

# Online Library Big Data SMACK A Guide To Apache Spark Mesos Akka Cassandra And Kafka Pdf Free Copy

Apache Apache Server for Windows Little Black Book Guide to Apache Programmer's Guide to Apache Thrift The Definitive Guide to Apache mod\_rewrite Spark: The Definitive Guide Apache Ignite Quick Start Guide Apache Superset Quick Start Guide HTTP: The Definitive Guide Apache Roller 4. 0, Beginner's Guide Big Data SMACK Apache Tomcat: The Definitive Guide

Apache Essentials Apache Apache Spark Quick Start Guide Kafka: The Definitive Guide Next-Generation Big Data Apache Administrator's Handbook Apache Server 2.0 Programmer's Guide to Apache Thrift Apache Tomcat 7 User Guide Hadoop 2 Quick-Start Guide Apache Pulsar in Action Data Pipelines with Apache Airflow Cassandra: The Definitive Guide Apache HTTP

Server Apache HTTP Server Apache Security Apache Hadoop 3 Quick Start Guide Apache Cookbook The Definitive Guide to Apache MyFaces and Facelets Apache Kafka Quick Start Guide Apache Buildr Apache Buildr Apache Solr Beginner's Guide Quickstart Apache Axis2 Machine Learning with Apache Spark Quick Start Guide Apache Solr Beginner's Guide

Mastering Spark with R

*Apache Server for Windows Little Black Book* Jul 20 2023  
Readers will find hundreds of immediate solutions for turning Apache into a powerhouse Web server. Key topics include setting up a virtual Web site, mastering security, creating optimized CGI scripts, tuning and customizing Apache servers, using the Apache Module API, setting up Apache as a proxy server, and setting up commerce sites.

**Apache Server 2.0** Jan 02 2022 Get up-to-speed quickly on running a network using Apache Server with this introductory guide. Aimed at administrators new to Apache,

this practical guide offers tips for running Apache on multiple platforms and contains clear, straightforward advice for administering Apache effectively in today's Web environment.

Apache Kafka Quick Start Guide Nov 19 2020 Process large volumes of data in real-time while building high performance and robust data stream processing pipeline using the latest Apache Kafka 2.0 Key Features Solve practical large data and processing challenges with Kafka Tackle data processing challenges like late events, windowing, and watermarking Understand real-time streaming applications processing using Schema

registry, Kafka connect, Kafka streams, and KSQLBook Description Apache Kafka is a great open source platform for handling your real-time data pipeline to ensure high-speed filtering and pattern matching on the fly. In this book, you will learn how to use Apache Kafka for efficient processing of distributed applications and will get familiar with solving everyday problems in fast data and processing pipelines. This book focuses on programming rather than the configuration management of Kafka clusters or DevOps. It starts off with the installation and setting up the development environment, before quickly moving on to performing fundamental

messaging operations such as validation and enrichment. Here you will learn about message composition with pure Kafka API and Kafka Streams. You will look into the transformation of messages in different formats, such as text, binary, XML, JSON, and AVRO. Next, you will learn how to expose the schemas contained in Kafka with the Schema Registry. You will then learn how to work with all relevant connectors with Kafka Connect. While working with Kafka Streams, you will perform various interesting operations on streams, such as windowing, joins, and aggregations. Finally, through KSQL, you will learn how to retrieve, insert,

modify, and delete data streams, and how to manipulate watermarks and windows. What you will learnHow to validate data with KafkaAdd information to existing data flowsGenerate new information through message compositionPerform data validation and versioning with the Schema RegistryHow to perform message Serialization and DeserializationHow to perform message Serialization and DeserializationProcess data streams with Kafka StreamsUnderstand the duality between tables and streams with KSQLWho this book is for This book is for developers who want to quickly master the

practical concepts behind Apache Kafka. The audience need not have come across Apache Kafka previously; however, a familiarity of Java or any JVM language will be helpful in understanding the code in this book.  
Big Data SMACK Oct 11 2022  
Learn how to integrate full-stack open source big data architecture and to choose the correct technology—Scala/Spark, Mesos, Akka, Cassandra, and Kafka—in every layer. Big data architecture is becoming a requirement for many different enterprises. So far, however, the focus has largely been on collecting, aggregating, and crunching large data sets in a

timely manner. In many cases now, organizations need more than one paradigm to perform efficient analyses. Big Data SMACK explains each of the full-stack technologies and, more importantly, how to best integrate them. It provides detailed coverage of the practical benefits of these technologies and incorporates real-world examples in every situation. This book focuses on the problems and scenarios solved by the architecture, as well as the solutions provided by every technology. It covers the six main concepts of big data architecture and how integrate, replace, and reinforce every layer: The language: Scala The engine:

