Apache Spark Machine Learning Blueprints-Alex Liu 2016-05-30 Develop a range of cutting-edge machine learning projects with Apache Spark using this actionable guide About This Book Customize Apache Spark and R to fit your analytical needs in customer research, fraud detection, risk analytics, and recommendation engine development. Develop a set of practical Machine Learning applications that can be implemented in real-life projects. A comprehensive, project-based guide to improve and refine your predictive models for practical implementation. Who This Book Is For If you are a data scientist, a data analyst, or an R and SPSS user with a good understanding of machine learning concepts, algorithms, and techniques, then this is the book for you. Some basic understanding of Spark and its core elements and application is required. What You Will Learn Set up Apache Spark and discover its impressive processing power Use Spark and R and unlock detailed business insights essential for decision making Build machine learning systems with Spark that can detect fraud and analyze financial risks. Build predictive models focusing on customer scoring and service ranking. Build a recommendation systems using SPSS on Apache Spark. Tackle parallel computing and find out how it can support your machine learning projects. Turn open data and communication data into actionable insights by making use of various forms of machine learning. In Detail There's a reason why Apache Spark has become one of the most popular tools in Machine Learning – its ability to handle huge datasets at an impressive speed means you can be much more responsive to the data at your disposal. This book shows you Spark at its very best, demonstrating how to connect it with R and unlock maximum value not only from the tool but also from your data. Packed with a range of project “blueprints” that demonstrate some of the most interesting challenges that Spark can help you tackle, you'll find out how to use Spark notebooks and access, clean, and join different datasets before putting your knowledge into practice with some real-world projects, in which you will see how Spark Machine Learning can help you with everything from fraud detection to analyzing customer attrition. You'll also find out how to build a recommendation engine using Spark's parallel computing powers. Style and Approach This book offers a step-by-step approach to setting up Apache Spark, and use other analytical tools with it to process Big Data and build machine learning projects. The initial chapters focus more on the theory aspect of machine learning with Spark, while each of the latter chapters focuses on building standalone projects using Spark. Hadoop Blueprints-Anurag Shrivastava 2016-09-30 Use Hadoop to solve business problems by learning from a rich set of real-life case studies. About This Book Solve real-world business problems using Hadoop and other Big Data technologies. Build efficient data lakes in Hadoop, and develop systems for various business cases like improving marketing campaigns, fraud detection, and more. Powerful packed with six case studies to get you going with Hadoop for Business Intelligence. Who This Book Is For If you are interested in building efficient business solutions using Hadoop, this is the book for you. This book assumes that you have basic knowledge of Hadoop, Java, and any scripting language. What You Will Learn Learn about the evolution of Hadoop as the big data platform. Understand the basics of Hadoop architecture Build a 360-degree view of your customer using Spark and Hive. Build and run classification models on Hadoop using BigML. Use Spark and Hadoop to build a fraud detection system. Develop a churn detection system using Java and MapReduce. Build an IoT-based data collection and visualization systemGet to grips with building a Hadoop-based Data Lake for large enterprises. Learn about the coexistence of NoSQL and In-Memory databases in the Hadoop ecosystem. In Detail If you have a basic understanding of Hadoop and want to put your knowledge to use to build fantastic Big Data solutions for business, then this book is for you. Build six real-life, end-to-end solutions using the tools in the Hadoop ecosystem, and take your knowledge of Hadoop to the next level. Start off by understanding various business problems which can be solved using Hadoop. You will also get acquainted with the common architectural patterns which are used to build Hadoop-based solutions. Build a 360-degree view of the customer by working with different types of data, and build an efficient fraud detection system for a financial institution. You will also develop a system in Hadoop to improve the effectiveness of marketing campaigns. Build a churn detection system for a telecom company. Develop an Internet of Things (IoT) system to monitor the environment in a factory, and build a data lake – all making use of the concepts and techniques mentioned in this book. The book covers other technologies and frameworks like Apache Spark, Hive, Sqoop, and more, and how they can be used in conjunction with Hadoop. You will be able to try out the solutions explained in the book and use the knowledge gained to extend them further in your own problem space. Style and approach This is an example-driven book where each chapter covers a simple business problem and describes its solution by explaining the structure of a dataset and tools required to process it. Every project is demonstrated with a step-by-step approach, and explained in a very easy-to-understand manner. Learning Spark-Jules S. Damji 2020-07-16 Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, you’ll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLLow. Frank Kane's Taming Big Data with Apache Spark and Python-Frank Kane 2017-06-30 Frank Kane's hands-on Spark training course, based on his bestselling Taming Big Data with Apache Spark and Python videos, now available in a book. Understand and analyze large data sets using Spark on a single system or on a cluster. About This Book Understand how Spark can be distributed across computing clusters. Develop and run Spark jobs efficiently using Python A hands-on tutorial by Frank Kane with over 15 real-world examples teaching you Big Data processing using Spark. Who This Book Is For If you have some programming experience in Python, and want to learn how to process large amounts of data using Apache Spark, Frank Kane’s Taming Big Data with Apache Spark and Python will also help you. What You Will Learn Find out how you can identify Big Data problems and Spark problems. Install and run Apache Spark on your computer or on a cluster. Analyze large data sets across many CPUs using Spark's Resilient Distributed Datasets. Implement machine learning on Spark using the MLlib library. Process continuous streams of data in real time using the Spark streaming module. Perform complex network analysis using Spark's GraphX library. Use Amazon's Elastic MapReduce service to run your Spark jobs on a cluster. In Detail Frank Kane's Taming Big Data with Apache Spark and Python is your companion to learning Apache Spark in a hands-on manner. Frank will start you off by teaching you how to set up Spark on a single system or on a cluster, and you'll soon move on to analyzing large data sets using Spark RDD, and developing and running effective Spark jobs using Python. Apache Spark has emerged as the next big thing in the Big Data domain – quickly rising from an ascending technology to an established superstar in just a matter of years. Spark allows you to quickly extract actionable insights from large amounts of data, on a real-time basis, making it an essential tool in many modern businesses. Frank has packed this book with over 15 interactive, fun-filled examples relevant to the real world, and he will empower you to understand the Spark ecosystem and implement production-grade real-time Spark projects with ease. Style and approach Frank Kane's Taming Big Data with Apache Spark and Python is a hands-on tutorial with over 15 real-
About the book Grokking Deep Reinforcement Learning uses engaging exercises to teach you how to build and train systems that explore and learn based on the responses of the environment. Grokking Deep Reinforcement Learning will be your definitive guide to batch and stream data processing with Apache Flink. The book begins with introducing the Apache Flink ecosystem, setting it up and using the DataSet and DataStream API for processing stream and batch data applications. As a result of your learning, you will cover various recipes for stream optimization and troubleshooting.

Learning Apache Flink-Tannay Deshpande 2017-02-20 Discover the definitive guide to crafting lightning-fast data processing for distributed systems with Apache Flink! About This Book Build your expertise in processing real-time data with Apache Flink and its ecosystem Gain insights into the workings of all components of Apache Flink such as FlinkML, Celly, and Table API Fill in real-world use cases Explain the capabilities like distributed data streaming, in-memory processing, pipelining and iteration operators to improve performance. Solve real world big data problems with real time in-memory and disk-based processing capabilities of Apache Flink. Who This Book Is For Big data developers who are looking to process batch and real-time data on distributed systems. Basic knowledge of Hadoop and big data is assumed. Reasonable knowledge of Java or Scala is expected. What You Will Learn Learn how to build to end to end real time analytics projects Integrate with existing big data stack and utilize existing infrastructure Build predictive analytics applications using FlinkML Use graph library to perform graph querying and search. Understand Flink’s - "Streaming First" architecture to implementing real streaming applications Learn Flink Logging and Monitoring best practices in order to efficiently design your data pipelines Explore the detailed processes to deploy Flink cluster on Amazon Web Services(AWS) and Google Cloud Platform (GCP). In Detail With the advent of massive computer systems, organizations are generating large amounts of data at an incredible speed. With the help of efficient tools and techniques, managing big data processing, Apache Flink, is designed to process continuous streams of data at a lightning fast pace. This book will be your definitive guide to batch and stream data processing with Apache Flink. The book begins with introducing the Apache Flink ecosystem, setting it up and using the DataSet and DataStream API for processing stream and batch applications. As a result of your learning, you will cover various recipes for stream optimization and troubleshooting. Data Analysis with Python offers a modern approach to data analysis so that you can work with the power of programming, complex algorithms, and AI. Use Python and its extensive libraries to experiment with different techniques and evaluate their benefits and limitations using real-world applications in a tutorial style. This book is perfect for Python and data science enthusiasts who wish to learn and implement various machine learning techniques. You'll move on to evolutionary computing, multibandit algorithms, and reinforcement learning. The