Bookmark File PDF One Level Training On Hands Cassandra Apache

Right here, we have countless books **One Level Training On Hands Cassandra Apache** and collections to check out. We additionally pay for variant types and as well as type of the books to browse. The suitable book, fiction, history, novel, scientific research, as with ease as various extra sorts of books are readily comprehensible here.

As this One Level Training On Hands Cassandra Apache, it ends taking place monster one of the favored book One Level Training On Hands Cassandra Apache collections that we have. This is why you remain in the best website to see the incredible ebook to have.

KEY=CASSANDRA - GRETCHEN CABRERA

APACHE CASSANDRA HANDS-ON TRAINING LEVEL ONE

CreateSpace Written in a clear, step-by-step manner, this 400-page course provides an excellent starting point for people that want to get into Apache Cassandra and learn best by doing. A high-quality, project-based, hands-on training courseware book, Apache Cassandra Hands-On Training Level One is designed to be used as the student book for a 2-day introductory level Cassandra course delivered by a Cassandra instructor. Having said that, this book can also be done as a self-paced training course. Recommended prerequisites for this training book are experience with databases, SQL, and programming. This hands-on training book takes people through the basics of working with Cassandra as they learn how to install Cassandra, create a database, create tables, insert, update, and delete data, and create an application, as well as create and modify a multiple-node cluster. Unit 1: Understanding What Cassandra is For Unit 2: Getting Started with the Architecture Unit 3: Installing Cassandra Unit 4: Communicating with Cassandra Unit 5: Creating a Database Unit 6: Creating a Table Unit 7: Inserting Data Unit 8: Modeling Data Unit 9: Creating an Application Unit 10: Updating and Deleting Data Unit 11: Selecting Hardware Unit 12: Adding Nodes to a Cluster Unit 13: Repairing Nodes Unit 14: Removing a Node Unit 15: Monitoring a Cluster Unit 16: Adding a Data Center As virtual machine images are used extensively throughout this hands-on course, including for the creation of a multiple-node Cassandra cluster, any computer used for the exercises in this course needs to be relatively high spec. Specifically, a computer with the following is needed: 64-bit operating system (Mac, Windows, or Linux) 8GB (or more) of RAM 30GB (or more) of free hard drive space Latest version of VMware Player installed and working A way to unzip files Acrobat Reader (or equivalent, for viewing a PDF file) For the full outline, and class files download, see ruthstryker.com/books/achotl1. For a sample unit, see ruthstryker.com/books/achotl1/achotl1_ch06_20140717.pdf (Unit 6) or ruthstryker.com/books/achotl1/achotl1_ch15_20140717.pdf (Unit 15). For the setup steps, see ruthstryker.com/books/achotl1/achotl1_apC_20140722.pdf (Appendix C). Student comments about the book: "Excellent starter course that has taken me from knowing nothing of Cassandra to feeling confident in setting up and using it." "Level covered in book is just right." "Course material was good. It had a wide range of labs and was very helpful in understanding the agenda." "Course material was well-written and easy to follow." "Excellent introduction into Cassandra filled with hands-on exercises for all topics." "Material covers the basics quite well."

CASSANDRA: THE DEFINITIVE GUIDE

DISTRIBUTED DATA AT WEB SCALE

"*O'Reilly Media, Inc.*" 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

PRACTICAL CASSANDRA

A DEVELOPER'S APPROACH

Pearson Education "Eric and Russell were early adopters of Cassandra at SimpleReach. In Practical Cassandra, you benefit from their experience in the trenches administering Cassandra, developing against it, and building one of the first CQL drivers. If you are deploying Cassandra soon, or you inherited a Cassandra cluster to tend, spend some time with the deployment, performance tuning, and maintenance chapters... If you are new to Cassandra, I highly recommend the chapters on data modeling and CQL." –From the Foreword by Jonathon Ellis, Apache Cassandra Chair Build and Deploy Massively Scalable, Super-fast Data Management Applications with Apache Cassandra Practical Cassandra is the first hands-on developer's guide to building Cassandra systems and applications that deliver breakthrough speed, scalability, reliability, and performance. Fully up to date, it reflects the latest versions of Cassandra–including Cassandra Query Language (CQL), which dramatically lowers the learning curve for Cassandra developers.

Pioneering Cassandra developers and Datastax MVPs Russell Bradberry and Eric Lubow walk you through every step of building a real production application that can store enormous amounts of structured, semi-structured, and unstructured data. Drawing on their exceptional expertise, Bradberry and Lubow share practical insights into issues ranging from querying to deployment, management, maintenance, monitoring, and troubleshooting. The authors cover key issues, from architecture to migration, and guide you through crucial decisions about configuration and data modeling. They provide tested sample code, detailed explanations of how Cassandra works "under the covers," and new case studies from three cutting-edge users: Ooyala, Hailo, and eBay. Coverage includes Understanding Cassandra's approach, architecture, key concepts, and primary use cases– and why it's so blazingly fast Getting Cassandra up and running on single nodes and large clusters Applying the new design patterns, philosophies, and features that make Cassandra such a powerful data store Leveraging CQL to simplify your transition from SQL-based RDBMSes Deploying and provisioning through the cloud or on bare-metal hardware Choosing the right configuration options for each type of workload Tweaking Cassandra to get maximum performance from your hardware, OS, and JVM Mastering Cassandra's essential tools for maintenance and monitoring Efficiently solving the most common problems with Cassandra deployment, operation, and application development

