
Frank Kanes Taming Big Data With Apache Spark And Python

PySpark Cookbook
Big Data Analysis with Python
Posthuman Glossary
Stream Processing with Apache Spark
Spark in Action
Perspectives on Time
Data Pipelines with Apache Airflow
Spark for Python Developers
Building Recommender Systems with Machine Learning and AI.
Thinking, Recording, and Writing History in the Ancient World
The Mandarins
Saints, Scholars, and Schizophrenics
The Art of Watching Films
The Flagellants
Learning PySpark
Scala Programming for Big Data Analytics
The Neurobiology of Painting
Hands-On Data Science and Python Machine Learning
A to Z of Sport
Frank Kane's Taming Big Data with Apache Spark and Python
Voices of Early Modern Japan
PolyBase Revealed
A Collection of Data Science Interview Questions Solved in Python and Spark
Applied Data Science Using PySpark
Learning Spark
Dictionary of German and English, English and German
Dub
Critical Survey of Mystery and Detective Fiction
Data Analytics with Spark Using Python
Deyo (Deyoe) Family
A Furrow Laid Bare
Spark: The Definitive Guide
Sexual Consent
Le Deuxième Sexe
PySpark Recipes
Unix Shell Programming
Buyology
Learning Spark
Cultural Anthropology
The Independent Woman

Frank Kanes Taming Big Data With Apache Spark And Python

Downloaded from archive.imba.com by guest

WILSON JULIAN

PySpark Cookbook Wesleyan University Press

Frank Kane's hands-on Spark training course, based on his bestselling Taming Big Data with Apache Spark and Python video, 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 with Spark Who This Book Is For If you are a data scientist or data analyst who wants to learn Big Data processing using Apache Spark and Python, this book is for you. 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 as 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 quickly 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-world examples carefully explained by Frank in a step-by-step manner. The examples vary in complexity, and you can move through them at your own pace.

Big Data Analysis with Python Simon and Schuster

Learn how to use, deploy, and maintain Apache Spark with this comprehensive guide, written by the creators of the open-source cluster-computing framework. With an emphasis on improvements and new features in Spark 2.0, authors Bill Chambers and Matei Zaharia break down Spark topics into distinct sections, each with unique goals. You'll explore the basic operations and common functions of Spark's structured APIs, as well as Structured Streaming, a new high-level API for building end-to-end streaming applications. Developers and system administrators will learn the fundamentals of monitoring, tuning, and debugging Spark, and explore machine learning techniques and scenarios for employing MLlib, Spark's

scalable machine-learning library. Get a gentle overview of big data and Spark Learn about DataFrames, SQL, and Datasets Spark's core APIs through worked examples Dive into Spark's low-level APIs, RDDs, and execution of SQL and DataFrames Understand how Spark runs on a cluster Debug, monitor, and tune Spark clusters and applications Learn the power of Structured Streaming, Spark's stream-processing engine Learn how you can apply MLlib to a variety of problems, including classification or recommendation

Posthuman Glossary Elsevier

THE A TO Z OF SPORT is unique: nothing as comprehensive has ever appeared before. It covers more than one hundred sports, with the treatment of some of the nation's favourites, such as football, rugby, cricket and athletics, almost meriting books in themselves. The sports covered include everything you can think of: from football to fencing, cricket to croquet, motor racing to marbles, all the way from adventure racing to wrestling - no sport is too big or small for inclusion. There is an introductory essay for each sport, explaining its history and rules. These are followed by the most comprehensive lists of tournaments, champions, cups, venues and participants ever assembled. And there are further sections on Sportspeople, the Television Sports Personality of the Year, Trophies, Sporting Quotations and Sporting Current Affairs. From remarkable facts to argument-settling information, the A TO Z OF SPORT is the indispensable reference book for every sports fan.