book includes a comprehensive overview of parallel computing in Scala and Akka followed by a description of Apache Spark and its ML library. With updated codes based on the latest version of Scala and comprehensive examples, this book will ensure that you have more than just a solid fundamental knowledge in machine learning with Scala. Style and approach This book is a comprehensive guide that covers advanced features of the Apache Flink, and communicates them with a practical understanding of the underlying concepts for how, when, and why to use them. Grokking Deep Reinforcement Learning-Miguel Morales 2020-11-10 Grokking Deep Reinforcement Learning uses engaging exercises to teach you how to build deep learning systems. This book combines annotated Python code with intuitive explanations to explore DRL techniques. You’ll see how algorithms function and learn to develop your own DRL agents using evaluative feedback. Suggestive feedback helps you develop an algorithm that can learn to solve the things that cause us to experience pain and failure. We embrace and build on the things that give us reward and success. This common pattern is the foundation of deep reinforcement learning: building machine learning systems that explore and learn based on the responses of the environment. Grokking Deep Reinforcement Learning introduces this powerful machine learning approach, using examples, illustrations, exercises, and crystal-clear teaching. You’ll love the perfectly paced teaching and the clever, engaging writing style as you dig into the core of Deep Reinforcement Learning and the techniques, and practical applications in this emerging field. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Learn by interacting with our environment, and the rewards or punishments we experience guide our future behavior. Deep reinforcement learning brings that same natural process to artificial intelligence, analyzing results to uncover the most efficient ways forward. DRL agents can improve by learning from experience, evaluating strategies, and comparing their performance with the best of the best. You’ll learn to build and train systems that explore and learn based on the responses of the environment. Grokking Deep Reinforcement Learning uses engaging exercises to teach you how to build deep learning systems. This book combines annotated Python code with intuitive explanations to explore DRL techniques. You’ll see how algorithms function and learn to develop your own DRL agents using evaluative feedback. What’s inside An introduction to reinforcement learning DRL agents with human-like behaviors Applying DRL to complex situations About the reader For developers with basic deep learning experience. About the author Miguel Morales works on reinforcement learning at Lockheed Martin and is an instructor for the Georgia Institute of Technology’s Reinforcement Learning and Control course. About Flink & Scala for Machine Learning-Patrick R. Nicolas 2017-09-26 Leverage Scala and Machine Learning to study and construct systems that can learn from data About This Book Explore a broad variety of data processing, machine learning, and genetic algorithms through diagrams, mathematical formulation, and updated source code in Scala Take your expertise in Scala programming to the next level by creating and customizing AI applications experiment with different techniques and evaluate their benefits and limitations using real-world applications in a tutorial style. This book is perfect for Python and data science enthusiasts who wish to learn and implement various machine learning techniques. You'll move on to evolutionary computing, multibandit algorithms, and reinforcement learning. Finally, the book includes a comprehensive overview of parallel computing in Scala and Akka followed by a description of Apache Spark and its ML library. With updated codes based on the latest version of Scala and comprehensive examples, this book will ensure that you have more than just a solid fundamental knowledge in machine learning with Scala. Style and approach This book is a comprehensive guide that covers advanced features of the Apache Flink, and communicates them with a practical understanding of the underlying concepts for how, when, and why to use them.
The authors leverage cutting-edge functional programming techniques and extend Spark with engineering/analysis approaches designed for Spark. They use Resilient Distributed Datasets (RDDs) for caching, functional Python, and program with the Spark API, including transformations and actions. They help you make the most of the Spark Cluster Architecture and develop Spark applications with Scala and how it fits into the Big Data landscape. They deploy and run Spark locally or in the cloud and interact with Spark from the shell. They embark on a new career in the booming area of Big Data. They learn how to use Spark SQL and DataFrames and understand the streaming capabilities of PySpark. They also get a thorough overview of machine learning capabilities of PySpark using ML and MLlib, graph processing using GraphFrames, and polyglot persistence using Blaze. Finally, they learn how to deploy your applications to the cloud using the spark-submit command. By the end of this book, you will have established a firm understanding of PySpark and programmed on PySpark 2.0. This book takes a very comprehensive, step-by-step approach so you understand how the Spark ecosystem can be used with Python to develop efficient, scalable solutions. Every chapter is standalone and written in a very easy-to-understand manner, with a focus on both the hows and the whys of each concept.
