Learning Real Time Processing With Spark Streaming

Concepts, Methodologies, Tools, and Applications
Learning Storm
with Practical Examples in MOA
Introduction to Apache Flink
Apache Spark 2: Data Processing and Real-Time Analytics
Master complex big data processing, stream analytics, and machine learning with Apache Spark
Big Data Processing with Apache Spark
Digital Transformation Technology
Crosslinguistic Influence and Second Language Learning
Fundamentals, Implementation, and Operation of Streaming Applications
Learning Real Time Processing with Spark Streaming
Frontiers in Massive Data Analysis
Business in Real-Time Using Azure IoT and Cortana Intelligence Suite
Encyclopedia of Ecology
Stream Processing with Apache Spark
Build Real Time Data Analytics on Google Cloud Platform
Machine Learning for Data Streams
Distributed Computing and Event Processing using Apache Spark, Flink, Storm, and Kafka
Apache Spark 2
Fundamentals, Implementations and Applications
Real-Time Digital Signal Processing
Scala for Machine Learning
Explore the concepts of functional programming, data streaming, and machine learning
Stream Processing with Apache Flink
4th International Conference on E-learning, Edutainment 2009, Banff, Canada, August 9-11, 2009, Proceedings
Data processing, ML algorithms, smart analytics, and more
Driving Your Digital Transformation
A Tapestry of Systems and AI-Based Theories and Methodologies
Practical Real-time Data Processing and Analytics
Learning Spark SQL
Kafka Streams - Real-time Stream Processing
Proceedings of ICMLIP 2019
Prediction in Second Language Processing and Learning
Unmanned Aerial Vehicles in Smart Cities
Advanced Analytics and Real-time Data Processing in Apache Spark
Learning Management Back from Machines

Interdisciplinary Approaches to Spatial Optimization Issues
Machine Learning and Information Processing
Deep Learning: Convergence to Big Data Analytics

KAITLYN SULLIVAN

Concepts, Methodologies, Tools, and Applications
BPB Publications
This book addresses the major challenges in realizing unmanned aerial vehicles (UAVs) in IoT-based Smart Cities. The challenges tackled vary from cost and energy efficiency to availability and service quality. The aim of this book is to focus on both the design and implementation aspects of the UAV-based approaches in IoT-enabled smart cities' applications that are enabled and supported by wireless sensor networks, 5G, and beyond. The contributors mainly focus on data delivery approaches and their performability aspects. This book is meant for readers of varying disciplines who are interested in implementing the smart planet/environments vision via wireless/wired enabling technologies. Involves the most up to date unmanned aerial vehicles (UAV) assessment and evaluation approaches Includes innovative operational ideas in agriculture, surveillance, rescue, etc. Pertains researchers, scientists, engineers and practitioners in the field of smart cities, IoT, and communications Fadi Al-Turjman received his Ph.D. from Queen's University, Canada. He is a full professor and a research center director at Near East University, Nicosia. He is a leading authority in the area of IoT and intelligent systems. His publication history spans over 250 publications in addition to his editorialship in top journals such as the IEEE Communication Surveys and Tutorials, and the Elsevier Sustaibable Cities and Society.
Learning Storm "O'Reilly Media, Inc."
As metropolises continue to see a growth in population, planners are continually searching for trending methods for utilizing space and seeking the best geographical arrangements for these cities. Professionals have continually used geographic information systems (GIS) to solve these issues; however, limitations in this technology remain prevalent. Integrating multiple-criteria decision analysis and evolutionary computing tools with GIS has created an array of robust solutions for spatial optimization problems in densely populated areas. Interdisciplinary Approaches to Spatial Optimization Issues is a pivotal reference source that provides vital research on advancements within the field of GIS and evolutionary solutions for spatial optimization issues. While highlighting topics such as computing machinery, vehicular routing, and operational research, this publication is ideally designed for practitioners, technicians, developers, academicians, students, government officials, planners, and researchers seeking current research on applications and improvements within spatial optimization and GIS.
with Practical Examples in MOA
Springer Nature

"Apache Storm is a distributed real-time processing engine. Created by Nathanmarz for Backtype and later open sourced under Apache License 2, it's a scalable and a fault-tolerant engine used to process a massive number of unbounded streams. In this course you will see how simple yet efficient Apache Storm is when it comes to real-time processing. In the course, you will learn about data processing types followed by Apache Storm and its features. You'll learn the core concepts of Apache Storm such as spouts, bolts, topology, and stream grouping, and set up Apache Storm in single-node and multi-node configurations. Also you'll explore how fault-tolerant Apache Storm is. Taking this course will kick-start your experience with Apache Storm; you'll create a scalable, fault-tolerant, real-time processing application while setting a strong base for the fundamentals of the real-time processing paradigm and Apache Storm."--Resource description page.