Stream Processing with Apache Spark Bloomsbury Publishing

Combine the power of Apache Spark and Python to build effective big data applications Key Features Perform effective data processing, machine learning, and analytics using PySpark Overcome challenges in developing and deploying Spark solutions using Python Explore recipes for efficiently combining Python and Apache Spark to process data Book Description Apache Spark is an open source framework for efficient cluster computing with a strong interface for data parallelism and fault tolerance. The PySpark Cookbook presents effective and time-saving recipes for leveraging the power of Python and putting it to use in the Spark ecosystem. You'll start by learning the Apache Spark architecture and how to set up a Python environment for Spark. You'll then get familiar with the modules available in PySpark and start using them effortlessly. In addition to this, you'll discover how to abstract data with RDDs and DataFrames, and understand the streaming capabilities of PySpark. You'll then move on to using ML and MLlib in order to solve any problems related to the machine learning capabilities of PySpark and use GraphFrames to solve graph-processing problems. Finally, you will explore how to deploy your applications to the cloud using the spark-submit command. By the end of this book, you will be able to use the Python API for Apache Spark to solve any problems associated with building data-intensive applications. What you will learn Configure a local instance of PySpark in a virtual environment Install and configure Jupyter in local and multi-node environments Create DataFrames from JSON and a dictionary using pyspark.sql Explore regression and clustering models available in the ML module Use DataFrames to transform data used for modeling Connect to PubNub and perform aggregations on streams Who this book is for The PySpark Cookbook is for you if you are a Python developer looking for hands-on recipes for using the Apache Spark 2.x ecosystem in the best possible way. A thorough understanding of Python (and some familiarity with Spark) will help you get the best out of the book.

Spark in Action Packt Publishing Ltd

Discover the capabilities of PySpark and its application in the realm of data science. This comprehensive guide with hand-picked examples of daily use cases will walk you through the end-to-end predictive model-building cycle with the latest techniques and tricks of the trade. Applied Data Science Using PySpark is divided into six sections which walk you through the book. In section 1, you start with the basics of PySpark focusing on data manipulation. We make you comfortable with the language and then build upon it to introduce you to the mathematical functions available off the shelf. In section 2, you will dive into the art of variable selection where we demonstrate various selection techniques available in PySpark. In section 3, we take you on a journey through machine learning algorithms, implementations, and fine-tuning techniques. We will also talk about different validation metrics and how to use them for picking the best models. Sections 4 and 5 go through machine learning pipelines and various methods available to operationalize the model and serve it through Docker/an API. In the final section, you will cover reusable objects for easy experimentation and learn some tricks that can help you optimize your programs and machine learning pipelines. By the end of this book, you will have seen the flexibility and advantages of PySpark in data science applications. This book is recommended to those who want to unleash the power of parallel computing by simultaneously working with big datasets. What You Will Learn Build an end-to-end predictive model Implement multiple variable selection techniques Operationalize models Master multiple algorithms and implementations Who This Book is For Data scientists and machine learning and deep learning engineers who want to learn and use PySpark for real-time analysis of streaming data.

Perspectives on Time CreateSpace

Thinking, Recording, and Writing History in the Ancient World presents a cross-cultural comparison of the ways in which ancient civilizations thought about the past and recorded their own histories. Written by an international group of scholars working in many disciplines Truly cross-cultural, covering historical thinking and writing in ancient or early cultures across in East, South, and West Asia, the Mediterranean, and the Americas Includes historiography shaped by religious perspectives, including Judaism, early Christianity, Islam, and Buddhism

Data Pipelines with Apache Airflow "O'Reilly Media, Inc."

If art, science, and the humanities have shared one thing, it was their common engagement with constructions and representations of the human. Under the pressure of new contemporary concerns, however, we are experiencing a "posthuman condition"; the combination of new developments—such as the neoliberal economics of global capitalism, migration, technological advances, environmental destruction on a mass scale, the perpetual war on terror and extensive security systems— with a troublesome reiteration of old, unresolved problems that mean the concept of the human as we had previously known it has undergone dramatic transformations. The Posthuman Glossary is a volume providing an outline of the critical terms of posthumanity in present-day artistic and intellectual work. It builds on the broad thematic topics of Anthropocene/Capitalocene, eco-sophies, digital activism, algorithmic cultures and security and the inhuman. It outlines potential artistic, intellectual, and activist itineraries of working through the complex reality of the 'posthuman condition', and creates an understanding of the altered meanings of art vis-à-vis critical present-day developments. It bridges missing links across disciplines, terminologies, constituencies and critical communities. This original work will unlock the terms of the posthuman for students and researchers alike.

