

Learning Real Time Processing With Spark Streaming

Electricity Pricing: Regulated, Deregulated and Smart Grid Systems presents proven methods for supplying uninterrupted, high-quality electrical power at a reasonable price to the consumer. Illustrating the evolution of the power market from a monopoly to an open access system, this essential text: Covers voltage stability analysis of longitudinal power supply systems using an artificial neural network (ANN) Explains how to improve performance using flexible alternating current transmission systems (FACTS) and high-voltage direct current (HVDC) Takes into account operating constraints as well as generation cost, line overload, and congestion for expected and inadvertent loading stress Goes beyond FACTS and HVDC to provide multi-objective optimization algorithms for the deregulated power market Proposes the use of stochastic optimization techniques in the smart grid, preparing the reader for future development Electricity Pricing: Regulated, Deregulated and Smart Grid Systems offers practical solutions for improving stability, reliability, and efficiency in real-time systems while optimizing electricity cost.

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.

Incorporating approaches from linguistics and psychology, The Handbook of Psycholinguistics explores language processing and language acquisition from an array of perspectives and features cutting edge research from cognitive science, neuroscience, and other related fields. The Handbook provides readers with a comprehensive review of the current state of the field, with an emphasis on research trends most likely to determine the shape of psycholinguistics in the years ahead. The chapters are organized into three parts, corresponding to the major areas of psycholinguistics: production, comprehension, and acquisition. The collection of chapters, written by a team of international scholars, incorporates multilingual populations and neurolinguistic dimensions. Each of the three sections also features an overview chapter in which readers are introduced to the different theoretical perspectives guiding research in the area covered in that section. Timely, comprehensive, and authoritative, The Handbook of Psycholinguistics is a valuable addition to the reference shelves of researchers in psychology, linguistics, and cognitive science, as well as advanced undergraduates and graduate students interested in how language works in the human mind and how language is acquired.

This book includes selected papers from the International Conference on Machine Learning and Information Processing (ICMLIP 2019), held at ISB & M School of Technology, Pune, Maharashtra, India, from December 27 to 28, 2019. It presents the latest developments and technical solutions in the areas of advanced computing and data sciences, covering machine learning, artificial intelligence, human-computer interaction, IoT, deep learning, image processing and pattern recognition, and signal and speech processing.

Building scalable and fault-tolerant streaming applications made easy with Spark streaming
About 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 For
This 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 application
In Detail
Using 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 approach
A 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.

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 not only help you get up and running with Spark, but will also take you farther down the road to becoming a data scientist.

"This course is your guide to performing real-time data analytics and stream processing with Spark. Use different components and tools such as HDFS, HBase, and Hive to process raw data. Learn how tools such as Hive and Pig aid in this process. In this course, you will start off by learning data analysis techniques with Hadoop using tools such as Hive. Furthermore, you will learn to apply these techniques in real-world big data applications. Also, you will delve into Spark and its related tools to perform real-time data analytics, streaming, and batch processing on your application. Finally, you'll learn how to extend your analytics solutions to the cloud."--Resource description page.

This book presents recent advances in intrusion detection systems (IDSs) using state-of-the-art deep learning methods. It also provides a systematic overview of classical machine learning and the latest developments in deep learning. In particular, it discusses deep learning applications in IDSs in different classes: generative, discriminative, and adversarial networks. Moreover, it compares various deep learning-based IDSs based on benchmarking datasets. The book also proposes two novel feature learning models: deep feature extraction and selection (D-FES) and fully unsupervised IDS. Further challenges and research directions are presented at the end of the book. Offering a comprehensive overview of deep learning-based IDS, the book is a valuable reference resource for undergraduate and graduate students, as well as researchers and practitioners interested in deep learning and intrusion detection. Further, the comparison of various deep-learning applications helps readers gain a basic understanding of machine learning, and inspires applications in IDS and other related areas in cybersecurity.

Many monitoring, forecasting, and control operations occur in settings where relationships among key measurements must be learned quickly. Examples are on-line industrial processes where influent material is not consistent over time, energy load or price forecasting where demand characteristics change rapidly, and health management where relationships among monitored variables must be learned for each patient-treatment combination. The solution presented is a new neuro-computing system that learns in real-time, even when data arrival rates are several million measurements per second. The book describes benefits and features of the system, statistical foundations for the system, and several related models. The book also describes available system software.