Introduction to Apache Flink Packt Pub Limited

Learn how to apply the principles of machine learning to time series modeling with this indispensable resource Machine Learning for Time Series Forecasting with Python is an incisive and straightforward examination of one of the most crucial elements of decision-making in finance, marketing, education, and healthcare: time series modeling. Despite the centrality of time series forecasting, few business analysts are familiar with the power or utility of applying machine learning to time series modeling. Author Francesca Lazzeri, a distinguished machine learning scientist and economist, corrects that deficiency by providing readers with comprehensive and approachable explanation and treatment of the application of machine learning to time series forecasting. Written for readers who have little to no experience in time series forecasting or machine learning, the book comprehensively covers all the topics necessary to: Understand time series forecasting concepts, such as stationarity, horizon, trend, and seasonality Prepare time series data for modeling Evaluate time series forecasting models' performance and accuracy Understand when to use neural networks instead of traditional time series models in time series forecasting Machine Learning for Time Series Forecasting with Python is full real-world examples, resources and concrete strategies to help readers explore and transform data and develop usable, practical time series forecasts. Perfect for entry-level data scientists, business analysts, developers, and researchers, this book is an invaluable and indispensable guide to the fundamental and advanced concepts of machine learning applied to time series modeling.

Apache Spark 2: Data Processing and Real-Time Analytics Partridge Publishing

Building scalable and fault-tolerant streaming applications made easy with Spark streamingAbout This Book• Process live data streams more efficiently with better fault recovery using Spark Streaming• Implement and deploy real-time log file analysis• Learn about integration with Advance Spark Libraries – GraphX, Spark SQL, and MLib.Who This Book Is ForThis book is

intended for big data developers with basic knowledge of Scala but no knowledge of Spark. It will help you grasp the basics of developing real-time applications with Spark and understand efficient programming of core elements and applications.What You Will Learn• Install and configure Spark and Spark Streaming to execute applications• Explore the architecture and components of Spark and Spark Streaming to use it as a base for other libraries• Process distributed log files in real-time to load data from distributed sources• Apply transformations on streaming data to use its functions• Integrate Apache Spark with the various advance libraries like MLib and GraphX• Apply production deployment scenarios to deploy your applicationIn DetailUsing practical examples with easy-to-follow steps, this book will teach you how to build real-time applications with Spark Streaming.Starting with installing and setting the required environment, you will write and execute your first program for Spark Streaming. This will be followed by exploring the architecture and components of Spark Streaming along with an overview of libraries/functions exposed by Spark. Next you will be taught about various client APIs for coding in Spark by using the use-case of distributed log file processing. You will then apply various functions to transform and enrich streaming data. Next you will learn how to cache and persist datasets. Moving on you will integrate Apache Spark with various other libraries/components of Spark like Mlib, GraphX, and Spark SQL. Finally, you will learn about deploying your application and cover the different scenarios ranging from standalone mode to distributed mode using Mesos, Yarn, and private data centers or on cloud infrastructure.Style and approachA Step-by-Step approach to learn Spark Streaming in a structured manner, with detailed explanation of basic and advance features in an easy-to-follow Style. Each topic is explained sequentially and supported with real world examples and executable code snippets that appeal to the needs of readers with the wide range of experiences.

