

Hadoop In 24 Hours Sams Teach Yourself

Hadoop in 24 Hours, Sams Teach Yourself Sams Publishing

Get a solid grounding in Apache Oozie, the workflow scheduler system for managing Hadoop jobs. With this hands-on guide, two experienced Hadoop practitioners walk you through the intricacies of this powerful and flexible platform, with numerous examples and real-world use cases. Once you set up your Oozie server, you'll dive into techniques for writing and coordinating workflows, and learn how to write complex data pipelines. Advanced topics show you how to handle shared libraries in Oozie, as well as how to implement and manage Oozie's security capabilities. Install and configure an Oozie server, and get an overview of basic concepts Journey through the world of writing and configuring workflows Learn how the Oozie coordinator schedules and executes workflows based on triggers Understand how Oozie manages data dependencies Use Oozie bundles to package several coordinator apps into a data pipeline Learn about security features and shared library management Implement custom extensions and write your own EL

functions and actions Debug workflows and manage Oozie's operational details

There is an easier way to build Hadoop applications. With this hands-on book, you'll learn how to use Cascading, the open source abstraction framework for Hadoop that lets you easily create and manage powerful enterprise-grade data processing applications—without having to learn the intricacies of MapReduce. Working with sample apps based on Java and other JVM languages, you'll quickly learn Cascading's streamlined approach to data processing, data filtering, and workflow optimization. This book demonstrates how this framework can help your business extract meaningful information from large amounts of distributed data. Start working on Cascading example projects right away

Model and analyze unstructured data in any format, from any source Build and test applications with familiar constructs and reusable components Work with the Scalding and Cascalog Domain-Specific Languages Easily deploy applications to Hadoop, regardless of cluster location or data size Build workflows that integrate several big data frameworks and processes Explore common use cases for Cascading, including features and tools that support them Examine a case study that uses a

dataset from the Open Data Initiative

This is the Rough Cut version of the printed book. With The world of data is changing rapidly. The growing demands of end users (Consumerization of IT) and availability of new types of data (Data explosion - 85% of this new data is coming from new data types e.g. sensors, RFIDs, WebLogs, high-definition video streaming, oil and gas exploration etc.) is causing a widening gap between our ability to store vast amounts of data and our ability to get meaningful insight and drive decision making based on this vast amount of data. This data explosion, combined with the fact that the cost of storage has practically gone to zero has landed us in a world where we need to have the ability to store all this data and get insight into it. This makes sense for companies to make better business decisions by enabling data scientists and other users to analyze huge volumes of transaction data as well as other data sources that may be left untapped by traditional business intelligence (BI) programs. On the analytics front there is a shift from traditional BI to predictive analytics as well - traditional BI helps customers to understand what has happened in past (rear view mirror) whereas predictive analysis allows customer to understand what

**would happen in future (forward-looking view). Predictive analysis has been effective in areas such as fraud detection, sales targeting, customer churn analysis, Ad Placement to increase revenue etc. This book is going to cover in detail about storing vast amount of data (big data) on hadoop on windows (in Windows Azure platform) and getting insight into it with familiar Microsoft BI tools. It addresses questions such as, "What is Big Data and how can Hadoop be used by an organization to tap into it? What are some of the important tools and technologies around the Hadoop ecosystem and Microsoft's partnership with Hortonworks?" From this book you will learn: Ease of installation, configuration and monitoring of Hadoop (HDInsight) cluster on cloud platform; Distributed storage and processing of unstructured data or big data; Programming to do big data analytics with MapReduce, Hive, PIG; Integration of Hadoop with Microsoft BI (MSBI) tools; Analyze and create visualization reports your with Microsoft Power BI. With Resilient Distributed Datasets, Spark SQL, Structured Streaming and Spark Machine Learning library
Practical Hive
Learning PySpark**

Hadoop in 24 Hours, Sams Teach Yourself Data-intensive Text Processing with MapReduce Data Science for Business