Spark (SQL, MLib, Streaming, GraphX) The container: Mesos, Docker The view: Akka The storage: Cassandra The message broker: Kafka What You Will Learn: Make big data architecture without using complex Greek letter architectures Build a cheap but effective cluster infrastructure Make queries, reports, and graphs that business demands Manage and exploit unstructured and No-SQL data sources Use tools to monitor the performance of your architecture Integrate all technologies and decide which ones replace and which ones reinforce Who This Book Is For: Developers, data architects, and data scientists looking to

integrate the most successful big data open stack architecture and to choose the correct technology in every layer

**Quickstart Apache Axis2** Jul 16 2020 This is a step by step practical guide for developing web services using Apache Axis2. There are lot of real-life examples, which makes this book a good learning material. This book is for Java developers who are interested in building web services using Apache Axis2. The book presumes that you are familiar with web standards like SOAP, WSDL and XML parsing.

**Apache Hadoop 3 Quick Start Guide** Feb 20 2021 A fast paced guide that will help

you learn about Apache Hadoop 3 and its ecosystem  
Key Features Set up, configure and get started with Hadoop to get useful insights from large data sets Work with the different components of Hadoop such as MapReduce, HDFS and YARN Learn about the new features introduced in Hadoop 3 Book Description Apache Hadoop is a widely used distributed data platform. It enables large datasets to be efficiently processed instead of using one large computer to store and process the data. This book will get you started with the Hadoop ecosystem, and introduce you to the main technical topics, including MapReduce, YARN, and HDFS.

The book begins with an overview of big data and Apache Hadoop. Then, you will set up a pseudo Hadoop development environment and a multi-node enterprise Hadoop cluster. You will see how the parallel programming paradigm, such as MapReduce, can solve many complex data processing problems. The book also covers the important aspects of the big data software development lifecycle, including quality assurance and control, performance, administration, and monitoring. You will then learn about the Hadoop ecosystem, and tools such as Kafka, Sqoop, Flume, Pig, Hive, and HBase. Finally, you will look at advanced

topics, including real time streaming using Apache Storm, and data analytics using Apache Spark. By the end of the book, you will be well versed with different configurations of the Hadoop 3 cluster. What you will learn Store and analyze data at scale using HDFS, MapReduce and YARN Install and configure Hadoop 3 in different modes Use Yarn effectively to run different applications on Hadoop based platform Understand and monitor how Hadoop cluster is managed Consume streaming data using Storm, and then analyze it using Spark Explore Apache Hadoop ecosystem components, such as Flume,

Sqoop, HBase, Hive, and Kafka  
Who this book is for Aspiring  
Big Data professionals who  
want to learn the essentials of  
Hadoop 3 will find this book to  
be useful. Existing Hadoop  
users who want to get up to  
speed with the new features  
introduced in Hadoop 3 will  
also benefit from this book.  
Having knowledge of Java  
programming will be an added  
advantage.

The Definitive Guide to Apache  
mod\_rewrite Apr 17 2023 \* The  
Definitive Guide to Apache  
mod\_rewrite is the first book  
devoted to mastering  
mod\_rewrite, Apache's most  
powerful (and confusing)  
module. \* It is authored by  
noted Apache expert and

Apache Software Foundation  
member Rich Bowen, who  
draws on several years of  
extensive experience  
administering the Apache  
server as well as speaking and  
writing regularly on the topic. \*  
This book covers all matters  
pertinent to mod\_rewrite:  
installation and configuration,  
basic and conditional rewrites,  
access control, maintaining  
virtual hosts, and proxying.  
Guide to Apache Jun 19 2023  
Apache Web Server is a hands-  
on, practical guide that can be  
used to master installation,  
configuration, and  
troubleshooting of the Apache  
Web server. The book takes a  
multiplatform approach,  
teaching the reader how to run

Apache with any Linux  
distribution or Windows. Key  
topics include installation,  
configuration, modularity,  
virtual hosting, security and  
SSL, access controls, Server  
Side Includes (SSI) and CGI,  
HTTP caching, and  
troubleshooting. The book also  
provides examples of sample  
commands and output, plus  
screen shots of related  
graphical utilities.