Spark for Python Developers Springer Science & Business Media

An introduction to issues of sexual consent, covering key strands of feminist thought, how sexual consent is negotiated in practice, the influence of popular culture, and more. The #MeToo movement has focused public attention on the issue of sexual consent. People of all genders, from all walks of life, have stepped forward to tell their stories of sexual harassment and violation. In a predictable backlash, others have taken to mass media to inquire plaintively if "flirting" is now forbidden. This volume in the MIT Press Essential Knowledge series offers a nuanced introduction to sexual consent by a writer who is both a scholar and an activist on this issue. It has become clear from discussions of the recent high-profile cases of Harvey Weinstein, Bill Cosby, and others that there is no clear agreement over what constitutes consent or non-consent and how they are expressed and perceived in sexual situations. This book presents key strands of feminist thought on the subject of sexual consent from across academic and activist communities and covers the history of research on consent in such fields as psychology and feminist legal studies. It discusses how sexual consent is negotiated in practice, from "No means no" to "Yes means yes," and describes what factors might limit individual agency in such negotiations. It examines how popular culture, including pornography, romance fiction, and sex advice manuals, shapes our ideas of consent; explores the communities at the forefront of consent activism; and considers what meaningful social change in this area might look like. Going beyond the conventional cisgender, heterosexual norm, the book lists additional resources for those seeking to improve their practice of consent, survivors of sexual violence, and readers who want to understand contemporary debates on this issue in more depth.

Building Recommender Systems with Machine Learning and AI. "O'Reilly Media, Inc."

Solve Data Analytics Problems with Spark, PySpark, and Related Open Source Tools Spark is at the heart of today's Big Data revolution, helping data professionals supercharge efficiency and performance in a wide range of data processing and analytics tasks. In this guide, Big Data expert Jeffrey Aven covers all you need to know to leverage Spark, together with its extensions, subprojects, and wider ecosystem. Aven combines a language-agnostic introduction to foundational Spark concepts with extensive programming examples utilizing the popular and intuitive PySpark development environment. This guide's focus on Python makes it widely accessible to large audiences of data professionals, analysts, and developers—even those with little Hadoop or Spark experience. Aven's broad coverage ranges from basic to advanced Spark programming, and Spark SQL to machine learning. You'll learn how to efficiently manage all forms of data with Spark: streaming, structured, semi-structured, and unstructured. Throughout, concise topic overviews quickly get you up to speed, and extensive hands-on exercises prepare you to solve real problems. Coverage includes: • Understand Spark's evolving role in the Big Data and Hadoop ecosystems • Create Spark clusters using various deployment modes • Control and optimize the operation of Spark clusters and applications • Master Spark Core RDD API programming techniques • Extend, accelerate, and optimize Spark routines with advanced API platform constructs, including shared variables, RDD storage, and partitioning • Efficiently integrate Spark with both SQL and nonrelational data stores • Perform stream processing and messaging with Spark Streaming and Apache Kafka • Implement predictive modeling with SparkR and Spark MLlib

Thinking, Recording, and Writing History in the Ancient World Univ of California Press

The book presents a basis for the interaction of the brain and nervous system with painting, music and literature, and a discussion of art from multiple facets - such as anatomy, migraine, illusion and evolutionary biology. The book explores several aspects of the neurobiology of painting, including evolutionary neurobiology, sensation vs. perception, the visual brain and how the mind works, and also explores the affects of brain disorders and trauma on artist, with a concluding chapter on Frida Kahlo and the spinal cord injury that influenced her painting.

The Mandarins Neerlandia, Alta. : Inheritance Publications

Summary The Spark distributed data processing platform provides an easy-to-implement tool for ingesting, streaming, and processing data from any source. In Spark in Action, Second Edition, you'll learn to take advantage of Spark's core features and incredible processing speed, with applications including real-time computation, delayed evaluation, and machine learning. Spark skills are a hot commodity in enterprises worldwide, and with Spark's powerful and flexible Java APIs, you can reap all the benefits without first learning Scala or Hadoop. Foreword by Rob Thomas. About the technology Analyzing enterprise data starts by reading, filtering, and merging files and streams from many sources. The Spark data processing engine handles this varied volume like a champ, delivering speeds 100 times faster than Hadoop systems. Thanks to SQL support, an intuitive interface, and a straightforward multilanguage API, you can use Spark without learning a complex new ecosystem. About the book Spark in Action, Second Edition, teaches you to create end-to-end analytics applications. In this entirely new book, you'll learn from interesting Java-based examples, including a complete data pipeline for processing NASA satellite data. And you'll discover Java, Python, and Scala code samples hosted on GitHub that you can explore and adapt, plus appendixes that give you a cheat sheet for installing tools and understanding Spark-specific terms. What's inside Writing Spark applications in Java Spark application architecture Ingestion through files, databases, streaming, and Elasticsearch Querying distributed datasets with Spark SQL About the reader This book does not assume previous experience with Spark, Scala, or Hadoop. About the author Jean-Georges Perrin is an experienced data and software architect. He is France's first IBM Champion and has been honored for 12 consecutive years. Table of Contents PART 1 - THE THEORY CRIPPLED BY AWESOME EXAMPLES 1 So, what is Spark, anyway? 2 Architecture and flow 3 The majestic role of the dataframe 4 Fundamentally lazy 5 Building a simple app for deployment 6 Deploying your simple app PART 2 - INGESTION 7 Ingestion from files 8 Ingestion from databases 9 Advanced ingestion: finding data sources and building your own 10 Ingestion through structured streaming PART 3 - TRANSFORMING YOUR DATA 11 Working with SQL 12 Transforming your data 13 Transforming entire documents 14 Extending transformations with user-defined functions 15 Aggregating your data PART 4 - GOING FURTHER 16 Cache and checkpoint: Enhancing Spark's performances 17 Exporting data and building full data pipelines 18 Exploring deployment