Build data-intensive applications locally and deploy at scale using the combined powers of Python and Spark 2.0 About This Book Learn why and how you can efficiently use Python to process data and build machine learning models in Apache Spark 2.0 Develop and deploy efficient, scalable real-time Spark solutions Take your understanding of using Spark with Python to the next level with this jump start guide Who This Book Is For If you are a Python developer who wants to learn about the Apache Spark 2.0 ecosystem, this book is for you. A firm understanding of Python is expected to get the best out of the book. Familiarity with Spark would be useful, but is not mandatory. What You Will Learn Learn about Apache Spark and the Spark 2.0 architecture Build and interact with Spark DataFrames using Spark SQL Learn how to solve graph and deep learning problems using GraphFrames and TensorFrames respectively Read, transform, and understand data and use it to train machine learning models Build machine learning models with MLlib and ML Learn how to submit your applications programmatically using spark-submit Deploy locally built applications to a cluster In Detail Apache Spark is an open source framework for efficient cluster computing with a strong interface for data parallelism and fault tolerance. This book will show you how to leverage the power of Python and put it to use in the Spark ecosystem. You will start by getting a firm understanding

Download Free Hadoop In 24 Hours Sams Teach Yourself

of the Spark 2.0 architecture and how to set up a Python environment for Spark. You will get familiar with the modules available in PySpark. You will learn how to abstract data with RDDs and DataFrames and understand the streaming capabilities of PySpark. Also, you will get a thorough overview of machine learning capabilities of PySpark using ML and MLlib, graph processing using GraphFrames, and polyglot persistence using Blaze. Finally, you will learn how to deploy your applications to the cloud using the spark-submit command. By the end of this book, you will have established a firm understanding of the Spark Python API and how it can be used to build data-intensive applications. **Style and approach** This book takes a very comprehensive, step-by-step approach so you understand how the Spark ecosystem can be used with Python to develop efficient, scalable solutions. Every chapter is standalone and written in a very easy-to-understand manner, with a focus on both the hows and the whys of each concept.

Third Edition: Thoroughly Updated and Expanded, with Extensive New Coverage! In just 24 sessions of one hour or less, you'll master the entire SAP project lifecycle, from planning through implementation and system administration through day-to-day operations. Using this book's straightforward, step-by-step approach, you'll gain a strong real-world foundation in both the technology and business essentials of today's SAP products and applications—from the ground up. Step-by-step instructions walk you through the most common questions, issues, and tasks you'll encounter with SAP.

Download Free Hadoop In 24 Hours Sams Teach Yourself

Case study-based exercises help you build and test your knowledge. By the Way notes present interesting pieces of information. Did You Know? tips offer advice or teach an easier way. Watch Out! cautions warn about potential problems. Learn how to... Understand SAP's newest products for enterprises and small-to-midsize businesses, and choose the right solutions for your company Discover how SAP integrates with Web services and service-oriented architecture Develop an efficient roadmap for deploying SAP in your environment Plan your SAP implementation from business, functional, technical, and project management perspectives Leverage NetWeaver 7.0 features to streamline development and integration, and reduce cost Walk through a step-by-step SAP technical installation Master basic SAP system administration and operations Perform essential tasks such as logon, session management, and printing Build SAP queries and reports Prepare for SAP upgrades and enhancements Develop your own personal career as an SAP professional Register your book at informit.com/title/9780137142842 for convenient access to updates and corrections as they become available.

Let Hadoop For Dummies help harness the power of your data and rein in the information overload Big data has become big business, and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed. Enter Hadoop and this easy-to-understand For Dummies guide. Hadoop For Dummies helps readers understand the value of big

Download Free Hadoop In 24 Hours Sams Teach Yourself

data, make a business case for using Hadoop, navigate the Hadoop ecosystem, and build and manage Hadoop applications and clusters. Explains the origins of Hadoop, its economic benefits, and its functionality and practical applications Helps you find your way around the Hadoop ecosystem, program MapReduce, utilize design patterns, and get your Hadoop cluster up and running quickly and easily Details how to use Hadoop applications for data mining, web analytics and personalization, large-scale text processing, data science, and problem-solving Shows you how to improve the value of your Hadoop cluster, maximize your investment in Hadoop, and avoid common pitfalls when building your Hadoop cluster From programmers challenged with building and maintaining affordable, scaleable data systems to administrators who must deal with huge volumes of information effectively and efficiently, this how-to has something to help you with Hadoop.

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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.