**Data Pipelines with Apache  
Airflow** Jul 28 2021 This book  
teaches you how to build and  
maintain effective data  
pipelines. You'll explore the  
most common usage patterns,  
including aggregating multiple  
data sources, connecting to and  
from data lakes, and cloud

deployment. --  
*Apache* Jun 07 2022 Describes the history of the Web server platform and covers downloading and compiling, configuring and running the program on UNIX, writing specialized modules, and establishing security routines.  
*Apache Buildr* Sep 17 2020 Do the Apache Buildr decisions we make today help people and the planet tomorrow? What is the purpose of Apache Buildr in relation to the mission? How much are sponsors, customers, partners, stakeholders involved in Apache Buildr? In other words, what are the risks, if Apache Buildr does not deliver successfully? How can you negotiate Apache Buildr

successfully with a stubborn boss, an irate client, or a deceitful coworker? What tools do you use once you have decided on a Apache Buildr strategy and more importantly how do you choose? Defining, designing, creating, and implementing a process to solve a business challenge or meet a business objective is the most valuable role... In EVERY company, organization and department. Unless you are talking a one-time, single-use project within a business, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex

enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc... - they are the people who rule the future. They are the person who asks the right questions to make Apache Buildr investments work better. This Apache Buildr All-Inclusive Self-Assessment enables You to be that person. All the tools you need to an in-depth Apache Buildr Self-

Assessment. Featuring 697 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Apache Buildr improvements can be made. In using the questions you will be better able to: - diagnose Apache Buildr projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent advances in Apache Buildr and process design strategies into practice according to best practice guidelines Using a Self-

Assessment tool known as the Apache Buildr Scorecard, you will develop a clear picture of which Apache Buildr areas need attention. Your purchase includes access details to the Apache Buildr self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows your organization exactly what to do next. Your exclusive instant access details can be found in your book. **Apache Ignite Quick Start Guide** Feb 15 2023 Build efficient, high-performance & scalable systems to process large volumes of data with Apache Ignite Key FeaturesUnderstand Apache Ignite's in-memory

technologyCreate High-Performance app components with IgniteBuild a real-time data streaming and complex event processing systemBook Description Apache Ignite is a distributed in-memory platform designed to scale and process large volume of data. It can be integrated with microservices as well as monolithic systems, and can be used as a scalable, highly available and performant deployment platform for microservices. This book will teach you to use Apache Ignite for building a high-performance, scalable, highly available system architecture with data integrity. The book takes you through the basics of Apache



Ignite and in-memory technologies. You will learn about installation and clustering Ignite nodes, caching topologies, and various caching strategies, such as cache aside, read and write through, and write behind. Next, you will delve into detailed aspects of Ignite's data grid: web session clustering and querying data. You will learn how to process large volumes of data using compute grid and Ignite's map-reduce and executor service. You will learn about the memory architecture of Apache Ignite and monitoring memory and caches. You will use Ignite for complex event processing, event streaming, and the time-

series predictions of opportunities and threats. Additionally, you will go through off-heap and on-heap caching, swapping, and native and Spring framework integration with Apache Ignite. By the end of this book, you will be confident with all the features of Apache Ignite 2.x that can be used to build a high-performance system architecture. What you will learn Use Apache Ignite's data grid and implement web session clustering Gain high performance and linear scalability with in-memory distributed data processing Create a microservice on top of Apache Ignite that can scale and

perform Perform ACID-compliant CRUD operations on an Ignite cache Retrieve data from Apache Ignite's data grid using SQL, Scan and Lucene Text query Explore complex event processing concepts and event streaming Integrate your Ignite app with the Spring framework Who this book is for The book is for Big Data professionals who want to learn the essentials of Apache Ignite. Prior experience in Java is necessary.

[The Definitive Guide to Apache MyFaces and Facelets](#) Dec 21 2020

**Programmer's Guide to Apache Thrift** May 18 2023  
Summary Programmer's Guide to Apache Thrift provides

comprehensive coverage of the Apache Thrift framework along with a developer's-eye view of modern distributed application architecture. Foreword by Jens Geyer. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Thrift-based distributed software systems are built out of communicating components that use different languages, protocols, and message types. Sitting between them is Thrift, which handles data serialization, transport, and service implementation. Thrift supports many client and server environments and a host of languages ranging from PHP

to JavaScript, and from C++ to Go. About the Book Programmer's Guide to Apache Thrift provides comprehensive coverage of distributed application communication using the Thrift framework. Packed with code examples and useful insight, this book presents best practices for multi-language distributed development. You'll take a guided tour through transports, protocols, IDL, and servers as you explore programs in C++, Java, and Python. You'll also learn how to work with platforms ranging from browser-based clients to enterprise servers. What's inside Complete coverage of Thrift's IDL Building and

serializing complex user-defined types Plug-in protocols, transports, and data compression Creating cross-language services with RPC and messaging systems About the Reader Readers should be comfortable with a language like Python, Java, or C++ and the basics of service-oriented or microservice architectures. About the Author Randy Abernethy is an Apache Thrift Project Management Committee member and a partner at RX-M. Table of Contents PART 1 - APACHE THRIFT OVERVIEW Introduction to Apache Thrift Apache Thrift architecture Building, testing, and debugging PART 2 -

