High Performance 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 describes techniques that can reduce data infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you’ll also learn how to make it sing. With this book, you’ll explore: How Spark SQL’s new interfaces improve performance over SQL’s RDD data structure The choice between data joins in Core Spark and Spark SQL Techniques for getting the most out of standard RDD transformations How to work around performance issues in Spark’s key/value pair paradigm Writing high-performance Spark code without Scala or the JVM How to test for functionality and performance when applying suggested improvements Using Spark MLlib and Spark ML community packages.
instead of mixing and matching tools like Hive, Hadoop, Mahout, and Storm community packages

**Learning Spark** - Holden Karau - 2015-01-28
Data in all domains is getting bigger. How can you work with it efficiently? Recently updated for Spark 1.3, this book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java, and Scala. This edition includes new information on Spark SQL, Spark Streaming, setup, and Maven coordinates. Written by the developers of Spark, this book will have data scientists and engineers up and running in n.
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 MLflow

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 describes techniques that can reduce data infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you’ll also learn how to make it sing. With this book, you’ll explore: How Spark SQL’s new interfaces improve performance over SQL’s RDD data structure The choice between data joins in Core Spark and Spark SQL Techniques for getting the most out of standard RDD transformations How to work around performance issues in Spark’s key/value pair paradigm Writing high-performance Spark code without Scala or the JVM How to test for functionality and performance when applying suggested improvements Using Spark MLlib and Spark ML machine learning libraries Spark’s Streaming components and external community packages
to teach you how to approach analytics problems by example. You’ll start infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you’ll also learn how to make it sing. With this book, you’ll explore: How Spark SQL’s new interfaces improve performance over SQL’s RDD data structure The choice between data joins in Core Spark and Spark SQL Techniques for getting the most out of standard RDD transformations How to work around performance issues in Spark’s key/value pair paradigm Writing high-performance Spark code without Scala or the JVM How to test for functionality and performance when applying suggested improvements Using Spark MLlib and Spark ML machine learning libraries Spark’s Streaming components and external community packages

**Advanced Analytics with Spark** - Sandy Ryza - 2015-04-02
In this practical book, four Cloudera data scientists present a set of self-contained patterns for performing large-scale data analysis with Spark. The authors bring Spark, statistical methods, and real-world data sets together to teach you how to approach analytics problems by example. You’ll start with an introduction to Spark and its ecosystem, and then dive into patterns that apply common techniques—classification, collaborative filtering, and anomaly detection among others—to fields such as genomics, security, and finance. If you have an entry-level understanding of machine learning and statistics, and you program in Java, Python, or Scala, you’ll find these patterns useful for working on your own data applications. Patterns include: Recommending music and the Audioscrobbler data set Predicting forest cover with decision trees Anomaly detection in network traffic with K-means clustering Understanding Wikipedia with Latent Semantic Analysis Analyzing co-occurrence networks with GraphX Geospatial and temporal data analysis on the New York City Taxi Trips data Estimating financial risk through Monte Carlo simulation Analyzing genomics data and the BDG project Analyzing neuroimaging data with PySpark and Thunder

**Data-intensive Text Processing with MapReduce** - Jimmy Lin - 2010
Our world is being revolutionized by data-driven methods: access to large amounts of data has generated new insights and opened exciting new opportunities in commerce, science, and computing applications. Processing the enormous quantities of data necessary for these advances requires large clusters, making distributed computing paradigms more crucial than ever. MapReduce is a programming model for expressing distributed computations on massive datasets and an execution framework for large-scale data processing on clusters of commodity servers. The programming model provides an easy-to-understand abstraction for designing scalable algorithms, while the execution framework transparently handles many system-level details, ranging from scheduling to synchronization to fault tolerance. This book focuses on MapReduce algorithm design, with an emphasis on text processing algorithms common in natural language processing, information retrieval, and machine learning. We introduce the notion of MapReduce design patterns, which represent general reusable solutions to commonly occurring problems across a variety of problem domains. This book not only intends to help the reader "think in MapReduce", but also discusses limitations of the programming model as well. This volume is a printed version of a work that appears in the Synthesis Digital Library of Engineering and Computer Science. Synthesis
Our world is being revolutionized by data-driven methods: access to large amounts of data has generated new insights and opened exciting new opportunities in commerce, science, and computing applications. Processing the enormous quantities of data necessary for these advances requires large clusters, making distributed computing paradigms more crucial than ever.