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 basic 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.

Immersive technology as an umbrella concept consists of multiple emerging technologies including augmented reality (AR), virtual reality (VR), gaming, simulation, and 3D printing. Research has shown immersive technology provides unique learning opportunities for experiential learning, multiple perspectives, and knowledge transfer. Due to its role in influencing learners' cognitive and affective processes, it is shown to have great potential in changing the educational landscape in the decades to come. However, there is a lack of general cognitive and affective theoretical framework to guide the diverse aspects of immersive technology research. In fact, lacking the cognitive and affective theoretical framework has begun to hamper the design and application of immersive technology in schools and related professional training. Cognitive and Affective Perspectives on Immersive Technology in Education is an essential research book that explores methods and implications for the design and implementation of upcoming immersive technologies in pedagogical and professional development settings. The book includes case studies that highlight the cognitive and affective processes in immersive technology as well as the successful applications of immersive technology in education. Featuring a wide range of topics such as curriculum design, K-12 education, and mobile learning, this book is ideal for academicians, educators, policymakers, curriculum developers, instructional designers, administrators, researchers, and students.

Big data has presented a number of opportunities across industries. With these opportunities come a number of challenges associated with handling, analyzing, and storing large data sets. One solution to this challenge is cloud computing, which supports a massive storage and computation facility in order to accommodate big data processing. Managing and Processing Big Data in Cloud Computing explores the challenges of supporting big data processing and cloud-based platforms as a proposed solution. Emphasizing a number of crucial topics such as data analytics, wireless networks, mobile clouds, and machine learning, this publication meets the research needs of data analysts, IT professionals, researchers, graduate students, and educators in the areas of data science, computer programming, and IT development.

This book offers a comprehensive overview of ensemble learning in the field of feature selection (FS), which consists of combining the output of multiple methods to obtain better results than any single method. It reviews various techniques for combining partial results, measuring diversity and evaluating ensemble performance. With the advent of Big Data, feature selection (FS) has become more necessary than ever to achieve dimensionality reduction. With so many methods available, it is difficult to choose the most appropriate one for a given setting, thus making the ensemble paradigm an interesting alternative. The authors first focus on the foundations of ensemble learning and classical approaches, before diving into the specific aspects of ensembles for FS, such as combining partial results, measuring diversity and evaluating ensemble performance. Lastly, the book shows examples of successful applications of ensembles for FS and introduces the new challenges that researchers now face. As such, the book offers a valuable guide for all practitioners, researchers and graduate students in the areas of machine

learning and data mining.

With the widespread interest in digital entertainment and the advances in the technologies of computer graphics, multimedia and virtual reality technologies, the new area of “Edutainment” has been accepted as a union of education and computer entertainment. Edutainment is recognized as an effective way of learning through a medium, such as a computer, software, games or AR/VR applications, that both educates and entertains. The Edutainment conference series was established and followed as a special event for the new interests in e-learning and digital entertainment. The main purpose of Edutainment conferences is the discussion, presentation, and information exchange of scientific and technological developments in the new community. The Edutainment conference series is a very interesting opportunity for researchers, engineers, and graduate students who wish to communicate at these international annual events. The conference series includes plenary invited talks, workshops, tutorials, paper presentation tracks, and panel discussions. The Edutainment conference series was initiated in Hangzhou, China in 2006. Following the success of the first (Edutainment 2006 in Hangzhou, China), the second (Edutainment 2007 in Hong Kong, China), and the third events (Edutainment 2008 in Nanjing, China), Edutainment 2009 was held August 9–11, 2009 in Banff, Canada. This year, we received 116 submissions from 25 different countries and regions - cluding Austria, Canada, China, Denmark, Finland, France, Germany, Greece, Hong Kong, Italy, Japan, Korea, Malaysia, Mexico, The Netherlands, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Taiwan, Trinidad and Tobago, UK, and USA.