MANAGEMENT TRAINING PROGRAMS IN HIGHER EDUCATION FOR THE FOURTH INDUSTRIAL REVOLUTION: EMERGING RESEARCH AND OPPORTUNITIES

EMERGING RESEARCH AND OPPORTUNITIES

IGI Global Under the Fourth Industrial Revolution, it is important that organizations recruit and retain managers with skills that allow them to focus on meeting strategic objectives. To achieve this, companies must focus on implementing strategic management that allows managers to manage and lead continuous change, creativity, innovation, learning, productivity, speed, and effectiveness of their activities that turn them into constant value generators. Managers must also possess personal competences such as initiative, resilience, assumption of risk, creativity, networking, empathy, negotiation, self-control, self-knowledge, and the ability to turn difficulties into opportunities. These capabilities are important in the context of the Fourth Industrial Revolution because they will allow organizations to cope with the highly changing environment that will enable them to consolidate their growth and profitability. Management Training Programs in Higher Education for the Fourth Industrial Revolution: Emerging Research and Opportunities delivers emerging research investigating empirical studies on the formation of management competences in higher education in the context of the Fourth Industrial Revolution in regards to its development and linkage with the business sector in order to offer educational strategies at the national and international level. Featuring coverage on a broad range of topics such as sustainable development, circular economics, and big data analysis, this book is ideally designed for academicians, educators, executives, managers, entrepreneurs, organizational development specialists, consultants, policymakers, researchers, administration professionals, and high-level students.

APACHE SPARK 2: DATA PROCESSING AND REAL-TIME ANALYTICS

MASTER COMPLEX BIG DATA PROCESSING, STREAM ANALYTICS, AND MACHINE LEARNING WITH APACHE SPARK

Packt Publishing Ltd Build efficient data flow and machine learning programs with this flexible, multi-functional open-source cluster-computing framework Key Features Master the art of real-time big data processing and machine learning Explore a wide range of use-cases to analyze large data Discover ways to optimize your work by using many features of Spark 2.x and Scala Book Description Apache Spark is an in-memory, cluster-based data processing system that provides a wide range of functionalities such as big data processing, analytics, machine learning, and more. With this Learning Path, you can take your knowledge of Apache Spark to the next level by learning how to expand Spark's functionality and building your own data flow and machine learning programs on this platform. You will work with the different modules in Apache Spark, such as interactive querying with Spark SQL, using DataFrames and datasets, implementing streaming analytics with Spark Streaming, and applying machine learning and deep learning techniques on Spark using MLlib and various external tools. By the end of this elaborately designed Learning Path, you will have all the knowledge you need to master Apache Spark, and build your own big data processing and analytics pipeline quickly and without any hassle. This Learning Path includes content from the following Packt products: Mastering Apache Spark 2.x by Romeo Kienzler Scala and Spark for Big Data Analytics by Md. Rezaul Karim, Sridhar Alla Apache Spark 2.x Machine Learning Cookbook by Siamak Amirghodsi, Meenakshi Rajendran, Broderick Hall, Shuen MeiCookbook What you will learn Get to grips with all the features of Apache Spark 2.x Perform highly optimized real-time big data processing Use ML and DL techniques with Spark MLlib and third-party tools Analyze structured and unstructured data using SparkSQL and GraphX Understand tuning, debugging, and monitoring of big data applications Build scalable and fault-tolerant streaming applications Develop scalable recommendation engines Who this book is for If you are an intermediate-level Spark developer looking to master the advanced capabilities and use-cases of Apache Spark 2.x, this Learning Path is ideal for you. Big data professionals who want to learn how to integrate and use the features of Apache Spark and build a strong big data pipeline will also find this Learning Path useful. To grasp the concepts explained in this Learning Path, you must know the fundamentals of Apache Spark and Scala.

TRINO: THE DEFINITIVE GUIDE

"O'Reilly Media, Inc." Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by Facebook, open source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn

about tools that will help you connect to Trino and query data Go deeper: Learn Trino's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more Put Trino in production: Secure Trino, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino

HANDS-ON DEEP LEARNING WITH APACHE SPARK

BUILD AND DEPLOY DISTRIBUTED DEEP LEARNING APPLICATIONS ON APACHE SPARK

Packt Publishing Ltd Speed up the design and implementation of deep learning solutions using Apache Spark Key FeaturesExplore the world of distributed deep learning with Apache SparkTrain neural networks with deep learning libraries such as BigDL and TensorFlowDevelop Spark deep learning applications to intelligently handle large and complex datasetsBook Description Deep learning is a subset of machine learning where datasets with several layers of complexity can be processed. Hands-On Deep Learning with Apache Spark addresses the sheer complexity of technical and analytical parts and the speed at which deep learning solutions can be implemented on Apache Spark. The book starts with the fundamentals of Apache Spark and deep learning. You will set up Spark for deep learning, learn principles of distributed modeling, and understand different types of neural nets. You will then implement deep learning models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) on Spark. As you progress through the book, you will gain hands-on experience of what it takes to understand the complex datasets you are dealing with. During the course of this book, you will use popular deep learning frameworks, such as TensorFlow, Deeplearning4j, and Keras to train your distributed models. By the end of this book, you'll have gained experience with the implementation of your models on a variety of use cases. What you will learnUnderstand the basics of deep learningSet up Apache Spark for deep learningUnderstand the principles of distribution modeling and different types of neural networksObtain an understanding of deep learning algorithmsDiscover textual analysis and deep learning with SparkUse popular deep learning frameworks, such as Deeplearning4j, TensorFlow, and KerasExplore popular deep learning algorithms Who this book is for If you are a Scala developer, data scientist, or data analyst who wants to learn how to use Spark for implementing efficient deep learning models, Hands-On Deep Learning with Apache Spark is for you. Knowledge of the core machine learning concepts and some exposure to Spark will be helpful.