Big Data Processing Made Simple

Hadoop in Practice

Sams Teach Yourself SAP in 24 Hours

Unreal Engine 4 Game Development in 24 Hours, Sams Teach Yourself

NoSQL For Dummies

The Workflow Scheduler for Hadoop

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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. This volume is a printed version of a work that appears in the Synthesis Digital Library of Engineering and Computer Science. Synthesis Lectures provide concise, original presentations of important research and development topics, published quickly, in digital and print formats. For more information visit www.morganclaypool.com

Get ready to unlock the power of your data. With the fourth edition of 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. Using Hadoop 2 exclusively, author Tom White presents new chapters on YARN and several Hadoop-related projects such as Parquet, Flume, Crunch, and Spark. You'll learn about recent changes to Hadoop, and explore new case studies on Hadoop's role in healthcare systems and genomics data processing. Learn fundamental components such as MapReduce, HDFS, and YARN

Explore MapReduce in depth, including steps for developing applications with it Set up and maintain a Hadoop cluster running HDFS and MapReduce on YARN Learn two data formats: Avro for data serialization and Parquet for nested data Use data ingestion tools such as Flume (for streaming data) and Sqoop (for bulk data transfer) Understand how high-level data processing tools like Pig, Hive, Crunch, and Spark work with Hadoop Learn the

Download Free Hadoop In 24 Hours Sams Teach Yourself

HBase distributed database and the ZooKeeper distributed configuration service

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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.

Expert Apache Cassandra Administration

Storage and Analysis at Internet Scale

What You Need to Know about Data Mining and Data-Analytic Thinking

Sams Teach Yourself Hadoop in 24 Hours

Managing Spark, YARN, and MapReduce

Get up to speed on the nuances of NoSQL databases and what they mean for your organization This easy to read guide to NoSQL databases provides the type of no-nonsense overview and analysis that you need to learn, including what NoSQL is and which database is right for you. Featuring specific evaluation criteria for NoSQL databases, along with a look into the pros and cons of the most popular options, NoSQL For Dummies provides the fastest and easiest way to dive into the details of this incredible technology. You'll gain an understanding of how to use NoSQL databases for mission-critical enterprise architectures and projects, and real-world examples reinforce the primary points to create an action-oriented resource for IT pros. If you're planning a big data project or platform, you probably already know you need to select a NoSQL database to complete your architecture. But with options flooding the market and updates and add-ons coming at a rapid pace, determining what you

require now, and in the future, can be a tall task. This is where NoSQL For Dummies comes in! Learn the basic tenets of NoSQL databases and why they have come to the forefront as data has outpaced the capabilities of relational databases Discover major players among NoSQL databases, including Cassandra, MongoDB, MarkLogic, Neo4J, and others Get an in-depth look at the benefits and disadvantages of the wide variety of NoSQL database options Explore the needs of your organization as they relate to the capabilities of specific NoSQL databases Big data and Hadoop get all the attention, but when it comes down to it, NoSQL databases are the engines that power many big data analytics initiatives. With NoSQL For Dummies, you'll go beyond relational databases to ramp up your enterprise's data architecture in no time. In just 24 sessions of one hour or less, learn how to build powerful, easy-to-maintain websites with Drupal—fast! Using this book's straightforward, step-by-step approach, you'll master every skill you'll need, from organizing sites and using Drupal's design themes to setting up search, polls, forums, and security. Each lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success! Step-by-step instructions carefully walk you through the most common Drupal 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. Jesse Feiler provides consulting services to small businesses and non-profits through his company, North Country Consulting (northcountryconsulting.com). His recent books include FileMaker Pro 10 In Depth, The Bento Book, How to Do Everything with Web 2.0 Mashups, and iWork '09 For Dummies®. Register your book at informit.com/register for convenient access to all sample project source code, as well as updates and corrections as they become available. Learn how to... Make the most of Drupal, whether you're building business, non-profit, or personal sites Download, install, and configure open source Drupal 6 and 7, and make sure it's working properly Plan and organize your sites so they are easy to manage—and friendly to both users and search engines Customize sites with Drupal's extensive library of optional open source modules Create barebone sites and home pages automatically Add text, images, links, and other essential site features Master Drupal's powerful Content Construction Kit (CCK) Incorporate new types of content, ranging from articles to e-commerce products Secure Drupal sites by managing users, permissions, roles, and user profiles Categorize and tag content, and implement full-text search Promote user involvement with polls, comments, forums, RSS feed aggregation, blogs, and newsletters Automate and simplify