MapReduce is a programming model for expressing distributed computations on massive datasets and an execution framework for large-scale data processing on clusters of commodity servers. The programming model provides an easy-to-understand abstraction for designing scalable algorithms, while the execution framework transparently handles many system-level details, ranging from scheduling to synchronization to fault tolerance. This book focuses on MapReduce algorithm design, with an emphasis on text processing algorithms common in natural language processing, information retrieval, and machine learning. We introduce the notion of MapReduce design patterns, which represent general reusable solutions to commonly occurring problems across a variety of problem domains. This book not only intends to help the reader "think in MapReduce", but also discusses limitations of the programming model as well. This volume is a printed version of a work that appears in the Synthesis Digital Library of Engineering and Computer Science. Synthesis Lectures provide concise, original presentations of important research and development topics, published quickly, in digital and print formats. For more information visit www.morganclaypool.com

This book covers a verity of topics, including in-memory data grid, highly available service grid, streaming (event processing for IoT and fast data) and in-memory computing use cases from high-performance computing to get performance gains. The book will be particularly useful for those, who have the following use cases: 1) You have a high volume of ACID transactions in your system. 2) You have database bottleneck in your application and want to solve the problem. 3) You want to develop and deploy Microservices in a distributed fashion. 4) You have an existing Hadoop ecosystem (OLAP) and want to improve the performance of map/reduce jobs without making any changes in your existing map/reduce jobs. 5) You want to share Spark RDD directly in-memory (without storing the state into the disk) 7) You are planning to process continuous never-ending streams and complex events of data. 8) You want to use distributed computations in parallel fashion to gain high performance.
Burchard found the answers. It turns out that just six deliberate habits give you the edge. Anyone can practice these habits and, when they do, extraordinary things happen in their lives, relationships, and careers. Which habits can help you achieve long-term success and vibrant well-being no matter your age, career, strengths, or personality? To become a high performer, you must seek clarity, generate energy, raise necessity, increase productivity, develop influence, and demonstrate courage. The art and science of how to do all this is what this book is about. Whether you want to get more done, lead others better, develop skill faster, or dramatically increase your sense of joy and confidence, the habits in this book will help you achieve it faster. Each of the six habits is illustrated by powerful vignettes, cutting-edge science, thought-provoking exercises, and real-world daily practices you can implement right now. If you’ve ever wanted a science-backed, heart-centered plan to living a better quality of life, it’s in your hands. Best of all, you can measure your progress. A link to a free professional assessment is included in the book.

**High Performance Habits** - Brendon Burchard - 2017-09-19

THESE HABITS WILL MAKE YOU EXTRAORDINARY. Twenty years ago, author Brendon Burchard became obsessed with answering three questions: 1. Why do some individuals and teams succeed more quickly than others and sustain that success over the long term? 2. Of those who pull it off, why are some miserable and others consistently happy on their journey? 3. What motivates people to reach for higher levels of success in the first place, and what practices help them improve the most After extensive original research and a decade as the world’s leading high performance coach, Burchard found the answers. It turns out that just six deliberate habits give you the edge. Anyone can practice these habits and, when they do, extraordinary things happen in their lives, relationships, and careers. Which habits can help you achieve long-term success and vibrant well-being no matter your age, career, strengths, or personality? To become a high performer, you must seek clarity, generate energy, raise necessity, increase productivity, develop influence, and demonstrate courage. The art and science of how to do all this is what this book is about. Whether you want to get more done, lead others better, develop skill faster, or dramatically increase your sense of joy and confidence, the habits in this book will help you achieve it faster. Each of the six habits is illustrated by powerful vignettes, cutting-edge science, thought-provoking exercises, and real-world daily practices you can implement right now. If you’ve ever wanted a science-backed, heart-centered plan to living a better quality of life, it’s in your hands. Best of all, you can measure your progress. A link to a free professional assessment is included in the book.