Next, we cover the charting and plotting features of Python in conjunction with Spark data processing. After that, we take a look at Spark’s stream processing, machine learning, and graph processing libraries. The last chapter is about the tools you learned from the preceding chapters and how to use them in a real-world application. By the end of this book, you will have all the knowledge you need to develop efficient large-scale applications using Apache Spark. Style and approach: Learn about Spark’s infrastructure with this practical tutorial. With the help of real-world use cases on the main features of Spark we offer an easy introduction to the framework. Mastering Spark with R-Javier Luraschi 2019-10-07 If you’re like most R users, you have deep knowledge and love for extending and application continues to develop a real-world Spark. This will make a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes Use statistical modeling and R integration into Spark for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including user-based, item-based, and KNN CF-Model-based methods including matrix factorization and SVD-Applying the mathematical explanation for implementing Recurrent Neural Networks (RNN) Get hands on practice of deep learning and their implementation with Hadoop. In Detail This book will teach you how to deploy large-scale dataset in deep neural networks with Hadoop for optimal performance. Starting with understanding what deep learning is, and what the various models associated with deep neural networks are, this book will then show you how to setup the Hadoop environment for deep learning. In this book, you will also learn how to overcome the challenges that you face while implementing distributed deep learning with large-scale unstructured datasets. The book will also show you how you can implement and parallelize the widely used deep learning models such as Deep Belief Networks, Convolutional Neural Networks, Recurrent Neural Networks, Restricted Boltzmann Machines and autocoder using the popular deep learning library deeplearning4j. Get in-depth mathematical explanations and visual representations to help you understand the design and implementation center around the techniques described, which will give you a more practical perspective, the book will also teach you the implementation of large-scale video processing, image processing and natural language processing on Hadoop. By the end of this book, you will know how to deploy various deep neural networks in distributed systems using Hadoop. Style and approach: This book takes a comprehensive, step-by-step approach to implement efficient deep learning models on Hadoop. It starts from the basics and moves to more advanced based collaboration and team-building. The algorithms will be presented with real-life case studies and research papers. Building Recommender Systems with Machine Learning and AI: Help People Discover New Products and Content with Deep Learning, Neural Networks, and Mach-Frank Kane 2018-08-11 Learn how to build recommender systems from one of Amazon’s pioneers in the field. Frank Kane spent over nine years in Amazon, where he managed and led the development of many of Amazon’s personalized product recommendation technologies.You’ve seen automated recommendations everywhere - on Netflix’s home page, on YouTube, and on Amazon as these machine learning algorithms learn about your unique interests, and show the best products or content tailored to you. These systems use extensive training data, and each employs out there, and by understanding how they work, you’ll become very valuable to them. This book is adapted from Frank’s popular online course published by Sundog Education, so you can expect lots of visual aids from its slides and a conversational, accessible tone throughout the book. The graphics and scripts from over 300 slides are included, and you have access to all of the source code associated with it as well. We’ll cover tried and true collaborative filtering techniques based on the neighborhood model and as they have been adapted to more modern techniques including matrix factorization and even deep learning with artificial neural networks. Along the way, you’ll learn from Frank’s extensive industry experience to understand the real-world challenges you’ll encounter when applying these algorithms at large scale and with real-world data. This book is very hands-on; you’ll develop your own framework for evaluating and combining many different recommendation algorithms together, and you’ll even build your own neural networks using TensorFlow to generate recommendations from real-world movie ratings from real people. We’ll cover: Building a recommendation engine Evaluating recommender systems-Content-based filtering using item attributes-Neighborhood-based collaborative filtering with user-based, item-based, and KNN CF-Model-based methods including matrix factorization and SVD-Applying deep learning, AI, and artificial neural networks to recommendations-Session-based recommendations with recursive neural networks-Scaling to massive data sets with Apache Spark machine learning, Amazon DSSTNE deep learning, and Spark MLlib-Autoregressive and factorization model with Hadoop-The fundamentals of recommendations-Case studies from YouTube and Netflix-Building hybrid recommender systems-Cardinality estimation with cluster analysis-Learning to rank with user-based and item-based methods. This comprehensive book takes you all the way from the early days of collaborative filtering, to bleeding-edge applications of deep neural networks and modern machine learning techniques for recommending the best items to every individual user. The coding exercises for this book use the Python programming language. We include an intro to Python if you’re new to it, but you’ll need some prior programming experience in order to use this book