site management with Triggers, Actions, and Rules

For many organizations, Hadoop is the first step for dealing with massive amounts of data. The next step? Processing and analyzing datasets with the Apache Pig scripting platform. With Pig, you can batch-process data without having to create a full-fledged application, making it easy to experiment with new datasets. Updated with use cases and programming examples, this second edition is the ideal learning tool for new and experienced users alike. You'll find comprehensive coverage on key features such as the Pig Latin scripting language and the Grunt shell. When you need to analyze terabytes of data, this book shows you how to do it efficiently with Pig. Delve into Pig's data model, including scalar and complex data types Write Pig Latin scripts to sort, group, join, project, and filter your data Use Grunt to work with the Hadoop Distributed File System (HDFS) Build complex data processing pipelines with Pig's macros and modularity features Embed Pig Latin in Python for iterative processing and other advanced tasks Use Pig with Apache Tez to build high-performance batch and interactive data processing applications Create your own load and store functions to handle data formats and storage mechanisms

Dive into the world of SQL on Hadoop and get the most out of your Hive data warehouses.

This book is your go-to resource for using Hive: authors Scott Shaw, Ankur Gupta, David Kjerrumgaard, and Andreas Francois Vermeulen

Download Free Hadoop In 24 Hours Sams Teach Yourself

take you through learning HiveQL, the SQL-like language specific to Hive, to analyze, export, and massage the data stored across your Hadoop environment. From deploying Hive on your hardware or virtual machine and setting up its initial configuration to learning how Hive interacts with Hadoop, MapReduce, Tez and other big data technologies, Practical Hive gives you a detailed treatment of the software. In addition, this book discusses the value of open source software, Hive performance tuning, and how to leverage semi-structured and unstructured data. What You Will Learn Install and configure Hive for new and existing datasets Perform DDL operations Execute efficient DML operations Use tables, partitions, buckets, and user-defined functions Discover performance tuning tips and Hive best practices Who This Book Is For Developers, companies, and professionals who deal with large amounts of data and could use software that can efficiently manage large volumes of input. It is assumed that readers have the ability to work with SQL.

Sams Teach Yourself Big Data Analytics with Microsoft HDInsight in 24 Hours

Dataflow Scripting with Hadoop

Hadoop: The Definitive Guide

Apache Spark in 24 Hours, Sams Teach Yourself Programming Hive

Beginning Apache Spark 2

The need to handle increasingly larger data volumes is one factor driving the adoption of a new class of

nonrelational “NoSQL” databases. Advocates of NoSQL databases claim they can be used to build systems that are more performant, scale better, and are easier to program. NoSQL Distilled is a concise but thorough introduction to this rapidly emerging technology. Pramod J. Sadalage and Martin Fowler explain how NoSQL databases work and the ways that they may be a superior alternative to a traditional RDBMS. The authors provide a fast-paced guide to the concepts you need to know in order to evaluate whether NoSQL databases are right for your needs and, if so, which technologies you should explore further. The first part of the book concentrates on core concepts, including schemaless data models, aggregates, new distribution models, the CAP theorem, and map-reduce. In the second part, the authors explore architectural and design issues associated with implementing NoSQL. They also present realistic use cases that demonstrate NoSQL databases at work and feature representative examples using Riak, MongoDB, Cassandra, and Neo4j. In addition, by drawing on Pramod Sadalage's pioneering work, NoSQL Distilled shows how to implement evolutionary design with schema migration: an essential technique for applying NoSQL databases. The book concludes by describing how NoSQL is ushering in a new age of Polyglot Persistence, where multiple data-storage worlds coexist, and

Download Free Hadoop In 24 Hours Sams Teach Yourself

architects can choose the technology best optimized for each type of data access.

An easy-to-follow Apache Hadoop administrator's guide filled with practical screenshots and explanations for each step and configuration. This book is great for administrators interested in setting up and managing a large Hadoop cluster. If you are an administrator, or want to be an administrator, and you are ready to build and maintain a production-level cluster running CDH5, then this book is for you. "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."--Publisher's description.