**Spark in Action** - Jean-Georges Perrin - 2020-05-12

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. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. 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
datasets with Spark SQL About the reader This book does not assume
majestic role of the dataframe 4 Fundamentally lazy 5 Building a simple app
dispatch 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

Spark in Action - Jean-Georges Perrin - 2020-05-12
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. Purchase of the print book includes a free eBook
in PDF, Kindle, and ePub formats from Manning Publications. 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
totally 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 appendices 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

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
dispatch 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

Kubeflow for Machine Learning - Trevor Grant - 2020-10-13
If you’re training a machine learning model but aren’t sure how to put it into
production, this book will get you there. Kubeflow provides a collection of
cloud native tools for different stages of a model’s lifecycle, from data
exploration, feature preparation, and model training to model serving. This
guide helps data scientists build production-grade machine learning
implementations with Kubeflow and shows data engineers how to make
models scalable and reliable. Using examples throughout the book, authors
Holden Karau, Trevor Grant, Ilan Filonenko, Richard Liu, and Boris
Lublinsky explain how to use Kubeflow to train and serve your machine
learning models on top of Kubernetes in the cloud or in a development
environment on-premises. Understand Kubeflow’s design, core components,
and the problems it solves Understand the differences between Kubeflow on
different cluster types Train models using Kubeflow with popular tools
including Scikit-learn, TensorFlow, and Apache Spark Keep your model up
to date with Kubeflow Pipelines Understand how to capture model training
metadata Explore how to extend Kubeflow with additional open source tools
Use hyperparameter tuning for training Learn how to serve your model in
production

high-performance-spark-best-practices-for-scaling-and-optimizing-apache-spark
high-performance-spark-best-practices-for-scaling-and-optimizing-apache-spark

its installation, programming and execution; Reviews the basics of Spark, 
**Kubeflow for Machine Learning** - Trevor Grant - 2020-10-13
If you’re training a machine learning model but aren’t sure how to put it into production, this book will get you there. Kubeflow provides a collection of cloud native tools for different stages of a model’s lifecycle, from data exploration, feature preparation, and model training to model serving. This guide helps data scientists build production-grade machine learning implementations with Kubeflow and shows data engineers how to make models scalable and reliable. Using examples throughout the book, authors Holden Karau, Trevor Grant, Ilan Filonenko, Richard Liu, and Boris Lublinsky explain how to use Kubeflow to train and serve your machine learning models on top of Kubernetes in the cloud or in a development environment on-premises. Understand Kubeflow's design, core components, and the problems it solves Understand the differences between Kubeflow on different cluster types Train models using Kubeflow with popular tools including Scikit-learn, TensorFlow, and Apache Spark Keep your model up to date with Kubeflow Pipelines Understand how to capture model training metadata Explore how to extend Kubeflow with additional open source tools Use hyperparameter tuning for training Learn how to serve your model in production

**Big Data Processing with Apache Spark** - Srini Penchikala -

**Big Data Processing with Apache Spark** - Srini Penchikala -

This timely text/reference describes the development and implementation of large-scale distributed processing systems using open source tools and technologies. Comprehensive in scope, the book presents state-of-the-art material on building high performance distributed computing systems, providing practical guidance and best practices as well as describing theoretical software frameworks. Features: describes the fundamentals of building scalable software systems for large-scale data processing in the new paradigm of high performance distributed computing; presents an overview of the Hadoop ecosystem, followed by step-by-step instruction on its installation, programming and execution; Reviews the basics of Spark, including resilient distributed datasets, and examines Hadoop streaming and working with Scalding; Provides detailed case studies on approaches to clustering, data classification and regression analysis; Explains the process of creating a working recommender system using Scalding and Spark.

