambari
Working with the Hadoop User Environment (HUE) Scaling,
securing, and troubleshooting Hadoop environments Integrating
Hadoop into the enterprise Deploying Hadoop in the cloud Getting

started with Apache Spark Step-by-step instructions walk students
through common questions, issues, and tasks; Q-and-As, Quizzes,
and Exercises build and test your knowledge; Did You Know? tips
offer insider advice and shortcuts; and Watch Out! alerts help
avoid pitfalls. By the time they're finished, they'll be comfortable
using Apache Hadoop to solve a wide spectrum of Big Data
problems.