Master complex big data processing, stream analytics, and machine learning with Apache Spark Packt Publishing Ltd

Harness the power of Scala to program Spark and analyze tonnes of data in the blink of an eye! About This Book Learn Scala's sophisticated type system that combines Functional Programming and object-oriented concepts Work on a wide array of applications, from simple batch jobs to stream processing and machine learning Explore the most common as well as some complex use-cases to perform large-scale data analysis with Spark Who This Book Is For Anyone who wishes to learn how to perform data analysis by harnessing the power of Spark will find this book extremely useful. No knowledge of Spark or Scala is assumed, although prior programming experience (especially with other JVM languages) will be useful to pick up concepts quicker. What You Will Learn Understand object-oriented & functional programming concepts of Scala In-depth understanding of Scala collection APIs Work with RDD and DataFrame to learn Spark's core abstractions Analysing

structured and unstructured data using SparkSQL and GraphX Scalable and fault-tolerant streaming application development using Spark structured streaming Learn machine-learning best practices for classification, regression, dimensionality reduction, and recommendation system to build predictive models with widely used algorithms in Spark MLlib & ML Build clustering models to cluster a vast amount of data Understand tuning, debugging, and monitoring Spark applications Deploy Spark applications on real clusters in Standalone, Mesos, and YARN In Detail Scala has been observing wide adoption over the past few years, especially in the field of data science and analytics. Spark, built on Scala, has gained a lot of recognition and is being used widely in productions. Thus, if you want to leverage the power of Scala and Spark to make sense of big data, this book is for you. The first part introduces you to Scala, helping you understand the object-oriented and functional programming concepts needed for Spark application development. It then moves on to Spark to cover the basic abstractions using RDD and DataFrame. This will help you develop scalable and fault-tolerant streaming applications by analyzing structured and unstructured data using SparkSQL, GraphX, and Spark structured streaming. Finally, the book moves on to some advanced topics, such as monitoring, configuration, debugging, testing, and deployment. You will also learn how to develop Spark applications using SparkR and PySpark APIs, interactive data analytics using Zeppelin, and in-memory data processing with Alluxio. By the end of this book, you will have a thorough understanding of Spark, and you will be able to perform full-stack data analytics with a feel that no amount of data is too big. Style and approach Filled with practical examples and use cases, this book will hot only help you get up and running with Spark, but will also take you farther down the road to becoming a data scientist.

Big Data Processing with Apache Spark

Packt Publishing Ltd

There is ample evidence that language users, including second-language (L2) users, can predict upcoming information during listening and reading. Yet it is still unclear when, how, and why language users engage in prediction, and what the relation is between prediction and learning. This volume presents a collection of current research, insights, and directions regarding the role of prediction in L2 processing and learning. The contributions in this volume specifically address how different (L1-based) theoretical models of prediction apply to or may be expanded to account for L2 processing, report new insights on factors (linguistic, cognitive, social) that modulate L2 users' engagement in prediction, and discuss the functions that prediction may or may not serve in L2 processing and learning. Taken together, this volume illustrates various fruitful approaches to investigating and accounting for differences in predictive processing within and across individuals, as well as across populations.

Digital Transformation Technology Springer Nature

"This comprehensive tutorial will acquaint you with all the aspects of real-time analytics with

Apache Spark, one of the trending Big Data processing frameworks on the market today. It will show you how to leverage the features of various components of the Spark framework to efficiently process, analyze, and visualize your data. You will learn how to implement the high velocity streaming operation for data processing in order to perform efficient analytics on your real-time data. You'll analyze data using machine learning techniques and graphs. You'll learn about Spark Streaming and create real-world streaming processing that address all the problems that need to be solved. You'll solve problems using Machine Learning techniques and find out about all the tools available in the MLlibtoolkit. You'll find out how to leverage Graphs to solve real-world problems. At the end of this video, you'll also see some useful Machine Learning algorithms with the help of Spark MLlib and will integrate Spark with R. We'll also make sure you're confident and prepared for graph processing, as you'll learn more about the GraphX API. By the end, you'll be well-versed in the aspects of real-time analytics and implement them with Apache Spark."--Resource description page.