PROGRAMMING HIVE

"*O'Reilly Media, Inc.*" Describes the features and functions of Apache Hive, the data infrastructure for Hadoop.

LEARNING SPARK

O'Reilly Media Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step 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

REWRITING LOGIC AND ITS APPLICATIONS

14TH INTERNATIONAL WORKSHOP, WRLA 2022, MUNICH, GERMANY, APRIL 2–3, 2022, REVISED SELECTED PAPERS

Springer Nature This book constitutes selected papers from the refereed proceedings of the 14th International Workshop on Rewriting Logic and Its Applications, WRLA 2022, held in Munich, Germany, in April 2022. The 9 full papers included in this book were carefully reviewed and selected from 13 submissions. They focus on topics in rewriting logic and its applications. The book also contains 2 invited papers, 2 invited tutorials and an experience report.

HANDS-ON DATA SCIENCE AND PYTHON MACHINE LEARNING

Packt Publishing Ltd This book covers the fundamentals of machine learning with Python in a concise and dynamic manner. It covers data mining and large-scale machine learning using Apache Spark. About This Book Take your first steps in the world of data science by understanding the tools and techniques of data analysis Train efficient Machine Learning models in Python using the supervised and unsupervised learning methods Learn how to use Apache Spark for processing Big Data efficiently Who This Book Is For If you are a budding data scientist or a data analyst who wants to analyze and gain actionable insights from data using Python, this book is for you. Programmers with some experience in Python who want to enter the lucrative world of Data Science will also find this book to be very useful, but you don't need to be an expert Python coder or mathematician to get the most from this book. What You Will Learn Learn how to clean your data and ready it for analysis Implement the popular clustering and regression methods in Python Train efficient machine learning models using decision trees and random forests Visualize the results of your analysis using Python's Matplotlib library Use Apache Spark's MLlib package to perform machine learning on large datasets In Detail Join Frank Kane, who worked on Amazon and IMDb's machine learning algorithms, as he guides you on your first steps into the world of data science. Hands-On Data Science and Python Machine Learning gives you the tools that you need to understand and explore the core topics in the field, and the confidence and practice to build and analyze your own machine learning models. With the help of interesting and easy-to-follow practical examples, Frank Kane explains potentially complex topics such as Bayesian methods and K-means clustering

in a way that anybody can understand them. Based on Frank's successful data science course, Hands-On Data Science and Python Machine Learning empowers you to conduct data analysis and perform efficient machine learning using Python. Let Frank help you unearth the value in your data using the various data mining and data analysis techniques available in Python, and to develop efficient predictive models to predict future results. You will also learn how to perform large-scale machine learning on Big Data using Apache Spark. The book covers preparing your data for analysis, training machine learning models, and visualizing the final data analysis. Style and approach This comprehensive book is a perfect blend of theory and hands-on code examples in Python which can be used for your reference at any time.

LEARNING APACHE CASSANDRA

Packt Publishing Ltd If you're an application developer familiar with SQL databases such as MySQL or Postgres, and you want to explore distributed databases such as Cassandra, this is the perfect guide for you. Even if you've never worked with a distributed database before, Cassandra's intuitive programming interface coupled with the step-by-step examples in this book will have you building highly scalable persistence layers for your applications in no time.

PROFESSIONAL NOSQL

John Wiley & Sons A hands-on guide to leveraging NoSQL databases NoSQL databases are an efficient and powerful tool for storing and manipulating vast quantities of data. Most NoSQL databases scale well as data grows. In addition, they are often malleable and flexible enough to accommodate semi-structured and sparse data sets. This comprehensive hands-on guide presents fundamental concepts and practical solutions for getting you ready to use NoSQL databases. Expert author Shashank Tiwari begins with a helpful introduction on the subject of NoSQL, explains its characteristics and typical uses, and looks at where it fits in the application stack. Unique insights help you choose which NoSQL solutions are best for solving your specific data storage needs. Professional NoSQL: Demystifies the concepts that relate to NoSQL databases, including column-family oriented stores, key/value databases, and document databases. Delves into installing and configuring a number of NoSQL products and the Hadoop family of products. Explains ways of storing, accessing, and querying data in NoSQL databases through examples that use MongoDB, HBase, Cassandra, Redis, CouchDB, Google App Engine Datastore and more. Looks at architecture and internals. Provides guidelines for optimal usage, performance tuning, and scalable configurations. Presents a number of tools and utilities relating to NoSQL, distributed platforms, and scalable processing, including Hive, Pig, RRDtool, Nagios, and more.