The groundbreaking Encyclopedia of Ecology provides an authoritative and comprehensive coverage of the complete field of ecology, from general to applied. It includes over 500 detailed entries, structured to provide the user with complete coverage of the core knowledge, accessed as intuitively as possible, and heavily cross-referenced. Written by an international team of leading experts, this revolutionary encyclopedia will serve as a one-stop-shop to concise, stand-alone articles to be used as a point of entry for undergraduate students, or as a tool for active researchers looking for the latest information in the field. Entries cover a range of topics, including: Behavioral Ecology Ecological Processes Ecological Modeling Ecological Engineering Ecological Indicators Ecological Informatics Ecosystems Ecotoxicology Evolutionary Ecology General Ecology Global Ecology Human Ecology System Ecology The first reference work to cover all aspects of ecology, from basic to applied Over 500 concise, stand-alone articles are written by prominent leaders in the field Article text is supported by full-color photos, drawings, tables, and other visual material Fully indexed and cross referenced with detailed references for further study Writing level is suited to both the expert and non-expert Available electronically on ScienceDirect shortly upon publication

Serious games provide a unique opportunity to fully engage students more than traditional teaching approaches. Understanding the best way to utilize these games and the concept of play in an educational setting is imperative for effectual learning in the 21st century. Gamification in Education: Breakthroughs in Research and Practice is an innovative reference source for the latest academic material on the different approaches and issues faced in integrating games within curriculums. Highlighting a range of topics, such as learning through play, virtual worlds, and educational computer games, this publication is ideally designed for educators, administrators, software designers, and stakeholders in all levels of education.

"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.

Learn how today's businesses can transform themselves by leveraging real-time data and advanced machine learning analytics. This book provides prescriptive guidance for architects and developers on the design and development of modern Internet of Things (IoT) and Advanced Analytics solutions. In addition, Business in Real-Time Using Azure IoT and Cortana Intelligence Suite offers patterns and practices for those looking to engage their customers and partners through Software-as-a-Service solutions that work on any device. Whether you're working in Health & Life Sciences, Manufacturing, Retail, Smart Cities and Buildings or Process Control, there exists a common platform from which you can create your targeted vertical solutions. Business in Real-Time Using Azure IoT and Cortana Intelligence Suite uses a reference architecture as a road map. Building on Azure's PaaS services, you'll see how a solution architecture unfolds that demonstrates a complete end-to-end IoT and Advanced Analytics scenario. What You'll Learn: Automate your software product life cycle using PowerShell, Azure Resource Manager Templates, and Visual Studio Team Services Implement smart devices using Node.JS and C# Use Azure Streaming Analytics to ingest millions of events Provide both "Hot" and "Cold" path outputs for real-time alerts, data transformations, and aggregation analytics Implement batch processing using Azure Data Factory Create a new form of Actionable Intelligence (AI) to drive mission critical business processes Provide rich Data Visualizations across a wide variety of mobile and web devices Who This Book is For: Solution Architects, Software Developers, Data Architects, Data Scientists, and CIO/CTA Technical Leadership Professionals

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 Sustainable Cities and Society.

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 Kinesis, 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.

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

Unleash the data processing and analytics capability of Apache Spark with the language of choice: Java About This Book Perform big data processing with Spark—without having

to learn Scala! Use the Spark Java API to implement efficient enterprise-grade applications for data processing and analytics Go beyond mainstream data processing by adding querying capability, Machine Learning, and graph processing using Spark Who This Book Is For If you are a Java developer interested in learning to use the popular Apache Spark framework, this book is the resource you need to get started. Apache Spark developers who are looking to build enterprise-grade applications in Java will also find this book very useful. What You Will Learn Process data using different file formats such as XML, JSON, CSV, and plain and delimited text, using the Spark core Library. Perform analytics on data from various data sources such as Kafka, and Flume using Spark Streaming Library Learn SQL schema creation and the analysis of structured data using various SQL functions including Windowing functions in the Spark SQL Library Explore Spark Mlib APIs while implementing Machine Learning techniques to solve real-world problems Get to know Spark GraphX so you understand various graph-based analytics that can be performed with Spark In Detail Apache Spark is the buzzword in the big data industry right now, especially with the increasing need for real-time streaming and data processing. While Spark is built on Scala, the Spark Java API exposes all the Spark features available in the Scala version for Java developers. This book will show you how you can implement various functionalities of the Apache Spark framework in Java, without stepping out of your comfort zone. The book starts with an introduction to the Apache Spark 2.x ecosystem, followed by explaining how to install and configure Spark, and refreshes the Java concepts that will be useful to you when consuming Apache Spark's APIs. You will explore RDD and its associated common Action and Transformation Java APIs, set up a production-like clustered environment, and work with Spark SQL. Moving on, you will perform near-real-time processing with Spark streaming, Machine Learning analytics with Spark MLlib, and graph processing with GraphX, all using various Java packages. By the end of the book, you will have a solid foundation in implementing components in the Spark framework in Java to build fast, real-time applications. Style and approach This practical guide teaches readers the fundamentals of the Apache Spark framework and how to implement components using the Java language. It is a unique blend of theory and practical examples, and is written in a way that will gradually build your knowledge of Apache Spark.