In just 24 lessons of one hour or less, Sams Teach Yourself Windows PowerShell in 24 Hours helps you streamline all facets of Windows administration, supercharging your effectiveness as an IT

Download Free Hadoop In 24 Hours Sams Teach Yourself

professional or power user. This book's straightforward, step-by-step approach shows you how to build and run scripts, extend Windows PowerShell reach, manage computers remotely, and automate a wide variety of tasks on any modern Windows server or client. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success! Step-by-step instructions carefully walk you through the most common PowerShell scripting tasks. 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, tips, and cautions point out shortcuts, pitfalls, and solutions. Learn how to... Install, configure, and explore Windows PowerShell (including updates for PowerShell 5) Leverage .NET's remarkable power and scope with easy-to-use cmdlets Build new scripts with the console, ISE visual tools, and other popular hosts Apply best practices for writing more reliable, flexible, team-friendly scripts Work effectively with the pipeline, objects, and data Extend Windows PowerShell reach via providers, drives, and output Run external tools like ping, ipconfig, and tracert from within PowerShell Remotely manage computers with basic and advanced remoting (WinRM), WMI, and Regex Configure Windows devices across the web with PSWA Sort, filter, measure, format, export, and convert script output

Download Free Hadoop In 24 Hours Sams Teach Yourself

Run Windows PowerShell flexibly, using background and scheduled jobs Customize your environment with profile scripts, alternate credentials, thirdparty tools, and packages Strengthen your control over Windows systems and services with Desired State Configuration Administer key tasks on SQL Server, SharePoint, and other Windows servers Master PowerShell skills needed to earn many Microsoft certifications

Hadoop For Dummies

Enterprise Data Workflows with Cascading

Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem

Data Analytics with Hadoop

Sams Teach Yourself

Microsoft SQL Server 2012 Unleashed

Buy the print version of Microsoft SQL Server 2012 Unleashed and get the eBook version for free! eBook version includes chapters 44-60 not included in the print. See inside the book for access code and details. With up-to-the-minute content, this is the industry's most complete, useful guide to SQL Server 2012. You'll find start-to-finish coverage of SQL Server's core database server and management capabilities: all the real-world information, tips, guidelines, and samples you'll need to create and manage complex database solutions. The additional

Download Free Hadoop In 24 Hours Sams Teach Yourself

online chapters add extensive coverage of SQL Server Integration Services, Reporting Services, Analysis Services, T-SQL programming, .NET Framework integration, and much more. Authored by four expert SQL Server administrators, designers, developers, architects, and consultants, this book reflects immense experience with SQL Server in production environments. Intended for intermediate-to-advanced-level SQL Server professionals, it focuses on the product's most complex and powerful capabilities, and its newest tools and features. Understand SQL Server 2012's newest features, licensing changes, and capabilities of each edition Manage SQL Server 2012 more effectively with SQL Server Management Studio, the SQLCMD command-line query tool, and Powershell Use Policy-Based Management to centrally configure and operate SQL Server Utilize the new Extended Events trace capabilities within SSMS Maximize performance by optimizing design, queries, analysis, and workload management Implement new best practices for SQL Server high availability Deploy AlwaysOn Availability Groups and Failover Cluster Instances to achieve enterprise-class availability and disaster recovery Leverage new business intelligence improvements, including

Download Free Hadoop In 24 Hours Sams Teach Yourself

Master Data Services, Data Quality Services and Parallel Data Warehouse Deliver better full-text search with SQL Server 2012's new Semantic Search Improve reporting with new SQL Server 2012 Reporting Services features Download the following from

informit.com/title/9780672336928: Sample databases and code examples

This is the eBook of the printed book and may not include any media, website access codes, or print supplements that may come packaged with the bound book. The Comprehensive, Up-to-Date Apache Hadoop Administration Handbook and Reference “Sam Alapati has worked with production Hadoop clusters for six years. His unique depth of experience has enabled him to write the go-to resource for all administrators looking to spec, size, expand, and secure production Hadoop clusters of any size.”

—Paul Dix, Series Editor In Expert Hadoop® Administration, leading Hadoop administrator Sam R. Alapati brings together authoritative knowledge for creating, configuring, securing, managing, and optimizing production Hadoop clusters in any environment. Drawing on his experience with large-scale Hadoop administration, Alapati integrates action-oriented advice with carefully researched

Download Free Hadoop In 24 Hours Sams Teach Yourself

explanations of both problems and solutions. He covers an unmatched range of topics and offers an unparalleled collection of realistic examples. Alapati demystifies complex Hadoop environments, helping you understand exactly what happens behind the scenes when you administer your cluster. You'll gain unprecedented insight as you walk through building clusters from scratch and configuring high availability, performance, security, encryption, and other key attributes. The high-value administration skills you learn here will be indispensable no matter what Hadoop distribution you use or what Hadoop applications you run. Understand Hadoop's architecture from an administrator's standpoint Create simple and fully distributed clusters Run MapReduce and Spark applications in a Hadoop cluster Manage and protect Hadoop data and high availability Work with HDFS commands, file permissions, and storage management Move data, and use YARN to allocate resources and schedule jobs Manage job workflows with Oozie and Hue Secure, monitor, log, and optimize Hadoop Benchmark and troubleshoot Hadoop