KAFKA: THE DEFINITIVE GUIDE

REAL-TIME DATA AND STREAM PROCESSING AT SCALE

"*O'Reilly Media, Inc.*" 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

FAST DATA PROCESSING SYSTEMS WITH SMACK STACK

Packt Publishing Ltd Combine the incredible powers of Spark, Mesos, Akka, Cassandra, and Kafka to build data processing platforms that can take on even the hardest of your data troubles! About This Book This highly practical guide shows you how to use the best of the big data technologies to solve your response-critical problems Learn the art of making cheap-yet-effective big data architecture without using complex Greek-letter architectures Use this easy-to-follow guide to build fast data processing systems for your organization Who This Book Is For If you are a developer, data architect, or a data scientist looking for information on how to integrate the Big Data stack architecture and how to choose the correct technology in every layer, this book is what you are looking for. What You Will Learn Design and implement a fast data Pipeline architecture Think and solve programming challenges in a functional way with Scala Learn to use Akka, the actors model implementation for the JVM Make on memory processing and data analysis with Spark to solve modern business demands Build a powerful and effective cluster infrastructure with Mesos and Docker Manage and consume unstructured and No-SQL data sources with Cassandra Consume and produce messages in a massive way with Kafka In Detail SMACK is an open source full stack for big data architecture. It is a combination of Spark, Mesos, Akka, Cassandra, and Kafka. This stack is the newest technique developers have begun to use to tackle critical real-time analytics for big data. This highly practical guide will teach you how to integrate these technologies to create a highly efficient data analysis system for fast data processing. We'll start off with an introduction to SMACK and show you when to use it. First you'll get to grips with functional thinking and problem solving using Scala. Next you'll come to understand the Akka architecture. Then you'll get to know how to improve the data structure architecture and optimize resources using Apache Spark. Moving forward, you'll learn how to perform linear scalability in databases with Apache Cassandra. You'll grasp the high throughput distributed messaging systems using Apache Kafka. We'll show you how to build a cheap but effective cluster infrastructure with Apache Mesos. Finally, you will deep dive into the different aspect of SMACK using a few case studies. By the end of the book, you will be able to integrate all the components of the SMACK stack and use them together to achieve

highly effective and fast data processing. Style and approach With the help of various industry examples, you will learn about the full stack of big data architecture, taking the important aspects in every technology. You will learn how to integrate the technologies to build effective systems rather than getting incomplete information on single technologies. You will learn how various open source technologies can be used to build cheap and fast data processing systems with the help of various industry examples

LEARNING SPARK

LIGHTNING-FAST BIG DATA ANALYSIS

"*O'Reilly Media, Inc.*" 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 including HDFS, Hive, JSON, and S3 Master advanced topics like data partitioning and shared variables

BIG DATA SMACK

A GUIDE TO APACHE SPARK, MESOS, AKKA, CASSANDRA, AND KAFKA

Apress 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

ADVANCES IN SPATIAL AND TEMPORAL DATABASES

12TH INTERNATIONAL SYMPOSIUM, SSTD 2011, MINNEAPOLIS, MN, USA, AUGUST 24-26, 2011. PROCEEDINGS

Springer Science & Business Media This volume constitutes the refereed proceedings of the 12th International Symposium on Spatial and Temporal Databases, SSTD 2011, held in Minneapolis, USA, in August 2011. The 24 revised full papers presented together with one keynote, 8 short papers, and 8 demonstration papers, were thoroughly reviewed and selected from a total of 63 research submissions, 21 vision and challenges submissions and 16 demonstration submissions. The papers are organized in topical sections on knowledge discovery; spatial networks; access methods; moving objects and sensor networks; multidimensional query processing; and temporal and streaming data.

ARMY

EXPERT APACHE CASSANDRA ADMINISTRATION

Apress 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.

SPARK: THE DEFINITIVE GUIDE

BIG DATA PROCESSING MADE SIMPLE

"O'Reilly Media, Inc." 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

BIG DATA ANALYTICS WITH JAVA

Packt Publishing Ltd Learn the basics of analytics on big data using Java, machine learning and other big data tools About This Book Acquire real-world set of tools for building enterprise level data science applications Surpasses the barrier of other languages in data science and learn create useful object-oriented codes Extensive use of Java compliant big data tools like apache spark, Hadoop, etc. Who This Book Is For This book is for Java developers who are looking to perform data analysis in production environment. Those who wish to implement data analysis in their Big data applications will find this book helpful. What You Will Learn Start from simple analytic tasks on big data Get into more complex tasks with predictive analytics on big data using machine learning Learn real time analytic tasks Understand the concepts with examples and case studies Prepare and refine data for analysis Create charts in order to understand the data See various real-world datasets In Detail This book covers case studies such as sentiment analysis on a tweet dataset, recommendations on a movielens dataset, customer segmentation on an ecommerce dataset, and graph analysis on actual flights dataset. This book is an end-to-end guide to implement analytics on big data with Java. Java is the de facto language for major big data environments, including Hadoop. This book will teach you how to perform analytics on big data with production-friendly Java. This book basically divided into two sections. The first part is an introduction that will help the readers get acquainted with big data environments, whereas the second part will contain a hardcore discussion on all the concepts in analytics on big data. It will take you from data analysis and data visualization to the core concepts and advantages of machine learning, real-life usage of regression and classification using Naive Bayes, a deep discussion on the concepts of clustering,and a review of simple neural networks on big data using deepLearning4j or plain Java Spark code. This book is a must-have book for Java developers who want to start learning big data analytics and want to use it in the real world. Style and approach The approach of book is to deliver practical learning modules in manageable content. Each chapter is a self-contained unit of a concept in big data analytics. Book will step by step builds the competency in the area of big data analytics. Examples using real world case studies to give ideas of real applications and how to use the techniques mentioned. The examples and case studies will be shown using both theory and code.