This volume includes a number of articles, some of which have previously appeared in Cognitive Science. The contributions as a whole present a fair sample of the late-1980s research and practice in connectionist models.

Design, process, and analyze large sets of complex data in real time About This Book Get acquainted with transformations and database-level interactions, and ensure the reliability of messages processed using Storm Implement strategies to solve the challenges of real-time data processing Load datasets, build queries, and make recommendations using Spark SQL Who This Book Is For If you are a Big Data architect, developer, or a programmer who wants to develop applications/frameworks to implement real-time analytics using open source technologies, then this book is for you. What You Will Learn Explore big data technologies and frameworks Work through practical challenges and use cases of real-time analytics versus batch analytics Develop real-world use cases for processing and analyzing data in real-time using the programming paradigm of Apache Storm Handle and process real-time transactional data Optimize and tune Apache Storm for varied workloads and production deployments Process and stream data with Amazon Kinesis and Elastic MapReduce Perform interactive and exploratory data analytics using Spark SQL Develop common enterprise architectures/applications for real-time and batch analytics In Detail Enterprise has been striving hard to deal with the challenges of data arriving in real time or near real time. Although there are technologies such as Storm and Spark (and many more) that solve the challenges of real-time data, using the appropriate technology/framework for the right business use case is the key to success. This book provides you with the skills required to quickly design, implement and deploy your real-time analytics using real-world examples of big data use cases. From the beginning of the book, we will cover the basics of varied real-time data processing frameworks and technologies. We will discuss and explain the differences between batch and real-time processing in detail, and will also explore the techniques and programming concepts using Apache Storm. Moving on, we'll familiarize you with "Amazon Kinesis" for real-time data processing on cloud. We will further develop your understanding of real-time analytics through a comprehensive review of Apache Spark along with the high-level architecture and the building blocks of a Spark program. You will learn how to transform your data, get an output from transformations, and persist your results using Spark RDDs, using an interface called Spark SQL to work with Spark. At the end of this book, we will introduce Spark Streaming, the streaming library of Spark, and will walk you through the emerging Lambda Architecture (LA), which provides a hybrid platform for big data processing by combining real-time and precomputed batch data to provide a near real-time view of incoming data. Style and approach This step-by-step is an easy-to-follow, detailed tutorial, filled with practical examples of basic and advanced features. Each topic is explained sequentially and supported by real-world examples and executable code snippets.

There's growing interest in learning how to analyze streaming data in large-scale systems such as web traffic, financial transactions, machine logs, industrial sensors, and many others. But analyzing data streams at scale has been difficult to do well—until now. This practical book delivers a deep introduction to Apache Flink, a highly innovative open source stream processor with a surprising range of capabilities. Authors Ellen Friedman and Kostas Tzoumas show technical and nontechnical readers alike how Flink is engineered to overcome significant tradeoffs that have limited the effectiveness of other approaches to stream processing. You'll also learn how Flink has the ability to handle both stream and batch data processing with one technology. Learn the consequences of not doing streaming well—in retail and marketing, IoT, telecom, and banking and finance Explore how to design data architecture to gain the best advantage from stream processing Get an overview of Flink's capabilities and features, along with examples of how companies use Flink, including in production Take a technical dive into Flink, and learn how it handles time and stateful computation Examine how Flink processes both streaming (unbounded) and batch (bounded) data without sacrificing performance

If you are a Java developer who wants to enter into the world of real-time stream processing applications using Apache Storm, then this book is for you. No previous experience in Storm is required as this book starts from the basics. After finishing this book, you will be able to develop not-so-complex Storm applications.