successfully. This book also provides a short introduction to deep learning, so you’ll need to be familiar with the field of artificial intelligence, but you’ll need to be able to understand new computer algorithms. Dive in, and learn about one of the most interesting and lucrative applications of machine learning and deep learning there is! Principles of Data Integration-AnHai Doan 2012-06-25 How do you approach answering queries when your data is stored in multiple databases that were designed independently by different people? This is first comprehensive book covering all the essential topics of data integration, from basic principles to the most advanced algorithms and applications. Offers a range of data integration solutions enabling you to focus on what is most relevant to the problem at hand. Enables you to build your own algorithms and implement your own data integration applications. Deep Learning with Hadoop-Dipayan Dev 2017-02-28 Build, implement and scale distributed deep learning models for large-scale datasets About This Book Get to grips with the deep learning concepts and set up Hadoop to scale across servers to accelerate your algorithm. This book provides a comprehensive tutorial to distributed deep learning with Hadoop Who This Book Is For If you are a data scientist who wants to learn how to perform deep learning on Hadoop, this is the book for you. Knowledge of the basic machine learning concepts and some understanding of Hadoop is required to make the best use of this book. What You Will Learn Explore Deep Learning and various models associated with it Understand the challenges of implementing distributed deep learning with Hadoop and how to overcome it Implement Convolutional Neural Networks (CNNs) and Restricted Boltzmann Machines (RBM) Understand the mathematical explanation for implementing Recurrent Neural Networks (RNN) Get hands on practice of deep learning and their implementation with Hadoop. In Detail This book will teach you how to deploy large-scale dataset in deep neural networks with Hadoop for optimal performance. Starting with understanding what deep learning is, and what the various models associated with deep neural networks are, this book will then show you how to setup the Hadoop environment for deep learning. In this book, you will also learn how to overcome the challenges that you face while implementing distributed deep learning with large-scale unstructured datasets. The book will also show you how you can implement and parallelize the widely used deep learning models such as Deep Belief Networks, Convolutional Neural Networks, Recurrent Neural Networks, Restricted Boltzmann Machines and autocoder using the popular deep learning library deeplearning4j. Get in-depth mathematical explanations and visual representations to help you understand the design and implementation center around the techniques described, which will give you a more practical perspective, the book will also teach you the implementation of large-scale video processing, image processing and natural language processing on Hadoop. By the end of this book, you will know how to deploy various deep neural networks in distributed systems using Hadoop. Style and approach: This book takes a comprehensive, step-by-step approach to implement efficient deep learning models on Hadoop. It starts from the basics and moves to more advanced based collaboration and team-building. The algorithms will be presented with real-life case studies and research papers.
that adapts fast enough to thrive. Clear away unnecessary governance processes, obsolete "command and control" decision-making, and better risk mitigation. This guide shows leaders how to use the breakthrough EDGE from their vast in-the-trenches experience that sustainable digital transformation requires far more than adopting world's leading agile pioneers and a coauthor of the Agile Manifesto. He, Linda Luu, and David Robinson know Digital Transformation is your guide to using this operating model for innovation. Jim Highsmith is one of the innovation happens at the edge of chaos: the messy, risky, and uncertain threshold between randomness and structure. However, there is no guaranteed formula for success. This book will provide you with the tools and strategies to achieve your goals. Register your product for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

High Performance Apache Spark on K8s 2017-05-25 Apache Spark is amazing when everything clicks. But if you haven’t seen the performance improvements you expected, or still don’t feel confident enough to use Spark in production, this practical book is for you. Authors Holden Karau and Rachel Warren demonstrate performance optimizations to help your Spark queries run faster and handle larger data sizes, while using fewer resources. Ideal for software engineers, data engineers, developers, and system administrators working with large-scale data applications, this book provides practical advice and concrete examples that will help you optimize your Spark code.

Data Science for Business, 2nd Edition 2019-06-07 Data science teams looking to turn research into useful analytics applications require not only the right tools, but also the right approach if they’re to succeed. With the revised second edition of this hands-on guide, up-and-coming data scientists will learn how to use the Agile Data Science development methodology to build data applications with Python, Apache Spark, Kafka, and other tools. Author Russell Jurney demonstrates how to use a variety of tools and programming languages such as Python, R, and Scala to create data-driven applications that can be deployed in the real world. This book covers how to use the popular Apache Spark framework, how to use the Apache Spark web application with Spark Streaming, and how to use the Apache Spark SQL library to create scalable data applications.