DATABASE INTERNALS

A DEEP DIVE INTO HOW DISTRIBUTED DATA SYSTEMS WORK

O'Reilly Media When it comes to choosing, using, and maintaining a database, understanding its internals is essential. But with so many distributed databases and tools available today, it's often difficult to understand what each one offers and how they differ. With this practical guide, Alex Petrov guides developers through the concepts behind modern database and storage engine internals. Throughout the book, you'll explore relevant material gleaned from numerous books, papers, blog posts, and the source code of several open source databases. These resources are listed at the end of parts one and two. You'll discover that the most significant distinctions among many modern databases reside in subsystems that determine how storage is organized and how data is distributed. This book examines: Storage engines: Explore storage classification and taxonomy, and dive into B-Tree-based and immutable Log Structured storage engines, with differences and use-cases for each Storage building blocks: Learn how database files are organized to build efficient storage, using auxiliary data structures such as Page Cache, Buffer Pool and Write-Ahead Log Distributed systems: Learn step-by-step how nodes and processes connect and build complex communication patterns Database clusters: Which consistency models are commonly used by modern databases and how distributed storage systems achieve consistency

GETTING STARTED WITH KUDU

PERFORM FAST ANALYTICS ON FAST DATA

"O'Reilly Media, Inc." Fast data ingestion, serving, and analytics in the Hadoop ecosystem have forced developers and architects to choose solutions using the least common denominator—either fast analytics at the cost of slow data ingestion or fast data ingestion at the cost of slow analytics. There is an answer to this problem. With the Apache Kudu column-oriented data store, you can easily

perform fast analytics on fast data. This practical guide shows you how. Begun as an internal project at Cloudera, Kudu is an open source solution compatible with many data processing frameworks in the Hadoop environment. In this book, current and former solutions professionals from Cloudera provide use cases, examples, best practices, and sample code to help you get up to speed with Kudu. Explore Kudu's high-level design, including how it spreads data across servers Fully administer a Kudu cluster, enable security, and add or remove nodes Learn Kudu's client-side APIs, including how to integrate Apache Impala, Spark, and other frameworks for data manipulation Examine Kudu's schema design, including basic concepts and primitives necessary to make your project successful Explore case studies for using Kudu for real-time IoT analytics, predictive modeling, and in combination with another storage engine

DESIGN PATTERNS FOR CLOUD NATIVE APPLICATIONS

"*O'Reilly Media, Inc.*" With the immense cost savings and scalability the cloud provides, the rationale for building cloud native applications is no longer in question. The real issue is how. With this practical guide, developers will learn about the most commonly used design patterns for building cloud native applications using APIs, data, events, and streams in both greenfield and brownfield development. You'll learn how to incrementally design, develop, and deploy large and effective cloud native applications that you can manage and maintain at scale with minimal cost, time, and effort. Authors Kasun Indrasiri and Sriskandarajah Suhothayan highlight use cases that effectively demonstrate the challenges you might encounter at each step. Learn the fundamentals of cloud native applications Explore key cloud native communication, connectivity, and composition patterns Learn decentralized data management techniques Use event-driven architecture to build distributed and scalable cloud native applications Explore the most commonly used patterns for API management and consumption Examine some of the tools and technologies you'll need for building cloud native systems

BIG DATA ARCHITECT'S HANDBOOK

A GUIDE TO BUILDING PROFICIENCY IN TOOLS AND SYSTEMS USED BY LEADING BIG DATA EXPERTS

Packt Publishing Ltd A comprehensive end-to-end guide that gives hands-on practice in big data and Artificial Intelligence Key Features Learn to build and run a big data application with sample code Explore examples to implement activities that a big data architect performs Use Machine Learning and AI for structured and unstructured data Book Description The big data architects are the "masters" of data, and hold high value in today's market. Handling big data, be it of good or bad quality, is not an easy task. The prime job for any big data architect is to build an end-to-end big data solution that integrates data from different sources and analyzes it to find useful, hidden insights. Big Data Architect's Handbook takes you through developing a complete, end-to-end big data pipeline, which will lay the foundation for you and provide the necessary knowledge required to be an architect in big data. Right from understanding the design considerations to implementing a solid, efficient, and scalable data pipeline, this book walks you through all the essential aspects of big data. It also gives you an overview of how you can leverage the power of various big data tools such as Apache Hadoop and ElasticSearch in order to bring them together and build an efficient big data solution. By the end of this book, you will be able to build your own design system which integrates, maintains, visualizes, and monitors your data. In addition, you will have a smooth design flow in each process, putting insights in action. What you will learn Learn Hadoop Ecosystem and Apache projects Understand, compare NoSQL database and essential software architecture Cloud infrastructure design considerations for big data Explore application scenario of big data tools for daily activities Learn to analyze and visualize results to uncover valuable insights Build and run a big data application with sample code from end to end Apply Machine Learning and AI to perform big data intelligence Practice the daily activities performed by big data architects Who this book is for Big Data Architect's Handbook is for you if you are an aspiring data professional, developer, or IT enthusiast who aims to be an all-round architect in big data. This book is your one-stop solution to enhance your knowledge and carry out easy to complex activities required to become a big data architect.