Crosslinguistic Influence and Second Language Learning Springer

No need to spend hours ploughing through endless data - let Spark, one of the fastest big data processing engines available, do the hard work for you. Key Features Get up and running with Apache Spark and Python Integrate Spark with AWS for real-time analytics Apply processed data streams to machine learning APIs of Apache Spark Book Description Processing big data in real time is challenging due to scalability, information consistency, and fault-tolerance. This book teaches you how to use Spark to make your overall analytical workflow faster and more efficient. You'll explore all core concepts and tools within the Spark ecosystem, such as Spark Streaming, the Spark Streaming API, machine learning extension, and structured streaming. You'll begin by learning data processing fundamentals using Resilient Distributed Datasets (RDDs), SQL, Datasets, and Dataframes APIs. After grasping these fundamentals, you'll move on to using Spark Streaming APIs to consume data in real time from TCP sockets, and integrate Amazon Web Services (AWS) for stream consumption. By the end of this book, you'll not only have understood how to use machine learning extensions and structured streams but you'll also be able to apply Spark in your own upcoming big data projects. What you will learn Write your own Python programs that can interact with Spark Implement data stream consumption using Apache Spark Recognize common operations in Spark to process known data streams Integrate Spark streaming with Amazon Web Services (AWS) Create a collaborative filtering model with the movielens dataset Apply processed data streams to Spark machine learning APIs Who this book is for Data Processing with Apache Spark is for you if you are a software engineer, architect, or IT professional who wants to explore distributed systems and big data analytics. Although you don't need any knowledge of Spark, prior experience of working with Python is recommended.

Downloading the example code for this book You can download the example code files for all Packt books you have purchased from your account at http://www.PacktPub.com. If you purchased this book elsewhere, you can visit http://www.PacktPub.com/support and register to have the files e-mailed directly to you.
National Academies Press
A practical guide to help you tackle different real-time data processing and analytics problems using the best tools for each scenario About This Book Learn about the various challenges in real-time data processing and use the right tools to overcome them This book covers popular tools and frameworks such as Spark, Flink, and Apache Storm to solve all your distributed processing problems A practical guide filled with examples, tips, and tricks to help you perform efficient Big Data processing in real-time Who This Book Is For If you are a Java developer who would like to be equipped with all the tools required to devise an end-to-end practical solution on real-time data streaming, then this book is for you. Basic knowledge of real-time processing would be helpful, and knowing the fundamentals of Maven, Shell, and Eclipse would be great. What You Will Learn Get an introduction to the established real-time stack Understand the key integration of all the components Get a thorough understanding of the basic building blocks for real-time solution designing Garnish the search and visualization aspects for your real-time solution Get conceptually and practically acquainted with real-time analytics Be well equipped to apply the knowledge and create your own solutions In Detail With the rise of Big Data, there is an increasing need to process large amounts of data continuously, with a shorter turnaround time. Real-time data processing involves continuous input, processing and output of data, with the condition that the time required for processing is as short as possible. This book covers the majority of the existing and evolving open source technology stack for real-time processing and analytics. You will get to know about all the real-time solution aspects, from the source to the presentation to persistence. Through this practical book, you'll be equipped with a clear understanding of how to solve challenges on your own. We'll cover topics such as how to set up components, basic executions, integrations, advanced use cases, alerts, and monitoring. You'll be exposed to the popular tools used in real-time processing today such as Apache Spark, Apache Flink, and Storm. Finally, you will put your knowledge to practical use by implementing all of the techniques in the form of a practical, real-world use case. By the end of this book, you will have a solid understanding of all the aspects of real-time data processing and analytics, and will know how to deploy the solutions in production environments in the best possible manner. Style and Approach In this practical guide to real-time analytics, each chapter begins with a basic high-level concept of the topic, followed by a practical, hands-on implementation of each concept, where you can see the working and execution of it. The book is written in a DIY style, with plenty of practical use cases, well-explained code examples, and