Data is bigger, arrives faster, and comes in a variety of formats—and it all needs

Download Free Hadoop In 24 Hours Sams Teach Yourself

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 walk-throughs, 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*

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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

MapReduce Design Patterns

Hadoop 2 Quick-Start Guide

Sams Teach Yourself Drupal in 24 Hours

Apache Oozie

Expert Hadoop 2 Administration

Windows PowerShell in 24 Hours, Sams Teach Yourself

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 MLlib, 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 MLlib 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*

Until now, design patterns for the MapReduce framework have been scattered among various research papers, blogs, and books. This handy guide brings together a unique collection of valuable MapReduce patterns that will save you time and effort regardless of the domain, language, or development framework you're using. Each pattern is explained in context, with pitfalls and caveats clearly

identified to help you avoid common design mistakes when modeling your big data architecture. This book also provides a complete overview of MapReduce that explains its origins and implementations, and why design patterns are so important. All code examples are written for Hadoop.

Summarization patterns: get a top-level view by summarizing and grouping data Filtering patterns: view data subsets such as records generated from one user Data organization

patterns: reorganize data to work with other systems, or to make MapReduce analysis easier Join patterns: analyze different datasets together to discover interesting

relationships Metapatterns: piece together several patterns to solve multi-stage problems, or to perform several analytics in the same job Input and output patterns: customize the way you use Hadoop to load or store data

"A clear exposition of MapReduce programs for common data processing patterns—this book is indispensable for anyone using Hadoop."

--Tom White, author of Hadoop: The Definitive Guide

In just 24 sessions of one hour or less, Sams Teach Yourself Go in 24 Hours will help new and experienced programmers build software that's simpler, more reliable, and far more scalable. This book's straightforward, step-by-step approach guides you from setting up your environment through testing and deploying powerful solutions. Using practical examples, expert Go developer George Ornbo walks you through Go's fundamental constructs, demonstrates its breakthrough features for concurrent and network programming, and illuminates Go's powerful new idioms. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success. Step-by-step instructions

Download Free Hadoop In 24 Hours Sams Teach Yourself

carefully walk you through the most common Go programming tasks and techniques Quizzes and exercises help you test your knowledge and stretch your skills Practical, hands-on examples show you how to apply what you learn Notes and Tips point out shortcuts, solutions, and problems to avoid Two bonus chapters available online: Hour 25, “Creating a RESTful JSON API,” and Hour 26 “Creating a TCP Chat Server” Learn how to... · Get productive quickly with Go development tools and web servers · Master core features, including strings, functions, structs, and methods · Work with types, variables, functions, and control structures · Make the most of Go’s arrays, slices, and maps · Write powerful concurrent software with Goroutines and channels · Handle program errors smoothly · Promote code reuse with packages · Master Go’s unique idioms for highly effective coding · Use regular expressions and time/date functions · Test and benchmark Go code · Write basic command-line programs, HTTP servers, and HTTP clients · Efficiently move Go code into production · Build basic TCP chat servers and JSON APIs Register your book at informit.com/register for convenient access to the two bonus chapters online, downloads, updates, and/or corrections as they become available.

A Brief Guide to the Emerging World of Polyglot Persistence NoSQL Distilled

Big Data Analytics with Microsoft HDInsight in 24 Hours, Big Data, Hadoop, and Microsoft Azure for Better Business Intelligence