PROGRAMMING APACHE THRIFT Moving bytes with transports Serializing data with protocols Apache Thrift IDL User-defined types Implementing services Handling exceptions Servers

PART 3 - APACHE THRIFT LANGUAGES Building clients and servers with C++ Building clients and servers with Java Building C# clients and servers with .NET Core and Windows Building Node.js clients and servers Apache Thrift and JavaScript Scripting Apache Thrift Thrift in the enterprise

**Next-Generation Big Data**  
Mar 04 2022 Utilize this practical and easy-to-follow guide to modernize traditional enterprise data warehouse and

business intelligence environments with next-generation big data technologies. Next-Generation Big Data takes a holistic approach, covering the most important aspects of modern enterprise big data. The book covers not only the main technology stack but also the next-generation tools and applications used for big data warehousing, data warehouse optimization, real-time and batch data ingestion and processing, real-time data visualization, big data governance, data wrangling, big data cloud deployments, and distributed in-memory big data computing. Finally, the book has an extensive and

detailed coverage of big data case studies from Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard. What You'll Learn Install Apache Kudu, Impala, and Spark to modernize enterprise data warehouse and business intelligence environments, complete with real-world, easy-to-follow examples, and practical advice Integrate HBase, Solr, Oracle, SQL Server, MySQL, Flume, Kafka, HDFS, and Amazon S3 with Apache Kudu, Impala, and Spark Use StreamSets, Talend, Pentaho, and CDAP for real-time and batch data ingestion and processing Utilize Trifacta, Alteryx, and Datameer for data

wrangling and interactive data processing Turbocharge Spark with Alluxio, a distributed in-memory storage platform Deploy big data in the cloud using Cloudera Director Perform real-time data visualization and time series analysis using Zoomdata, Apache Kudu, Impala, and Spark Understand enterprise big data topics such as big data governance, metadata management, data lineage, impact analysis, and policy enforcement, and how to use Cloudera Navigator to perform common data governance tasks Implement big data use cases such as big data warehousing, data warehouse optimization, Internet of Things, real-time

data ingestion and analytics, complex event processing, and scalable predictive modeling Study real-world big data case studies from innovative companies, including Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard Who This Book Is For BI and big data warehouse professionals interested in gaining practical and real-world insight into next-generation big data processing and analytics using Apache Kudu, Impala, and Spark; and those who want to learn more about other advanced enterprise topics **Apache Solr Beginner's Guide** Aug 17 2020 Search is everywhere, yet it is one of the

most misunderstood functionalities of the IT industry. In Apache Solr, author Xavier Morera guides you through the basics of this highly popular enterprise search tool. You'll learn how to set up an index and how to make it searchable, then query it with a simple enterprise search. Explanations for precision and recall are also included to help you ensure that relevant, accurate results have been returned. Custom UIs using SolrItas and SolrNet are also covered. This updated and expanded second edition of Book provides a user-friendly introduction to the subject, Taking a clear structural framework, it guides the reader

through the subject's core elements. A flowing writing style combines with the use of illustrations and diagrams throughout the text to ensure the reader understands even the most complex of concepts. This succinct and enlightening overview is a required reading for all those interested in the subject . We hope you find this book useful in shaping your future career & Business.

[Apache Cookbook](#) Jan 22 2021  
This is a collection of problems, solutions, and practical examples for webmasters, web administrators, programmers, and anyone who works with Apache.

[Apache Solr Beginner's Guide](#)  
May 14 2020 Written in a

friendly, example-driven format, the book includes plenty of step-by-step instructions and examples that are designed to help you get started with Apache Solr. This book is an entry level text into the wonderful world of Apache Solr. The book will center around a couple of simple projects such as setting up Solr and all the stuff that comes with customizing the Solr schema and configuration. This book is for developers looking to start using Apache Solr who are stuck or intimidated by the difficulty of setting it up and using it. For anyone wanting to embed a search engine in their site to help users navigate around the mammoth data

available this book is an ideal starting point. Moreover, if you are a data architect or a project manager and want to make some key design decisions, you will find that every example included in the book contains ideas usable in real-world contexts.