MACHINE LEARNING

HANDS-ON FOR DEVELOPERS AND TECHNICAL PROFESSIONALS

John Wiley & Sons Dig deep into the data with a hands-on guide to machine learning with updated examples and more! Machine Learning: Hands-On for Developers and Technical Professionals provides hands-on instruction and fully-coded working examples for the most common machine learning techniques used by developers and technical professionals. The book contains a breakdown of each ML variant, explaining how it works and how it is used within certain industries, allowing readers to incorporate the presented techniques into their own work as they follow along. A core tenant of machine learning is a strong focus on data preparation, and a full exploration of the various types of learning algorithms illustrates how the proper tools can help any developer extract information and insights from existing data. The book includes a full complement of Instructor's Materials to facilitate use in the classroom, making this resource useful for students and as a professional reference. At its core, machine learning is a mathematical, algorithm-based technology that forms the basis of historical data mining and modern big data science. Scientific analysis of big data requires a working knowledge of machine learning, which forms predictions based on known properties learned from training data. Machine Learning is an accessible, comprehensive guide for the non-mathematician, providing clear guidance that allows readers to: Learn the languages of machine learning including Hadoop, Mahout, and Weka Understand decision trees, Bayesian networks, and artificial neural networks Implement Association Rule, Real Time, and Batch learning Develop a strategic plan for safe, effective, and efficient machine learning By learning to construct a system that can learn from data, readers can increase their utility across industries. Machine learning sits at the core of deep dive data analysis and visualization, which is increasingly in demand as companies discover the goldmine hiding in their existing data. For the tech professional involved in data science, Machine Learning: Hands-On for Developers and Technical Professionals provides the skills and techniques required to dig deeper.

PURE-BRED DOGS, AMERICAN KENNEL GAZETTE

CASSANDRA ADMIN DBA GUIDE

Independently Published What value this book provides? This book absolutely provides tremendous value in terms its usefulness. This book takes away the pain associated with learning and mastering Cassandra. All complexity has been digested by the author and simplified for the reader with very useful and practical information that can be absorbed just by glancing through the pages. Years of author's experience and practical knowledge in Cassandra has been gifted to the reader in this book with great diligence and generosity. If you are planning to undergo expensive Cassandra training, think again, spending few hours with this book will change your mind, this book has been written with great care to reduce the learning curve. The aim of this book is multi fold, just to mention a few: Provide confidence to the reader in Cassandra concepts and architecture Provide a flexible, practical framework and context customizable for various situations Provide practical guidance to manage Cassandra platforms of various hues, sizes, shapes Provide real world examples to reduce guess work Provide executable query statements and command line statements at every step Provide practical outcomes to help the reader to gain instance understanding of what to expect Provide insights into making Cassandra environment robust and scalable Provide tricks and tips to implement and maintain seamlessly Provide security and vulnerability mitigation tips and steps Provide best practices to follow for optimal Cassandra use There is no doubt - this book makes the reader very productive Cassandra professional in very short span of time. This book essentially bridges the training gap as the industry is moving fast to take full advantage of what Cassandra can offer to fulfill emerging business needs. This book can be very helpful to Data administrators, Developers, Data modelers/Application Architects, Project Managers and Data Custodians. This book has range of topics, many are listed here: Cassandra concepts and architecture Cassandra Installation and Configuration Cassandra system architecture depicting gossip protocol, replication, consistency, tombstones, hinted handoff, compactions, repairs, memtables, commit log, read and write path functions Cassandra oriented data modelling Cassandra QL (CQL) tutorial Handling of Primary and Partition keys in Cassandra covering No joins, Static columns and TTL aspects Configuring authentication, authorization to access Cassandra in addition, steps to set up node-node and client-node SSL Configuring nodes addition, removal, decommission in single token and vnode setup modes in Cassandra Instructions to add new data center and delink the existing data center from a multi-dc cluster arrangement Cassandra backup and recovery functions with real examples of restoring tables after truncation events Cassandra utilities such as sstabledump, sstablemetadata, sstablesplit, cqlsh and cassandra-stress Troubleshooting methods such as Node down, Read latency and Recover truncated table Upgrading Cassandra to higher versions Additional Cassandra architecture II methods such as Read and Write path, Compactions and Repairs

BEGINNING APACHE CASSANDRA DEVELOPMENT

Apress Beginning Apache Cassandra Development introduces you to one of the most robust and best-performing NoSQL database platforms on the planet. Apache Cassandra is a document database following the JSON document model. It is specifically designed to manage large amounts of data across many commodity servers without there being any single point of failure. This design approach makes Apache Cassandra a robust and easy-to-implement platform when high availability is needed. Apache Cassandra can be used by developers in Java, PHP, Python, and JavaScript—the primary and most commonly used languages. In Beginning Apache Cassandra Development, author and Cassandra expert Vivek Mishra takes you through using Apache Cassandra from each of these primary languages. Mishra also covers the Cassandra Query Language (CQL), the Apache Cassandra analog to SQL. You'll learn to develop applications sourcing data from Cassandra, query that data, and deliver it at speed to your application's users. Cassandra is one of the leading NoSQL databases, meaning you get unparalleled throughput and performance without the sort of processing overhead that comes with traditional proprietary databases. Beginning Apache Cassandra Development will therefore help you create applications that generate search results quickly, stand up to high levels of demand, scale as your user base grows, ensure operational simplicity, and—not least—provide delightful user experiences.