An Introduction for Data Scientists

Data Analytics with Spark Using Python

Go in 24 Hours, Sams Teach Yourself

Download Free Hadoop In 24 Hours Sams Teach Yourself

Get Started Fast with Apache Hadoop® 2, YARN, and Today's Hadoop Ecosystem With Hadoop 2.x and YARN, Hadoop moves beyond MapReduce to become practical for virtually any type of data processing. Hadoop 2.x and the Data Lake concept represent a radical shift away from conventional approaches to data usage and storage. Hadoop 2.x installations offer unmatched scalability and breakthrough extensibility that supports new and existing Big Data analytics processing methods and models. Hadoop® 2 Quick-Start Guide is the first easy, accessible guide to Apache Hadoop 2.x, YARN, and the modern Hadoop ecosystem. Building on his unsurpassed experience teaching Hadoop and Big Data, author Douglas Eadline covers all the basics you need to know to install and use Hadoop 2 on personal computers or servers, and to navigate the powerful technologies that complement it. Eadline concisely introduces and explains every key Hadoop 2 concept, tool, and service, illustrating each with a simple "beginning-to-end" example and identifying trustworthy, up-to-date resources for learning more. This guide is ideal if you want to learn about Hadoop 2 without getting mired in technical details. Douglas Eadline will bring you up to speed quickly, whether you're a user, admin, devops specialist, programmer, architect, analyst, or data scientist. Coverage Includes Understanding what Hadoop 2 and YARN do, and how they improve on Hadoop 1 with MapReduce Understanding Hadoop-based Data Lakes versus RDBMS Data Warehouses Installing Hadoop 2 and core services on Linux machines, virtualized sandboxes, or clusters Exploring the Hadoop Distributed File System (HDFS) Understanding the essentials of

Download Free Hadoop In 24 Hours Sams Teach Yourself

MapReduce and YARN application programming
Simplifying programming and data movement with Apache Pig, Hive, Sqoop, Flume, Oozie, and HBase
Observing application progress, controlling jobs, and managing workflows
Managing Hadoop efficiently with Apache Ambari—including recipes for HDFS to NFSv3 gateway, HDFS snapshots, and YARN configuration
Learning basic Hadoop 2 troubleshooting, and installing Apache Hue and Apache Spark

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

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 MLlib 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

Download Free Hadoop In 24 Hours Sams Teach Yourself

including HDFS, Hive, JSON, and S3 Master advanced topics like data partitioning and shared variables Starter Kit Includes C++ compiler and IDE for Windows, Mac & Linux In just 24 lessons of one hour or less, you can learn the basics of programming with C++—one of the most popular and powerful programming languages ever created. Using a straightforward, step-by-step approach, this fast and friendly tutorial teaches you everything you need to know, from installing and using a compiler, to debugging the programs you've created, to what's coming in C++0x, the next version of C++. Each lesson builds on what you've already learned, giving you a solid understanding of the basics of C++ programming concepts and techniques. Step-by-step instructions carefully walk you through the most common C++ programming tasks Quizzes and Exercises at the end of each chapter help you test yourself to make sure you're ready to go on Starter Kit software provides everything you need to create and compile C++ programs on any platform—Windows, Mac or Linux Learn how to... Install and use a C++ compiler for Windows, Mac OS X or Linux Build object-oriented programs in C++ Master core C++ concepts such as functions, classes, arrays, and pointers Add rich functionality with linked lists and templates Debug your programs for flawless code Learn exception and error-handling techniques Discover what's new in C++0x, the next version of C++ Jesse Liberty is the author of numerous books on software development, including best selling titles on C++ and .NET. He is the president of Liberty Associates, Inc. where he provides custom programming, consulting, and training. Rogers Cadenhead

Download Free Hadoop In 24 Hours Sams Teach Yourself

is a web application developer who has written many books on Internet-related topics, including Teach Yourself Java in 24 Hours. He maintains this book's official website at <http://cplusplus.cadenhead.org>. CD-ROM Includes C++ compiler Visual development environment for Windows, Mac and Linux Source code for the book's examples Register your book at informit.com/register for convenient access to updates and corrections as they become available

Lightning-Fast Big Data Analysis

Spark: The Definitive Guide

Streamlined Enterprise Data Management and Analysis

Tableau For Dummies

Building Effective Algorithms and Analytics for Hadoop and Other Systems

Sams Teach Yourself C++ in 24 Hours

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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 MLlib and Spark ML machine learning libraries Spark's Streaming components and external community packages

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

Download Free Hadoop In 24 Hours Sams Teach Yourself

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.

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
Make your data work for you! Tableau For Dummies brings order to the chaotic world of data. Understanding your data and organizing it into formats and visualizations that make sense to you are crucial to making a real impact on your business with the information that's already at your fingertips. This easy-to-use reference explores the user interface, and guides you through the process of connecting your data sources to the software. Additionally, this approachable, yet comprehensive text shows you how to use graphs, charts, and other images to bring visual interest to your data, how to create dashboards from multiple data sources, and how to export the visualizations that you have developed into multiple formats that translate into positive change for your business. The mission of Tableau Software is to grant you access to data that, when put into action, will help you build your company. Learning to use the data available to you helps you make informed, grounded business decisions that can spell success for your company. Navigate the user interface to efficiently access the features you need Connect to various spreadsheets, databases, and other data sources to create a multi-dimensional snapshot of your business Develop visualizations with easy to use drag and drop features Start building your data with templates and sample workbooks to spark your creativity and help you organize your information Tableau For Dummies is a step-by-step resource that helps you make sense of the data landscape—and put your data to work in support of your business.