### **Hadoop 2 Quick-Start Guide**

Sep 29 2021 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 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 *Cassandra: The Definitive*

*Guide* Jun 26 2021 Imagine what you could do if scalability wasn't a problem. With this hands-on guide, you'll learn how the Cassandra database management system handles hundreds of terabytes of data while remaining highly available across multiple data centers. This expanded second edition—updated for Cassandra 3.0—provides the technical details and practical examples you need to put this database to work in a production environment. Authors Jeff Carpenter and Eben Hewitt demonstrate the advantages of Cassandra's non-relational design, with special attention to data modeling. If you're a developer, DBA, or application

architect looking to solve a database scaling issue or future-proof your application, this guide helps you harness Cassandra's speed and flexibility. Understand Cassandra's distributed and decentralized structure Use the Cassandra Query Language (CQL) and cqlsh—the CQL shell Create a working data model and compare it with an equivalent relational model Develop sample applications using client drivers for languages including Java, Python, and Node.js Explore cluster topology and learn how nodes exchange data Maintain a high level of performance in your cluster Deploy Cassandra on site, in the Cloud, or with

Docker Integrate Cassandra with Spark, Hadoop, Elasticsearch, Solr, and Lucene *Apache* Sep 10 2022

**Spark: The Definitive Guide** Mar 16 2023 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  
**Mastering Spark with R** Apr 12 2020 If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute

problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats



with ease from within Spark  
Learn about alternative  
modeling frameworks for graph  
processing, geospatial analysis,  
and genomics at scale Dive into  
advanced topics including  
custom transformations, real-  
time data processing, and  
creating custom Spark  
extensions

**Apache Superset Quick Start  
Guide** Jan 14 2023 Integrate  
open source data analytics and  
build business intelligence on  
SQL databases with Apache  
Superset. The quick, intuitive  
nature for data visualization in  
a web application makes it easy  
for creating interactive  
dashboards. Key Features Work  
with Apache Superset's rich set  
of data visualizations Create

interactive dashboards and  
data storytelling Easily explore  
dataBook Description Apache  
Superset is a modern, open  
source, enterprise-ready  
business intelligence (BI) web  
application. With the help of  
this book, you will see how  
Superset integrates with  
popular databases like  
Postgres, Google BigQuery,  
Snowflake, and MySQL. You  
will learn to create real time  
data visualizations and  
dashboards on modern web  
browsers for your organization  
using Superset. First, we look  
at the fundamentals of  
Superset, and then get it up  
and running. You'll go through  
the requisite installation,  
configuration, and deployment.

Then, we will discuss different  
columnar data types, analytics,  
and the visualizations available.  
You'll also see the security  
tools available to the  
administrator to keep your data  
safe. You will learn how to  
visualize relationships as  
graphs instead of coordinates  
on plain orthogonal axes. This  
will help you when you upload  
your own entity relationship  
dataset and analyze the dataset  
in new, different ways. You will  
also see how to analyze  
geographical regions by  
working with location data.  
Finally, we cover a set of  
tutorials on dashboard designs  
frequently used by analysts,  
business intelligence  
professionals, and developers.

What you will learn  
Get to grips with the fundamentals of data exploration using Superset  
Set up a working instance of Superset on cloud services like Google Compute Engine  
Integrate Superset with SQL databases  
Build dashboards with Superset  
Calculate statistics in Superset for numerical, categorical, or text data  
Understand visualization techniques, filtering, and grouping by aggregation  
Manage user roles and permissions in Superset  
Work with SQL Lab  
Who this book is for  
This book is for data analysts, BI professionals, and developers who want to learn Apache

Superset. If you want to create interactive dashboards from SQL databases, this book is what you need. Working knowledge of Python will be an advantage but not necessary to understand this book.

#### Tomcat: The Definitive Guide

Aug 09 2022 Jakarta  
Tomcat is not only the most commonly used open source servlet engine today, it's become the de facto standard by which other servlet engines are measured. Powerful and flexible, it can be used as a stand-alone web server or in conjunction with another server, like Apache or IIS, to run servlets or JSPs. But mastery of Tomcat is not easy: because it's as complex as it is

complete. Tomcat: The Definitive Guide answers vexing questions that users, administrators, and developers alike have been asking. This concise guide provides much needed information to help harness Tomcat's power and wealth of features. Tomcat: The Definitive Guide offers something for everyone who uses Tomcat. System and network administrators will find detailed instructions on installation, configuration, and maintenance. For users, it supplies insightful information on how to deploy Tomcat. And seasoned enterprise Java developers will have a complete reference to setting up, running, and using this

powerful softwareThe book begins with an introduction to the Tomcat server and includes an overview of the three types of server configurations: stand-alone, in-process, and out-of-process. The authors show how directories are laid out, cover the initial setup, and describe how to set the environment variables and modify the configuration files, concluding with common errors, problems, and solutions. In subsequent chapters, they cover: The server.xml configuration file Java Security manager Authentication schemes and Tomcat users The Secure Socket Layer (SSL) Tomcat JDBC Realms Installing servlets and Java Server Pages