CASSANDRA DESIGN PATTERNS

Packt Publishing Ltd Build real-world, industry-strength data storage solutions with time-tested design methodologies using Cassandra About This Book Explore design patterns which co-exist with legacy data stores, migration from RDBMS, and caching technologies with Cassandra Learn about design patterns and use Cassandra to provide consistency, availability, and partition tolerance guarantees for applications Handle temporal data for analytical purposes Who This Book Is For This book is intended for big data developers who are familiar with the basics of Cassandra and wish to understand and utilize Cassandra design patterns to develop real-world big data solutions. Prior knowledge of RDBMS solutions is assumed. What You Will Learn Enable Cassandra to co-exist with RDBMS and other legacy data stores Explore various design patterns to build effective and robust storage solutions Migrate from RDBMS-based data stores and caching solutions to Cassandra Understand the behaviour of Cassandra when trying to balance the needs of consistency, availability, and partition tolerance Deal with time stamps related to data effectively See how Cassandra can be used in analytical use cases Apply the design patterns covered in this book in real-world use cases In Detail There are many NoSQL data stores used by big data applications. Cassandra is one of the most widely used NoSQL data stores that is frequently used by a huge number of heavy duty Internet-scale applications. Unlike the RDBMS world, the NoSQL landscape is very diverse and there is no one way to model data stores. This mandates the need to have good solutions to commonly seen data store design problems. Cassandra addresses such common problems simply. If you are new to Cassandra but well-versed in RDBMS modeling and design, then it is natural to model data in the same way in Cassandra, resulting in poorly performing applications and losing the real purpose of Cassandra. If you want to learn to make the most of Cassandra, this book is for you. This book starts with strategies to integrate Cassandra with other legacy data stores and progresses to the ways in which a migration from RDBMS to Cassandra can be accomplished. The journey continues with ideas to migrate data from cache solutions to Cassandra. With this, the stage is set and the book moves on to some of the most commonly seen problems in applications when dealing with consistency, availability, and partition tolerance guarantees. Cassandra is exceptionally good at dealing with temporal data and patterns such as the time-series pattern and log pattern, which are covered next. Many NoSQL data stores fail miserably when a huge amount of data is read for analytical purposes, but Cassandra is different in this regard. Keeping analytical needs in mind, you'll walk through different and interesting design patterns. No theoretical discussions

are complete without a good set of use cases to which the knowledge gained can be applied, so the book concludes with a set of use cases you can apply the patterns you've learned. Style and approach This book is written in very simple language and an engaging style complete with examples in every chapter and real-world use cases at the end of the book.

NOSQL DISTILLED

A BRIEF GUIDE TO THE EMERGING WORLD OF POLYGLOT PERSISTENCE

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

BEGINNING DATABASE DESIGN SOLUTIONS

John Wiley & Sons The vast majority of software applications use relational databases that virtually every application developer must work with. This book introduces you to database design, whether you're a DBA or database developer. You'll discover what databases are, their goals, and why proper design is necessary to achieve those goals. Additionally, you'll master how to structure the database so it gives good performance while minimizing the chance for error. You will learn how to decide what should be in a database to meet the application's requirements.

COMPUTERWORLD

For more than 40 years, Computerworld has been the leading source of technology news and information for IT influencers worldwide. Computerworld's award-winning Web site (Computerworld.com), twice-monthly publication, focused conference series and custom research form the hub of the world's largest global IT media network.

APACHE SQOOP COOKBOOK

UNLOCKING HADOOP FOR YOUR RELATIONAL DATABASE

"*O'Reilly Media, Inc.*" Integrating data from multiple sources is essential in the age of big data, but it can be a challenging and time-consuming task. This handy cookbook provides dozens of ready-to-use recipes for using Apache Sqoop, the command-line interface application that optimizes data transfers between relational databases and Hadoop. Sqoop is both powerful and bewildering, but with this cookbook's problem-solution-discussion format, you'll quickly learn how to deploy and then apply Sqoop in your environment. The authors provide MySQL, Oracle, and PostgreSQL database examples on GitHub that you can easily adapt for SQL Server, Netezza, Teradata, or other relational systems. Transfer data from a single database table into your Hadoop ecosystem Keep table data and Hadoop in sync by importing data incrementally Import data from more than one database table Customize transferred data by calling various database functions Export generated, processed, or backed-up data from Hadoop to your database Run Sqoop within Oozie, Hadoop's specialized workflow scheduler Load data into Hadoop's data warehouse (Hive) or database (HBase) Handle installation, connection, and syntax issues common to specific database vendors

ARROW OF THE HEART

Amber Quill Press Riding across the Arizona desert in 1884 and carrying close to a million dollars worth of twenty dollar gold pieces, Cassidy Howard is transported to the present with an Apache arrow in her back. Sheriff Jesse White Feather, a full-blood Apache, wants to know why his family marking is on the arrow and who in his family wants to harm Cassidy. When someone in the present tries to steal her gold and her life, Cassidy must put her trust in Jesse so that, together, they can discover the culprit's identity. As they do, they also discover the magic that binds their two worlds together. Cassidy finds the true meaning of home, while Jesse finds the roots he's lost. During the investigation, Jesse is drawn to the unusual woman and realizes he is giving away his heart, which is the only gift she wants from him...