Next Generation Systems Programming with Golang

Programming Pig

Cloudera Administration Handbook

Big Data Analytics with Microsoft HDInsight in 24 Hours, Sams Teach Yourself

Learning Spark

A Guide to Hadoop's Data Warehouse System

Follow this handbook to build, configure, tune, and secure Apache Cassandra databases. Start with the installation of Cassandra and move on to the creation of a single instance, and then a cluster of Cassandra databases. Cassandra is increasingly a key player in many big data environments, and this book shows you how to use Cassandra with Apache Spark, a popular big data processing framework. Also covered are day-to-day topics of importance such as the backup and recovery of Cassandra databases, using the right compression and compaction strategies, and loading and unloading data. Expert Apache Cassandra Administration provides numerous step-by-step examples starting with the basics of a Cassandra database, and going all the way through backup and recovery, performance optimization, and monitoring and securing the data. The book serves as an authoritative and comprehensive guide to the building and management of simple to complex Cassandra databases. The book: Takes you through building a Cassandra database from installation of the software and creation of a single database, through to complex clusters and data centers Provides numerous examples of actual commands in a real-life Cassandra environment that show how to confidently configure, manage, troubleshoot, and tune Cassandra databases Shows how to use the Cassandra configuration

properties to build a highly stable, available, and secure Cassandra database that always operates at peak efficiency What You'll Learn Install the Cassandra software and create your first database Understand the Cassandra data model, and the internal architecture of a Cassandra database Create your own Cassandra cluster, step-by-step Run a Cassandra cluster on Docker Work with Apache Spark by connecting to a Cassandra database Deploy Cassandra clusters in your data center, or on Amazon EC2 instances Back up and restore mission-critical Cassandra databases Monitor, troubleshoot, and tune production Cassandra databases, and cut your spending on resources such as memory, servers, and storage Who This Book Is For Database administrators, developers, and architects who are looking for an authoritative and comprehensive single volume for all their Cassandra administration needs. Also for administrators who are tasked with setting up and maintaining highly reliable and high-performing Cassandra databases. An excellent choice for big data administrators, database administrators, architects, and developers who use Cassandra as their key data store, to support high volume online transactions, or as a decentralized, elastic data store.

Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an

MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between business stakeholders and data scientists, but also how participate intelligently in your company's data science projects. You'll also discover how to think data-analytically, and fully appreciate how data science methods can support business decision-making. Understand how data science fits in your organization—and how you can use it for competitive advantage Treat data as a business asset that requires careful investment if you're to gain real value Approach business problems data-analytically, using the data-mining process to gather good data in the most appropriate way Learn general concepts for actually extracting knowledge from data Apply data science principles when interviewing data science job candidates

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.

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,

Download Free Hadoop In 24 Hours Sams Teach Yourself

Linux-or all of them! Sams Teach Yourself Unreal Engine 4 Game Development in 24 Hours' 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. Organize new projects and work with the Gameplay Framework Master Unreal's units and control systems Import 3D models and work with the Static Mesh Editor Create new landscapes and use Unreal's foliage system Bring characters and creatures to life with the Persona Editor Apply materials and build lighting Integrate and modify audio with the Unreal Sound Cue Editor Craft particle effects and simulate physics Set up and react to player inputs Build levels and entirely new worlds Get started with powerful Blueprint visual scripting system Script an arcade game from start to finish Create events that respond to player actions Spawn Actors during gameplay Design and create action-based encounters Optimize games for mobile devices and touch-based inputs Build menus with Unreal's UMG UI Designer Prepare your game for deployment Step-by-step instructions carefully walk you through the most common Unreal Engine 4 game development tasks. 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. All the project files and assets you'll need are available for download, including "before-and-after" files

Download Free Hadoop In 24 Hours Sams Teach Yourself

demonstrating initial setup and proper completion for every exercise.

High Performance Spark

Best Practices for Scaling and Optimizing Apache Spark