Integrating Tomcat with Apache Advanced Tomcat configuration and much more.Tomcat: The Definitive Guide covers all major platforms, including Windows, Solaris, Linux, and Mac OS X, contains details on Tomcat configuration files, and has a quick-start guide to get developers up and running with Java servlets and JavaServer Pages. If you've struggled with this powerful yet demanding technology in the past, this book will provide the answers you need.

[Apache Security](#) Mar 24 2021 "The complete guide to securing your Apache web server"--Cover.

**Apache Administrator's**

**Handbook** Feb 03 2022 Apache Administrator's Handbook is a comprehensive manual on the Apache Web server. Designed to be a hands-on, comprehensive desk reference, it covers everything from installation to configuration, performance tuning, security, logging, and dynamic content. The book covers Apache 2.0 and also provides information on migrating from Apache 1.3 and earlier.

**Apache Buildr** Oct 19 2020 Do the Apache Buildr decisions we make today help people and the planet tomorrow? What is the purpose of Apache Buildr in relation to the mission? How much are sponsors, customers,

partners, stakeholders involved in Apache Buildr? In other words, what are the risks, if Apache Buildr does not deliver successfully? How can you negotiate Apache Buildr successfully with a stubborn boss, an irate client, or a deceitful coworker? What tools do you use once you have decided on a Apache Buildr strategy and more importantly how do you choose? Defining, designing, creating, and implementing a process to solve a business challenge or meet a business objective is the most valuable role... In EVERY company, organization and department. Unless you are talking a one-time, single-use project within a business, there

should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc... - they are the people who rule the future. They are the person who asks the right questions to make Apache

Buildr investments work better. This Apache Buildr All-Inclusive Self-Assessment enables You to be that person. All the tools you need to an in-depth Apache Buildr Self-Assessment. Featuring 697 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Apache Buildr improvements can be made. In using the questions you will be better able to: - diagnose Apache Buildr projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies

aligned with overall goals - integrate recent advances in Apache Buildr and process design strategies into practice according to best practice guidelines Using a Self-Assessment tool known as the Apache Buildr Scorecard, you will develop a clear picture of which Apache Buildr areas need attention. Your purchase includes access details to the Apache Buildr self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows your organization exactly what to do next. Your exclusive instant access details can be found in your book. [Apache Tomcat 7 User Guide](#) Oct 31 2021 Apache Tomcat (or

Jakarta Tomcat or simply Tomcat) is an open source servlet container developed by the Apache Software Foundation (ASF). Tomcat implements the Java Servlet and the JavaServer Pages (JSP) specifications.

**Apache Essentials** Jul 08 2022 \* Covers the most up-to-date versions of the technology. \* Teaches subject using real world tutorials. \* Subject is taught in a very concise, no-nonsense fashion, wasting little time on history and other fluff. Book is therefore compact and portable. It gives you just what you need. \* Demystifies what is sometimes thought of as an intimidating subject. \* Covers Apache for Windows,

Mac/Unix, and Linux. \* Full book support available online and via e-mail.

*Programmer's Guide to Apache Thrift* Dec 01 2021 Programmer's Guide to Apache Thrift provides comprehensive coverage of distributed application communication using the Thrift framework. Packed with code examples and useful insight, this book presents best practices for multi-language distributed development. You'll take a guided tour through transports, protocols, IDL, and servers as you explore programs in C++, Java, and Python. You'll also learn how to work with platforms ranging from browser-based clients to

enterprise servers.

### **Apache Roller 4. 0,**

**Beginner's Guide** Nov 12  
2022 A comprehensive, step-by-step guide on how to set up, customize, and market your blog using Apache Roller.