Saints, Scholars, and Schizophrenics Routledge

ALERT: Before you purchase, check with your instructor or review your course syllabus to ensure that you select the correct ISBN. Several versions of Pearson's MyLab & Mastering products exist for each title, including customized versions for individual schools, and registrations are not transferable. In addition, you may need a CourseID, provided by your instructor, to register for and use Pearson's MyLab & Mastering products. Packages Access codes for Pearson's MyLab & Mastering products may not be included when purchasing or renting from companies other than Pearson; check with the

seller before completing your purchase. Used or rental books If you rent or purchase a used book with an access code, the access code may have been redeemed previously and you may have to purchase a new access code. Access codes Access codes that are purchased from sellers other than Pearson carry a higher risk of being either the wrong ISBN or a previously redeemed code. Check with the seller prior to purchase. -- Anthropology in today's world. Through clear writing, a balanced theoretical approach, and engaging examples, Cultural Anthropology stresses the importance of social inequality and human rights, the environment, culture change and applied aspects of anthropology. Rich examples of gender, ethnicity, race, class, and age thread through the topical coverage of economic systems, the life-cycle, health, kinship, social organization, politics, language, religion, and expressive culture. In addition, the last two chapters address how migration is changing world cultures and how the importance of local cultural values and needs are shaping international development policies and programs. Note: MyAnthroLab does not come automatically packaged with this text. To purchase MyAnthroLab, please visit: www.myanthrolab.com or you can purchase a valuepack of the text + MyAnthroLab (at no additional cost): ValuePack ISBN-10: 0205949509 / ValuePack ISBN-13: 9780205949502

The Art of Watching Films Little Brown Uk

A concise guide to implementing Spark Big Data analytics for Python developers, and building a real-time and insightful trend tracker data intensive app About This Book • Set up real-time streaming and batch data intensive infrastructure using Spark and Python • Deliver insightful visualizations in a web app using Spark (PySpark) • Inject live data using Spark Streaming with real-time events Who This Book Is For This book is for data scientists and software developers with a focus on Python who want to work with the Spark engine, and it will also benefit Enterprise Architects. All you need to have is a good background of Python and an inclination to work with Spark. What You Will Learn • Create a Python development environment powered by Spark (PySpark), Blaze, and Bokeh • Build a real-time trend tracker data intensive app • Visualize the trends and insights gained from data using Bokeh • Generate insights from data using machine learning through Spark MLLIB • Juggle with data using Blaze • Create training data sets and train the Machine Learning models • Test the machine learning models on test datasets • Deploy the machine learning algorithms and models and scale it for real-time events In Detail Looking for a cluster computing system that provides high-level APIs? Apache Spark is your answer—an open source, fast, and general purpose cluster computing system. Spark's multi-stage memory primitives provide performance up to 100 times faster than Hadoop, and it is also well-suited for machine learning algorithms. Are you a Python developer inclined to work with Spark engine? If so, this book will be your companion as you create data-intensive app using Spark as a processing engine, Python visualization libraries, and web frameworks such as Flask. To begin with, you will learn the most effective way to install the Python development environment powered by Spark, Blaze, and Bokeh. You will then find out how to connect with data stores such as MySQL, MongoDB, Cassandra, and Hadoop. You'll expand your skills throughout, getting familiarized with the various data sources (Github, Twitter, Meetup, and Blogs), their data structures, and solutions to effectively tackle complexities. You'll explore datasets using iPython Notebook and will discover how to optimize the data models and pipeline. Finally, you'll get to know how to create training datasets and train the machine learning models. By the end of the book, you will have created a real-time and insightful trend tracker data-intensive app with Spark. Style and approach This is a comprehensive guide packed with easy-to-follow examples that will take your skills to the next level and will get you up and running with Spark.

