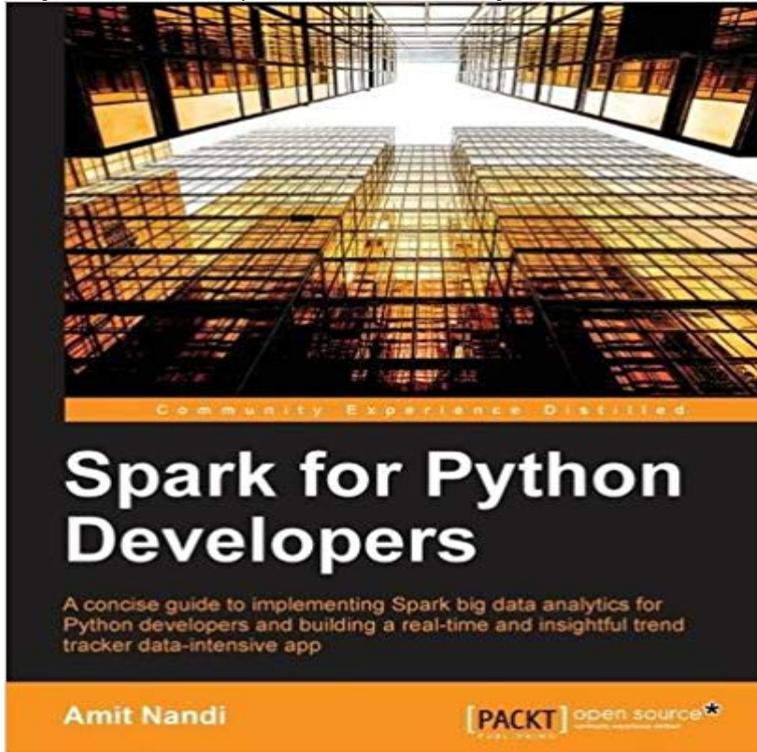


Spark for Python Developers



Key Features

- 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

Book Description

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. Sparks 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.

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

MLLIBJuggle with data using BlazeCreate training data sets and train the Machine Learning modelsTest the machine learning models on test datasetsDeploy the machine learning algorithms and models and scale it for real-time eventsAbout the AuthorAmit Nandi studied physics at the Free University of Brussels in Belgium, where he did his research on computer generated holograms. Computer generated holograms are the key components of an optical computer, which is powered by photons running at the speed of light. He then worked with the university Cray supercomputer, sending batch jobs of programs written in Fortran. This gave him a taste for computing, which kept growing. He has worked extensively on large business reengineering initiatives, using SAP as the main enabler. He focused for the last 15 years on start-ups in the data space, pioneering new areas of the information technology landscape. He is currently focusing on large-scale data-intensive applications as an enterprise architect, data engineer, and software developer. He understands and speaks seven human languages. Although Python is his computer language of choice, he aims to be able to write fluently in seven computer languages too.

Table of Contents

Setting Up a Spark Virtual Environment

Building Batch and Streaming Apps with Spark

Juggling Data with Spark

Learning from Data Using Spark

Streaming Live Data with Spark

Visualizing Insights and Trends

Introduction #. Apache Spark is an open source cluster computing framework originally developed in the AMPLab at University of California, Spark for Python Developers. A concise guide to implementing Spark big data analytics for Python developers and building a real-time and insightful trend tracker data intensive app

1536 Spark & Python Developer jobs available. See salaries, compare reviews, easily apply, and get hired. New Spark & Python Developer careers are added

Spark for Python Developers: A concise guide to implementing Spark big data analytics for Python developers and building a real-time and insightful trend tracker data intensive app

Read Spark for Python Developers by Amit Nandi with Rakuten Kobo. A concise guide to implementing Spark Big Data analytics for Python developers, and building a real-time and insightful trend tracker data intensive app

AboutNote 0.0/5. Retrouvez Spark for Python Developers: A concise guide to implementing Spark big data analytics for Python developers and building a real-time and insightful trend tracker data intensive app

One search API that essentially allows us

to retrieve past tweets based on search terms. This is how we have - Selection from Spark for Python Developers
Desired Experience: Pig, Hive, Microsoft SQL Server, Perl, JSON, , MySQL, Informatica, JIRA, Hadoop, Oracle, HBase, Tableau, Java, Spark, DataSpark for Python Developers. Are you a Python developer inclined to work with Spark engine? If so, this book will be your companion as you create
Search Python spark developer jobs. Get the right Python spark developer job with company ratings & salaries. 1283 open jobs for Python spark developer. 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