The Great Machine Learning Blunders and How to Avoid Them 2019-05-09 Machine learning is a rapidly evolving field, and it can be challenging to keep up with the latest trends and techniques. This book provides a comprehensive overview of the most common mistakes that machine learning practitioners make, along with practical advice on how to avoid them. Whether you're a seasoned data scientist or just getting started, this book is a valuable resource for anyone looking to improve their machine learning models and avoid common pitfalls. The Great Machine Learning Blunders and How to Avoid Them is an essential read for anyone working with data and machine learning.
understanding of using AWS services to design, build, secure, and maintain analytics solutions that provide insight from data. It assesses an examinee’s ability to define AWS data analytics services and understand how they integrate with each other, and explain how AWS data analytics services fit in the data lifecycle of collection, storage, processing, and visualization. The book focuses on the following domains: Collection • Storage and Data Management • Processing • Analysis and Visualization • Data Security This is your opportunity to take the next step in your career by expanding and validating your skills on the AWS cloud. AWS is the frontrunner in cloud computing products and services, and the AWS Certified Data Analytics Study Guide: Specialty exam will get you fully prepared through expert content, and real-world knowledge, key exam essentials, chapter review questions, and much more. Written by an AWS subject-matter expert, this study guide covers exam concepts, and provides key review on exam topics. Readers will also have access to Sybex’s superior online interactive learning environment and test bank, including chapter tests, practice exams, a glossary of key terms, and electronic flashcards. And included with this version of the book, XtremeLabs virtual labs that run from your browser. The registration code is included with the book and gives you 6 months of unlimited access to XtremeLabs AWS Certified Data Analytics Labs with 3 unique lab modules based on the book.

Practical Real-time Data Processing and Analytics-Shilpi Saxena 2017-09-28 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.

Big Data Analytics: Systems, Algorithms, Applications-C.S.R. Prabhu 2019-10-14 This book provides a comprehensive survey of techniques, technologies and applications of Big Data and its analysis. The Big Data phenomenon is increasingly impacting all sectors of business and industry, producing an emerging new information ecosystem. On the applications front, the book offers detailed descriptions of various application areas for Big Data Analytics in the important domains of Social Semantic Web Mining, Banking and Financial Services, Capital Markets, Insurance, Advertisement, Recommendation Systems, Bio-Informatics, the IoT and Fog Computing, before delving into issues of security and privacy. With regard to machine learning techniques, the book presents all the standard algorithms for learning – including supervised, semi-supervised and unsupervised techniques such as clustering and reinforcement learning techniques to perform collective Deep Learning. Multi-layered and nonlinear learning for Big Data are also covered. In turn, the book highlights real-life case studies on successful implementations of Big Data Analytics at large IT companies such as Google, Facebook, LinkedIn and Microsoft. Multi-sectorial case studies on domain-based companies such as Deutsche Bank, the power provider Opower, Delta Airlines and a Chinese City Transportation application represent a valuable addition. Given its comprehensive coverage of Big Data Analytics, the book offers a unique resource for undergraduate and graduate students, researchers, educators and IT professionals alike.

Thoughtful Machine Learning with Python-Matthew Kirk 2017-01-16 Gain the confidence you need to apply machine learning in your daily work. With this practical guide, author Matthew Kirk shows you how to integrate and test machine learning algorithms in your code, without the academic subtext. Featuring graphs and highlighted code examples throughout, the book features tests with Python’s Numpy, Pandas, Scikit-Learn, and SciPy data science libraries. If you’re a software engineer or business analyst interested in data science, this book will help you: Reference real-world examples to test each algorithm through engaging, hands-on exercises Apply test-driven development (TDD) to write and run tests before you start coding Explore techniques for improving your machine-learning models with data extraction and feature development Watch out for the risks of machine learning, such as underfitting or overfitting data Work with K-Nearest Neighbors, neural networks, clustering, and other algorithms Transforming Healthcare with Big Data and AI-Mingbo Gong 2020-04-01 Healthcare and technology are at a convergence point where significant changes are poised to take place. The vast and complex requirements of medical record keeping, coupled with stringent patient privacy laws, create an incredibly unwieldy maze of health data needs. While the past decade has seen giant leaps in AI, machine learning, wearable technologies, and data mining capacities that have enabled quantities of data to be accumulated, processed, and shared around the globe. Transforming Healthcare with Big Data and AI examines the crossroads of these two fields and looks to the future of leveraging advanced technologies and developing data ecosystems to the healthcare field. This book is the product of the Transforming Healthcare with Data conference, held at the University of Southern California. Many speakers and digital healthcare industry leaders contributed multidisciplinary expertise to chapters in this book. Authors’ backgrounds range from data scientists, healthcare experts, university professors, and digital healthcare entrepreneurs. If you have an understanding of data technologies and are interested in the future of Big Data and AI, in healthcare, this book will provide a wealth of insights into the new landscape of healthcare.