**Guide to High Performance Distributed Computing** - K.G. Srinivasa -

**Guide to High Performance Distributed Computing** - K.G. Srinivasa -

**Guide to High Performance Distributed Computing** - K.G. Srinivasa -

**The Challenger Sale** - Matthew Dixon - 2011-11-10
What’s the secret to sales success? If you’re like most business leaders, you’d say it’s fundamentally about relationships-and you’d be wrong. The best salespeople don’t just build relationships with customers. They challenge them. The need to understand what top-performing reps are doing that their average performing colleagues are not drove Matthew Dixon, Brent Adamson, and their colleagues at Corporate Executive Board to investigate the skills, behaviors, knowledge, and attitudes that matter most for high performance. And what they discovered may be the biggest shock to conventional sales wisdom in decades. Based on an exhaustive study of thousands of sales reps across multiple industries and geographies, The Challenger Sale argues that classic relationship building is a losing approach, especially when it comes to selling complex, large-scale business-
Challengers approach customers with unique insights about how they can save or make money. They tailor their sales message to the customer's specific needs and objectives. Rather than acquiescing to the customer's every demand or objection, they are assertive, pushing back when necessary and taking control of the sale. The things that make Challengers unique are replicable and teachable to the average sales rep. Once you understand how to identify the Challengers in your organization, you can model their approach and embed it throughout your sales force. The authors explain how almost any average-performing rep, once equipped with the right tools, can successfully reframe customers' expectations and deliver a distinctive purchase experience that drives higher levels of customer loyalty and, ultimately, greater growth.

**The Challenger Sale** - Matthew Dixon - 2011-11-10
What’s the secret to sales success? If you’re like most business leaders, you’d say it’s fundamentally about relationships—and you’d be wrong. The best salespeople don’t just build relationships with customers. They challenge them. The need to understand what top-performing reps are doing that their average performing colleagues are not drove Matthew Dixon, Brent Adamson, and their colleagues at Corporate Executive Board to investigate the skills, behaviors, knowledge, and attitudes that matter most for high performance. And what they discovered may be the biggest shock to conventional sales wisdom in decades. Based on an exhaustive study of thousands of sales reps across multiple industries and geographies, The Challenger Sale argues that classic relationship building is a losing approach, especially when it comes to selling complex, large-scale business-to-business solutions. The authors’ study found that every sales rep in the world falls into one of five distinct profiles, and while all of these types of reps can deliver average sales performance, only one-the Challenger-delivers consistently high performance. Instead of bludgeoning customers with endless facts and features about their company and products,
installing tools and understanding Spark-specific terms. What's inside streaming jobs in almost the same way you write batch jobs. Authors Gerard Maas and François Garillot help you explore the theoretical underpinnings of Apache Spark. This comprehensive guide features two sections that compare and contrast the streaming APIs Spark now supports: the original Spark Streaming library and the newer Structured Streaming API. Learn fundamental stream processing concepts and examine different streaming architectures Explore Structured Streaming through practical examples; learn different aspects of stream processing in detail Create and operate streaming jobs and applications with Spark Streaming; integrate Spark Streaming with other Spark APIs Learn advanced Spark Streaming techniques, including approximation algorithms and machine learning algorithms Compare Apache Spark to other stream processing projects, including Apache Storm, Apache Flink, and Apache Kafka Streams

Spark in Action, Second Edition - Jean-Georges Perrin - 2020-06-02
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. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. 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

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 performance 17 Exporting data and building full data pipelines 18 Exploring deployment
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

Practical Apache Spark - Subhashini Chellappan - 2019-01-02
Work with Apache Spark using Scala to deploy and set up single-node, multi-node, and high-availability clusters. This book discusses various components of Spark such as Spark Core, DataFrames, Datasets and SQL, Spark Streaming, Spark MLib, and R on Spark with the help of practical code snippets for each topic. Practical Apache Spark also covers the integration of Apache Spark with Kafka with examples. You’ll follow a learn-to-do-by-yourself approach to learning – learn the concepts, practice the code snippets in Scala, and complete the assignments given to get an overall exposure. On completion, you’ll have knowledge of the functional programming aspects of Scala, and hands-on expertise in various Spark components. You’ll also become familiar with machine learning algorithms with real-time usage. What You Will Learn
Discover the functional programming features of Scala
Understand the complete architecture of Spark and its components
Integrate Apache Spark with Hive and Kafka
Use Spark SQL, DataFrames, and Datasets to process data using traditional SQL queries
Work with different machine learning concepts and libraries using Spark’s MLib packages
Who This Book Is For
Developers and professionals who deal with batch and stream data processing.