relevant screenshots and diagrams.
Fundamentals, Implementation, and Operation of Streaming Applications Packt Publishing Ltd
The book Kafka Streams - Real-time Stream Processing helps you understand the stream processing in general and apply that skill to Kafka streams programming. This book is focusing mainly on the new generation of the Kafka Streams library available in the Apache Kafka 2.x. The primary focus of this book is on Kafka Streams. However, the book also touches on the other Apache Kafka capabilities and concepts that are necessary to grasp the Kafka Streams programming. Who should read this book? Kafka Streams: Real-time Stream Processing is written for software engineers willing to develop a stream processing application using Kafka Streams library. I am also writing this book for data architects and data engineers who are responsible for designing and building the organization's data-centric infrastructure. Another group of people is the managers and architects who do not directly work with Kafka implementation, but they work with the people who implement Kafka Streams at the ground level. What should you already know? This book assumes that the reader is familiar with the basics of Java programming language. The source code and examples in this book are using Java 8, and I will be using Java 8 lambda syntax, so experience with lambda will be helpful. Kafka Streams is a library that runs on Kafka. Having a good fundamental knowledge of Kafka is essential to get the most out of Kafka Streams. I will touch base on the mandatory Kafka concepts for those who are new to Kafka. The book also assumes that you have some familiarity and experience in running and working on the Linux operating system.
Learning Real Time Processing with Spark Streaming Newnes
The book describes the emergence of big data technologies and the role of Spark in the entire big data stack. It compares Spark and Hadoop and identifies the shortcomings of Hadoop that have been overcome by Spark. The book mainly focuses on the in-depth architecture of Spark and our understanding of Spark RDDs and how RDD complements big data's immutable nature, and solves it with lazy evaluation, cacheable and type inference. It also addresses advanced topics in Spark, starting with the basics of Scala and the core Spark framework, and exploring Spark data frames, machine learning using Mllib, graph analytics using Graph X and real-time processing with Apache Kafka, AWS Kenisis, and Azure Event Hub. It then goes on to investigate Spark using PySpark and R. Focusing on the current big data stack, the book examines the interaction with current big data tools, with Spark being the core processing layer for all types of data. The book is intended for data engineers and scientists working on massive datasets and big data technologies in the cloud. In addition to industry professionals, it is helpful for aspiring data processing professionals and students working in big data processing and cloud computing environments.
Frontiers in Massive Data Analysis Springer
This book presents deep

learning techniques, concepts, and algorithms to classify and analyze big data. Further, it offers an introductory level understanding of the new programming languages and tools used to analyze big data in real-time, such as Hadoop, SPARK, and GRAPHX. Big data analytics using traditional techniques face various challenges, such as fast, accurate and efficient processing of big data in real-time. In addition, the Internet of Things is progressively increasing in various fields, like smart cities, smart homes, and e-health. As the enormous number of connected devices generate huge amounts of data every day, we need sophisticated algorithms to deal, organize, and classify this data in less processing time and space. Similarly, existing techniques and algorithms for deep learning in big data field have several advantages thanks to the two main branches of the deep learning, i.e. convolution and deep belief networks. This book offers insights into these techniques and applications based on these two types of deep learning. Further, it helps students, researchers, and newcomers understand big data analytics based on deep learning approaches. It also discusses various machine learning techniques in concatenation with the deep learning paradigm to support high-end data processing, data classifications, and real-time data processing issues. The classification and presentation are kept quite simple to help the readers and students grasp the basics concepts of various deep learning paradigms and frameworks. It mainly focuses on theory rather than the mathematical background of the deep learning concepts. The book consists of 5 chapters, beginning with an introductory explanation of big data and deep learning techniques, followed by integration of big data and deep learning techniques and lastly the future directions.

Business in Real-Time Using Azure IoT and Cortana Intelligence Suite Engineering Science Reference Step-by-step guide to different data movement and processing techniques, using Google Cloud Platform Services DESCRIPTION Modern businesses are awash with data, making data-driven decision-making tasks increasingly complex. As a result, relevant technical expertise and analytical skills are required to do such tasks. This book aims to equip you with enough knowledge of Cloud Computing in conjunction with Google Cloud Data platform to succeed in the role of a Cloud data expert. The current market is trending towards the latest cloud technologies, which is the need of the hour. Google being the pioneer, is dominating this space with the right set of cloud services being offered as part of GCP (Google Cloud Platform). At this juncture, this book will be very vital and will cover all the services that are being offered by GCP, putting emphasis on Data services. This book starts with sophisticated knowledge on Cloud Computing. It also explains different types of data services/technology and machine learning algorithm/Pre-Trained API through real-business problems, which are built on the Google Cloud Platform (GCP). With some of the latest business examples and hands-on guide, this book will enable the developers