*Apache HTTP Server* May 26  
2021 How do the Apache HTTP Server results compare with the performance of your competitors and other organizations with similar offerings? Can Management personnel recognize the monetary benefit of Apache HTTP Server? Where do ideas that reach policy makers and planners as proposals for Apache HTTP Server strengthening and reform actually originate? How do we

go about Securing Apache HTTP Server? Think about the people you identified for your Apache HTTP Server project and the project responsibilities you would assign to them. what kind of training do you think they would need to perform these responsibilities effectively? Defining, designing, creating, and implementing a process to solve a business challenge or meet a business objective is the most valuable role... In EVERY company, organization and department. Unless you are talking a one-time, single-use project within a business, there should be a process. Whether that process is managed and implemented by humans, AI, or

a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc... - they are the people who rule the future. They are the person who asks the right questions to make Apache HTTP Server investments work better. This Apache HTTP Server All-Inclusive Self-

Assessment enables You to be that person. All the tools you need to an in-depth Apache HTTP Server Self-Assessment. Featuring 700 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Apache HTTP Server improvements can be made. In using the questions you will be better able to: - diagnose Apache HTTP Server projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent

advances in Apache HTTP Server and process design strategies into practice according to best practice guidelines Using a Self-Assessment tool known as the Apache HTTP Server Scorecard, you will develop a clear picture of which Apache HTTP Server areas need attention. Your purchase includes access details to the Apache HTTP Server self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows your organization exactly what to do next. Your exclusive instant access details can be found in your book.

**Machine Learning with**

**Apache Spark Quick Start Guide** Jun 14 2020 Combine advanced analytics including Machine Learning, Deep Learning Neural Networks and Natural Language Processing with modern scalable technologies including Apache Spark to derive actionable insights from Big Data in real-time Key FeaturesMake a hands-on start in the fields of Big Data, Distributed Technologies and Machine LearningLearn how to design, develop and interpret the results of common Machine Learning algorithmsUncover hidden patterns in your data in order to derive real actionable insights and business valueBook Description Every

person and every organization in the world manages data, whether they realize it or not. Data is used to describe the world around us and can be used for almost any purpose, from analyzing consumer habits to fighting disease and serious organized crime. Ultimately, we manage data in order to derive value from it, and many organizations around the world have traditionally invested in technology to help process their data faster and more efficiently. But we now live in an interconnected world driven by mass data creation and consumption where data is no longer rows and columns restricted to a spreadsheet, but an organic and evolving asset

in its own right. With this realization comes major challenges for organizations: how do we manage the sheer size of data being created every second (think not only spreadsheets and databases, but also social media posts, images, videos, music, blogs and so on)? And once we can manage all of this data, how do we derive real value from it? The focus of Machine Learning with Apache Spark is to help us answer these questions in a hands-on manner. We introduce the latest scalable technologies to help us manage and process big data. We then introduce advanced analytical algorithms applied to real-world use cases in order to

uncover patterns, derive actionable insights, and learn from this big data. What you will learn Understand how Spark fits in the context of the big data ecosystem Understand how to deploy and configure a local development environment using Apache Spark Understand how to design supervised and unsupervised learning models Build models to perform NLP, deep learning, and cognitive services using Spark ML libraries Design real-time machine learning pipelines in Apache Spark Become familiar with advanced techniques for processing a large volume of data by applying machine learning algorithms Who this book is for This book is aimed



at Business Analysts, Data Analysts and Data Scientists who wish to make a hands-on start in order to take advantage of modern Big Data technologies combined with Advanced Analytics.

### **Kafka: The Definitive Guide**

Apr 05 2022 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

### **Apache Spark Quick Start Guide**

May 06 2022 A practical guide for solving complex data processing challenges by applying the best optimizations techniques in Apache Spark.

Key Features Learn about the core concepts and the latest developments in Apache SparkMaster writing efficient big data applications with Spark's built-in modules for SQL, Streaming, Machine Learning and Graph analysisGet introduced to a variety of optimizations based on the actual experienceBook Description Apache Spark is a flexible framework that allows processing of batch and real-time data. Its unified engine has made it quite popular for big data use cases. This book will help you to get started with Apache Spark 2.0 and write big data applications for a variety of use cases. It will also introduce you to Apache Spark

- one of the most popular Big Data processing frameworks. Although this book is intended to help you get started with Apache Spark, but it also focuses on explaining the core concepts. This practical guide provides a quick start to the Spark 2.0 architecture and its components. It teaches you how to set up Spark on your local machine. As we move ahead, you will be introduced to resilient distributed datasets (RDDs) and DataFrame APIs, and their corresponding transformations and actions. Then, we move on to the life cycle of a Spark application and learn about the techniques used to debug slow-running applications. You will also go

through Spark's built-in modules for SQL, streaming, machine learning, and graph analysis. Finally, the book will lay out the best practices and optimization techniques that are key for writing efficient Spark applications. By the end of this book, you will have a sound fundamental understanding of the Apache Spark framework and you will be able to write and optimize Spark applications. What you will learnLearn core concepts such as RDDs, DataFrames, transformations, and moreSet up a Spark development environmentChoose the right APIs for your applicationsUnderstand Spark's architecture and the

execution flow of a Spark application Explore built-in modules for SQL, streaming, ML, and graph analysis Optimize your Spark job for better performance Who this book is for If you are a big data enthusiast and love processing huge amount of data, this book is for you. If you are data engineer and looking for the best optimization techniques for your Spark applications, then you will find this book helpful. This book also helps data scientists who want to implement their machine learning algorithms in Spark. You need to have a basic understanding of any one of the programming languages such as Scala, Python or Java.