AWS CERTIFIED SYSOPS ADMINISTRATOR OFFICIAL STUDY GUIDE

ASSOCIATE EXAM

John Wiley & Sons Comprehensive, interactive exam preparation and so much more The AWS Certified SysOps Administrator Official Study Guide: Associate Exam is a comprehensive exam preparation resource. This book bridges the gap between exam preparation and real-world readiness, covering exam objectives while guiding you through hands-on exercises based on situations you'll likely encounter as an AWS Certified SysOps Administrator. From deployment, management, and operations to migration, data flow, cost control, and beyond, this guide will help you internalize the processes and best practices associated with AWS. The Sybex interactive

online study environment gives you access to invaluable preparation aids, including an assessment test that helps you focus your study on areas most in need of review, and chapter tests to help you gauge your mastery of the material. Electronic flashcards make it easy to study anytime, anywhere, and a bonus practice exam gives you a sneak preview so you know what to expect on exam day. Cloud computing offers businesses a cost-effective, instantly scalable IT infrastructure. The AWS Certified SysOps Administrator - Associate credential shows that you have technical expertise in deployment, management, and operations on AWS. Study exam objectives Gain practical experience with hands-on exercises Apply your skills to real-world scenarios Test your understanding with challenging review questions Earning your AWS Certification is much more than just passing an exam—you must be able to perform the duties expected of an AWS Certified SysOps Administrator in a real-world setting. This book does more than coach you through the test: it trains you in the tools, procedures, and thought processes to get the job done well. If you're serious about validating your expertise and working at a higher level, the AWS Certified SysOps Administrator Official Study Guide: Associate Exam is the resource you've been seeking.

REAL-TIME ANALYTICS

TECHNIQUES TO ANALYZE AND VISUALIZE STREAMING DATA

John Wiley & Sons Construct a robust end-to-end solution for analyzing and visualizing streaming data Real-time analytics is the hottest topic in data analytics today. In Real-Time Analytics: Techniques to Analyze and Visualize Streaming Data, expert Byron Ellis teaches data analysts technologies to build an effective real-time analytics platform. This platform can then be used to make sense of the constantly changing data that is beginning to outpace traditional batch-based analysis platforms. The author is among a very few leading experts in the field. He has a prestigious background in research, development, analytics, real-time visualization, and Big Data streaming and is uniquely qualified to help you explore this revolutionary field. Moving from a description of the overall analytic architecture of real-time analytics to using specific tools to obtain targeted results, Real-Time Analytics leverages open source and modern commercial tools to construct robust, efficient systems that can provide real-time analysis in a cost-effective manner. The book includes: A deep discussion of streaming data systems and architectures Instructions for analyzing, storing, and delivering streaming data Tips on aggregating data and working with sets Information on data warehousing options and techniques Real-Time Analytics includes in-depth case studies for website analytics, Big Data, visualizing streaming and mobile data, and mining and visualizing operational data flows. The book's "recipe" layout lets readers quickly learn and implement different techniques. All of the code examples presented in the book, along with their related data sets, are available on the companion website.

APACHE SPARK IMPLEMENTATION ON IBM Z/OS

IBM Redbooks The term big data refers to extremely large sets of data that are analyzed to reveal insights, such as patterns, trends, and associations. The algorithms that analyze this data to provide these insights must extract value from a wide range of data sources, including business data and live, streaming, social media data. However, the real value of these insights comes from their timeliness. Rapid delivery of insights enables anyone (not only data scientists) to make effective decisions, applying deep intelligence to every enterprise application. Apache Spark is an integrated analytics framework and runtime to accelerate and simplify algorithm development, depoyment, and realization of business insight from analytics. Apache Spark on IBM® z/OS® puts the open source engine, augmented with unique differentiated features, built specifically for data science, where big data resides. This IBM Redbooks® publication describes the installation and configuration of IBM z/OS Platform for Apache Spark for field teams and clients. Additionally, it includes examples of business analytics scenarios.

HANDS-ON INDUSTRIAL INTERNET OF THINGS

CREATE A POWERFUL INDUSTRIAL IOT INFRASTRUCTURE USING INDUSTRY 4.0

Packt Publishing Ltd Build a strong and efficient IoT infrastructure at industrial and enterprise level by mastering Industrial IoT network Key FeaturesGain hands-on experience working with industrial architectureExplore the potential of cloud-based Industrial IoT platforms, analytics, and protocolsImprove business models and transform your workforce with Industry 4.0Book Description We live in an era where advanced automation is used to achieve accurate results. To set up an automation environment, you need to first configure a network that can be accessed anywhere and by any device. This book is a practical guide that helps you discover the technologies and use cases for Industrial Internet of Things (IIOT). Hands-On Industrial Internet of Things takes you through the implementation of industrial processes and specialized control devices and protocols. You'll study the process of identifying and connecting to different industrial data sources gathered from different sensors. Furthermore, you'll be able to connect these sensors to cloud network, such as AWS IoT, Azure IoT, Google IoT, and OEM IoT platforms, and extract data from the cloud to your devices. As you progress through the chapters, you'll gain hands-on experience in using open source Node-Red, Kafka, Cassandra, and Python. You will also learn how to develop streaming and batch-based Machine Learning algorithms. By the end of this book, you will have mastered the features of Industry 4.0 and be able to build stronger, faster, and more reliable IoT infrastructure in your Industry. What you will learnExplore industrial processes, devices, and protocolsDesign and implement the I-IoT network flowGather and transfer industrial data in a secure wayGet to grips with popular cloud-based platformsUnderstand diagnostic analytics to answer critical workforce questionsDiscover the Edge device and understand Edge and Fog computingImplement equipment and process management to achieve business-specific goalsWho this book is for If you're an IoT architect, developer, or stakeholder working with architectural aspects of Industrial Internet of Things, this book is for you.