entering the data analytics fields to implement an end-to-end data pipeline, using GCP Data services. Through the course of the book, you will come across multiple industry-wise use cases, like Building Datawarehouse using Big Query, a sample real-time data analytics solution on machine learning and Artificial Intelligence that helped with the business decision, by employing a variety of data science approaches on Google Cloud environment. Whether your business is at the early stage of cloud implementation in its journey or well on its way to digital transformation, Google Cloud's solutions and technologies will always help chart a path to success. This book can be used to develop the GCP concepts in an easy way. It contains many examples showcasing the implementation of a GCP service. It enables the learning of the basic and advance concepts of Google Cloud Data Platform. This book is divided into 7 chapters and provides a detailed description of the core concepts of each of the Data services offered by Google Cloud. KEY FEATURES Learn the basic concept of Cloud Computing along with different Cloud service provides with their supported Models (IaaS/PaaS/SaaS) Learn the basics of Compute Engine, App Engine, Container Engine, Project and Billing setup in the Google Cloud Platform Learn how and when to use Cloud DataFlow, Cloud DataProc and Cloud DataPrep Build real-time data pipeline to support real-time analytics using Pub/Sub messaging service Setting up a fully managed GCP Big Data Cluster using Cloud DataProc for running Apache Spark and Apache Hadoop clusters in a simpler, more cost-efficient manner Learn how to use Cloud Data Studio for visualizing the data on top of Big Query Implement and understand real-world business scenarios for Machine Learning, Data Pipeline Engineering WHAT WILL YOU LEARN By the end of the book, you will have come across different data services and platforms offered by Google Cloud, and how those services/features can be enabled to serve business needs. You will also see a few case studies to put your knowledge to practice and solve business problems such as building a real-time streaming pipeline engine, Scalable Data Warehouse on Cloud, fully managed Hadoop cluster on Cloud and enabling TensorFlow/Machine Learning API’s to support real-life business problems. Remember to practice additional examples to master these techniques. WHO IS THIS BOOK FOR This book is for professionals as well as graduates who want to build a career in Google Cloud data analytics technologies. While no prior knowledge of Cloud Computing or related technologies is assumed, it will be helpful to have some data background and experience. One stop shop for those who wish to get an initial to advance understanding of the GCP data platform. The target audience will be data engineers/professionals who are new, as well as those who are acquainted with the tools and techniques related to cloud and data space. ● Individuals who have basic data understanding (i.e. Data and cloud) and have done some work in the field of data analytics, can refer/use this book to master their knowledge/understanding. ● The highlight of this

book is that it will start with the basic cloud computing fundamentals and will move on to cover the advance concepts on GCP cloud data analytics and hence can be referred across multiple different levels of audiences. Table of Contents 1. GCP Overview and Architecture 2. Data Storage in GCP 3. Data Processing in GCP with Pub/Sub and Dataflow 4. Data Processing in GCP with DataPrep and Dataflow 5. Big Query and Data Studio 6. Machine Learning with GCP 7. Sample Use cases and Examples

Encyclopedia of Ecology MIT Press

During the 1990s the computing industry has witnessed many advances in mobile and enterprise computing. Many of these advances have been made possible by developments in the areas such as modeling, simulation, and artificial intelligence. Within the different areas of enterprise computing - such as manufacturing, health organisation, and commerce - the need for a disciplined, multifaceted, and unified approach to modeling and simulation has become essential. This new book provides a forum for scientists, academics, and professionals to present their latest research findings from the various fields: artificial intelligence, collaborative/distributed computing, modeling, and simulation.

Stream Processing with Apache Spark Packt Publishing Ltd

This book presents cutting-edge research on innovative system interfaces, highlighting both lifecycle development and human–technology interaction, especially in virtual, augmented and mixed reality systems. It describes advanced methodologies and tools for evaluating and improving interface usability, and discusses new models, case studies and good practices. The book addresses the human, hardware, and software factors in the process of developing interfaces for optimizing total system performance, while minimizing costs. It also highlights the forces currently shaping the nature of computing and systems, such as the importance of portability and technologies for reducing power requirements; the need for better assimilation of computation in the environment; and solutions to promote computer and system accessibility for people with special needs. Based on the AHFE 2020 Virtual Conference on Human Factors and Systems Interaction, held on July 16–20, 2020, the book offers a timely survey and a practice-oriented guide for systems interface users and developers alike.