*Apache Pulsar in Action* Aug 29 2021 Distributed applications demand reliable, high-performance messaging. The Apache Pulsar server-to-server messaging system provides a secure, stable platform without the need for a stream processing engine like Spark. Contributed by Yahoo to the Apache Foundation, Pulsar is mature and battle-tested, handling millions of messages per second for over three years at Yahoo. *Apache Pulsar in Action* is a comprehensive and practical guide to building high-traffic applications with Pulsar, delivering extreme levels of speed and durability. about the technology Pulsar is a streaming messaging system

designed for high performance server-to-server messaging. Built and tested under intense conditions at Yahoo, Pulsar has been proven in production and can handle millions of messages per second. Now free and open-source, Pulsar's unique architecture helps solve some of the challenges of modern development. Pulsar avoids latency in streaming data transmission, making it a powerful tool for IoT Edge analytics. Its unified messaging model improves the performance of microservices architecture, and its tiered storage capabilities allow for larger volumes of data to be handled without fear of data loss. Pulsar's flexible API

interface works with Java, C++, Python, and Go, making it easy to incorporate Pulsar into your stack. about the book Apache Pulsar in Action is a hands-on guide to building scalable streaming messaging systems for distributed applications and microservices systems. You'll start with Pulsar's fundamentals, each illustrated by real-world examples, as you get to grips with Pulsar's unique architecture. Pulsar contributor David Kjerrumgaard teaches the skills you need to deploy a Pulsar server, ingest data from third-party systems, and deploy lightweight computing logic with simple functions. You'll learn to employ Pulsar's

seamless scalability through relatable case studies, including an IOT analytics application that can be deployed within a resource constrained environment and a microservices application based on Pulsar functions. At the end of this practical book, you'll be ready to fully take advantage of Pulsar to create high-traffic message-driven applications. what's inside Publish from Apache Pulsar into third-party data repositories and platforms Design and develop Apache Pulsar functions Perform interactive SQL queries against data stored in Apache Pulsar Examples of Pulsar-based microservices that you can

download and try yourself about the reader Written for experienced Java developers. No prior knowledge of Pulsar is needed. about the author David Kjerrumgaard is the Director of Solution Architecture at Streamlio, and a contributor to the Apache Pulsar and Apache NiFi projects.

**Apache** Aug 21 2023 Describes the history of the Web server platform and covers downloading and compiling, configuring and running the program on UNIX, writing specialized modules, and establishing security routines.

**Apache HTTP Server** Apr 24 2021 How do the Apache HTTP Server results compare with the performance of your

competitors and other organizations with similar offerings? Can Management personnel recognize the monetary benefit of Apache HTTP Server? Where do ideas that reach policy makers and planners as proposals for Apache HTTP Server strengthening and reform actually originate? How do we go about Securing Apache HTTP Server? Think about the people you identified for your Apache HTTP Server project and the project responsibilities you would assign to them. what kind of training do you think they would need to perform these responsibilities effectively? Defining, designing, creating, and

implementing a process to solve a business challenge or meet a business objective is the most valuable role... In EVERY company, organization and department. Unless you are talking a one-time, single-use project within a business, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment

empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc... - they are the people who rule the future. They are the person who asks the right questions to make Apache HTTP Server investments work better. This Apache HTTP Server All-Inclusive Self-Assessment enables You to be that person. All the tools you need to an in-depth Apache HTTP Server Self-Assessment. Featuring 700 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Apache HTTP Server improvements

can be made. In using the questions you will be better able to: - diagnose Apache HTTP Server projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent advances in Apache HTTP Server and process design

strategies into practice according to best practice guidelines Using a Self-Assessment tool known as the Apache HTTP Server Scorecard, you will develop a clear picture of which Apache HTTP Server areas need attention. Your purchase includes access details to the Apache HTTP Server self-assessment dashboard download which gives you your

dynamically prioritized projects-ready tool and shows your organization exactly what to do next. Your exclusive instant access details can be found in your book.

*HTTP: The Definitive Guide*  
Dec 13 2022 Covers topics including HTTP methods and status codes, optimizing proxies, designing web crawlers, content negotiation, and load-balancing strategies.