The Flagellants Macmillan

Get to grips with processing large volumes of data and presenting it as engaging, interactive insights using Spark and Python. Key Features Get a hands-on, fast-paced introduction to the Python data science stack Explore ways to create useful metrics and statistics from large datasets Create detailed analysis reports with real-world data Book Description Processing big data in real time is challenging due to scalability, information inconsistency, and fault tolerance. Big Data Analysis with Python teaches you how to use tools that can control this data avalanche for you. With this book, you'll learn practical techniques to aggregate data into useful dimensions for posterior analysis, extract statistical measurements, and transform datasets into features for other systems. The book begins with an introduction to data manipulation in Python using pandas. You'll then get familiar with statistical analysis and plotting techniques. With multiple hands-on activities in store, you'll be able to analyze data that is distributed on several computers by using Dask. As you progress, you'll study how to aggregate data for plots when the entire data cannot be accommodated in memory. You'll also explore Hadoop (HDFS and YARN), which will help you tackle larger datasets. The book also covers Spark and explains how it interacts with other tools. By the end of this book, you'll be able to bootstrap your own Python environment, process large files, and manipulate data to generate statistics, metrics, and graphs. What you will learn Use Python to read and transform data into different formats Generate basic statistics and metrics using data on disk Work with computing tasks distributed over a cluster Convert data from various sources into storage or querying formats Prepare data for statistical analysis, visualization, and machine learning Present data in the form of effective visuals Who this book is for Big Data Analysis with Python is designed for Python developers, data analysts, and data scientists who want to get hands-on with methods to control

data and transform it into impactful insights. Basic knowledge of statistical measurements and relational databases will help you to understand various concepts explained in this book.

Learning PySpark Apress

Accompanying CD-ROM provides short film clips that reinforce the key concepts and topics in each chapter.

Scala Programming for Big Data Analytics O'Reilly Media

Automated recommendations are everywhere: Netflix, Amazon, YouTube, and more. Recommender systems learn about your unique interests and show the products or content they think you'll like best. Discover how to build your own recommender systems from one of the pioneers in the field. Frank Kane spent over nine years at Amazon, where he led the development of many of the company's personalized product recommendation technologies. In this course, he covers recommendation algorithms based on neighborhood-based collaborative filtering and more modern techniques, including matrix factorization and even deep learning with artificial neural networks. Along the way, you can learn from Frank's extensive industry experience and understand the real-world challenges of applying these algorithms at a large scale with real-world data. You can also go hands-on, developing your own framework to test algorithms and building your own neural networks using technologies like Amazon DSSTNE, AWS SageMaker, and TensorFlow.

The Neurobiology of Painting O'Reilly Media

Quickly find solutions to common programming problems encountered while processing big data. Content is presented in the popular problem-solution format. Look up the programming problem that you want to solve. Read the solution. Apply the solution directly in your own code. Problem solved! PySpark Recipes covers Hadoop and its shortcomings. The architecture of Spark, PySpark, and RDD are presented. You will learn to apply RDD to solve day-to-day big data problems. Python and NumPy are included and make it easy for new learners of PySpark to understand and adopt the model. What You Will Learn Understand the advanced features of PySpark2 and SparkSQL Optimize your code Program SparkSQL with Python Use Spark Streaming and Spark MLlib with Python Perform graph analysis with GraphFrames Who This Book Is For Data analysts, Python programmers, big data enthusiasts

Hands-On Data Science and Python Machine Learning Currency

"The Flagellants is the story of the romantic relationship between Ideal and Jimson. After a brief prologue establishing Ideal's childhood connection to a black community called "the Bottom," the novel unfolds as a series of arguments between the couple, representing the historical gender conflicts between black men and women."--eNotes.

A to Z of Sport Vintage

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

Frank Kane's Taming Big Data with Apache Spark and Python Apress

The classic manifesto of the liberated woman, this book explores every facet of a woman's life.

Related with Frank Kanes Taming Big Data With Apache Spark And Python:

• F O G Math Examples : [click here](#)