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.

The four-volume set LNCS 8925, 8926, 8927 and 8928 comprises the thoroughly refereed post-workshop proceedings of the Workshops that took place in conjunction with the 13th European Conference on Computer Vision, ECCV 2014, held in Zurich, Switzerland, in September 2014. The 203 workshop papers were carefully reviewed and selected for inclusion in the proceedings. They were presented at workshops with the following themes: where computer vision meets art; computer vision in vehicle technology; spontaneous facial behavior analysis; consumer depth cameras for computer vision; "chlearn" looking at people: pose, recovery, action/interaction, gesture recognition; video event categorization, tagging and retrieval towards big data; computer vision with local binary pattern variants; visual object tracking challenge; computer vision + ontology applies cross-disciplinary technologies; visual perception of affordance and functional visual primitives for scene analysis; graphical models in computer vision; light fields for computer vision; computer vision for road scene understanding and autonomous driving; soft biometrics; transferring and adapting source knowledge in computer vision; surveillance and re-identification; color and photometry in computer vision; assistive computer vision and robotics; computer vision problems in plant phenotyping; and non-rigid shape analysis and deformable image alignment. Additionally, a panel discussion on video segmentation is included.

This volume of Advances in Intelligent and Soft Computing contains accepted papers presented at SOCO 2010 held in the beautiful and historic city of Guimarães, Portugal, June 2010. The global purpose of SOCO conferences has been to provide a broad and interdisciplinary forum for soft computing and associated paradigms, which are playing increasingly important roles in an important number of industrial and environmental applications fields. Soft computing represents a collection or set of computational techniques in machine learning, computer science and some engineering disciplines, which investigate, simulate and analyze very complex issues and phenomena. This workshop is mainly focused on its industrial and environmental applications. The SOCO 2010 is the 5th International Workshop on Soft Computing Models in Industrial Applications and provides interesting opportunities to present and discuss the latest theoretical advances and real world applications in this multidisciplinary research field. This volume presents the papers accepted for the 2010 edition, both for the main event and the Special Sessions. SOCO 2010 Special Sessions are a very useful tool in order to complement the regular program with new or emerging topics of particular interest to the participating community. Special Sessions that emphasize on multi-disciplinary and transversal aspects, as well as cutting-edge topics were especially encouraged and welcome. SOCO 2010 included a total of 3 Special Sessions: Ensemble Learning and Formation Fusion for Industrial Applications; Soft Computing for Service Management; Hybrid Intelligent Systems and Applications.

"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 MLlib toolkit. 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.

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.

Processing data streams has raised new research challenges over the last few years. This book provides the reader with a comprehensive overview of stream data processing, including famous prototype implementations like the Nile system and the TinyOS operating system. Applications in security, the natural sciences, and education are presented. The huge bibliography offers an excellent starting point for further reading and future research.

Learning Real Time Processing with Spark Streaming

This book provides readers with a comprehensive and recent exposition in deep learning and its multidisciplinary applications, with a concentration on advances of deep learning architectures. The book discusses various artificial intelligence (AI) techniques based on deep learning architecture with applications in natural language processing, semantic knowledge, forecasting and many more. The authors shed light on various applications that can benefit from the use of deep learning in pattern recognition, person re-identification in surveillance videos, action recognition in videos, image and video captioning. The book also highlights how deep learning concepts can be interwoven with more modern concepts to yield applications in multidisciplinary fields. Presents a comprehensive look at deep learning and its multidisciplinary applications, concentrating on advances of deep learning architectures; Includes a survey of deep learning problems and solutions, identifying the main open issues, innovations and latest technologies; Shows industrial

deep learning in practice with examples/cases, efforts, challenges, and strategic approaches.

This book introduces the advantages of parallel processing and details how to use it to deal with common signal processing and control algorithms. The text includes examples and end-of-chapter exercises, and case studies to put theoretical concepts into a practical context.

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 Mei Cookbook

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.

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.

[Copyright: 9aaa7a8e596ffbf12ae4017f33350c4a](#)