Build Real Time Data Analytics on Google Cloud Platform John Benjamins Publishing Company

Get started with Apache Flink, the open source framework that powers some of the world’s largest stream processing applications. With this practical book, you’ll explore the fundamental concepts of parallel stream processing and discover how this technology differs from traditional batch data processing. Longtime Apache Flink committers Fabian Hueske and Vasia Kalavri show you how to implement scalable streaming applications with Flink’s DataStream API and continuously run and maintain these applications in operational environments. Stream processing is ideal for many use cases, including

low-latency ETL, streaming analytics, and real-time dashboards as well as fraud detection, anomaly detection, and alerting. You can process continuous data of any kind, including user interactions, financial transactions, and IoT data, as soon as you generate them. Learn concepts and challenges of distributed stateful stream processing Explore Flink's system architecture, including its event-time processing mode and fault-tolerance model Understand the fundamentals and building blocks of the DataStream API, including its time-based and statefuloperators Read data from and write data to external systems with exactly-once consistency Deploy and configure Flink clusters Operate continuously running streaming applications

Machine Learning for Data Streams Packt Publishing Ltd

Technology driven witty solutions to everyday Managerial Problems Like it is often told "Solutions at your doorstep", we are completely surrounded by profound managerial solutions waiting to be unearthed from our everyday machines in the form of phones, computers, safety devices, automobile etc. The world of machines abounds with managerial thoughts and solutions. This inspiring book provides us with a new approach in problem solving and addresses the diverse challenges faced in managerial functions today. "Learning Management Back From Machines", is the wonderful story of Krish and his latest creation, MANU – an advanced hyper-intelligent, direct-neural interface-capable humanoid, which helps Krish along in deriving managerial solutions from fellow-machines and machine-processes alike. In the process of learning and observing the history of various technological marvels along with the need for these inventions, we discover a whole new dimension of creative intelligence and learning, waiting to reveal itself all over again. The book is aimed at understanding the core essence of how machines have been made to work and help us discover new and innovative solutions to our everyday social and managerial problems. • RELIGIONS TEACH US MANAGEMENT. • STORIES AND FABLES TEACH US MANAGEMENT. • MANAGEMENT THEORIES TEACH US MANAGEMENT. • NOW EVERYDAY MACHINES WILL TEACH US MANAGEMENT

Distributed Computing and Event Processing using Apache Spark, Flink, Storm, and Kafka Learning Real Time Processing with Spark StreamingBuilding scalable and fault-tolerant streaming applications made easy with Spark streamingAbout This Book• Process live data streams more efficiently with better fault recovery using Spark Streaming• Implement and deploy real-time log file analysis• Learn about integration with Advance Spark Libraries – GraphX, Spark SQL, and MLib.Who This Book Is ForThis book is intended for big data developers with basic knowledge of Scala but no knowledge of Spark. It will help you grasp the basics of developing real-time applications with Spark and understand efficient programming of core elements and applications.What You Will Learn• Install and configure Spark and Spark Streaming to execute applications• Explore the architecture and components of Spark and Spark Streaming to use it as a base for other libraries• Process

distributed log files in real-time to load data from distributed sources• Apply transformations on streaming data to use its functions• Integrate Apache Spark with the various advance libraries like MLib and GraphX• Apply production deployment scenarios to deploy your applicationIn DetailUsing practical examples with easy-to-follow steps, this book will teach you how to build real-time applications with Spark Streaming.Starting with installing and setting the required environment, you will write and execute your first program for Spark Streaming. This will be followed by exploring the architecture and components of Spark Streaming along with an overview of libraries/functions exposed by Spark. Next you will be taught about various client APIs for coding in Spark by using the use-case of distributed log file processing. You will then apply various functions to transform and enrich streaming data. Next you will learn how to cache and persist datasets. Moving on you will integrate Apache Spark with various other libraries/components of Spark like Mlib, GraphX, and Spark SQL. Finally, you will learn about deploying your application and cover the different scenarios ranging from standalone mode to distributed mode using Mesos, Yarn, and private data centers or on cloud infrastructure.Style and approachA Step-by-Step approach to learn Spark Streaming in a structured manner, with detailed explanation of basic and advance features in an easy-to-follow Style. Each topic is explained sequentially and supported with real world examples and executable code snippets that appeal to the needs of readers with the wide range of experiences.Practical Real-time Data Processing and AnalyticsDistributed Computing and Event Processing using Apache Spark, Flink, Storm, and Kafka
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.
Apache Spark 2 Springer Nature
Learning Real Time Processing with Spark Streaming