Raise Your Game - Alan Stein - 2019-01-08
Performance coach Alan Stein Jr. shares the secret principles used by world-class performers that will help you improve your productivity and achieve higher levels of success. High achievers are at the top of their game because of the discipline they have during the unseen hours. They have made a commitment to establish, tweak, and repeat positive habits in everything they do. RAISE YOUR GAME examines the top leaders in sports
Performance coach Alan Stein Jr. shares the secret principles used by world-class performers that will help you improve your productivity and achieve higher levels of success. High achievers are at the top of their game because of the discipline they have during the unseen hours. They have made a commitment to establish, tweak, and repeat positive habits in everything they do. RAISE YOUR GAME examines the top leaders in sports and business and proves that success is a result of the little things we do all the time. The basic principles provided in RAISE YOUR GAME are simple, but not easy. We live in an instantly downloadable world that encourages us to skip steps. We are taught to chase what's hot, flashy and sexy and ignore what's basic. But the basics work. They always have and they always will. RAISE YOUR GAME will inspire and empower you to commit to the fundamentals, create a winning mindset, and progress into new levels of success.

Beginning Apache Spark 2 - Hien Luu - 2018-08-16
Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud technologies. Beginning Apache Spark 2 gives you an introduction to Apache Spark and shows you how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applications with Spark Structured Streaming. Furthermore, you'll learn the fundamentals of Spark ML for machine learning and much more. After you read this book, you will have the fundamentals to become proficient in using Apache Spark and know when and how to apply it to your big data applications. What You Will Learn Understand Spark unified data processing platform How to run Spark in Spark Shell or Databricks Use and manipulate RDDs Deal with structured data using Spark SQL through its operations and advanced functions Build real-time applications using Spark Structured Streaming Develop intelligent applications with the Spark Machine Learning library Who This Book Is For Programmers and developers active in big data, Hadoop, and Java but who are new to the Apache Spark platform.

The Best Team Wins - Adrian Gostick - 2018-02-13
The New York Times bestselling authors of The Carrot Principle and All In deliver a breakthrough, groundbreaking guide for building today's most collaborative teams—so any organization can operate at peak performance. A massive shift is taking place in the business world. In today's average company, up to eighty percent of employees' days are now spent working in teams. And yet the teams most people find themselves in are nowhere near effective. RAISE YOUR GAME will inspire and empower you to commit to the fundamentals, create a winning mindset, and progress into new levels of success.

Beginning Apache Spark 2 - Hien Luu - 2018-08-16
Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud technologies. Beginning Apache Spark 2 gives you an introduction to Apache Spark and shows you how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applications with Spark Structured Streaming. Furthermore, you'll learn the fundamentals of Spark ML for machine learning and much more. After you read this book, you will have the fundamentals to become proficient in using Apache Spark and know when and how to apply it to your big data applications. What You Will Learn Understand Spark unified data processing platform How to run Spark in Spark Shell or Databricks Use and manipulate RDDs Deal with structured data using Spark SQL through its operations and advanced functions Build real-time applications using Spark Structured Streaming Develop intelligent applications with the Spark Machine Learning library Who This Book Is For Programmers and developers active in big data, Hadoop, and Java but who are new to the Apache Spark platform.

The Best Team Wins - Adrian Gostick - 2018-02-13
The New York Times bestselling authors of The Carrot Principle and All In deliver a breakthrough, groundbreaking guide for building today's most collaborative teams—so any organization can operate at peak performance. A massive shift is taking place in the business world. In today's average company, up to eighty percent of employees’ days are now spent working in teams. And yet the teams most people find themselves in are nowhere near effective. RAISE YOUR GAME will inspire and empower you to commit to the fundamentals, create a winning mindset, and progress into new levels of success.
engagement; ways to promote healthy discord and spark innovation; and outright dissension, and dysfunctional teams drain employees’ energy, enthusiasm, and creativity. Now Adrian Gostick and Chester Elton share the proven ways managers can build cohesive, productive teams, despite the distractions and challenges every business is facing. In The Best Team Wins, Gostick and Elton studied more than 850,000 employee engagement surveys to develop their “Five Disciplines of Team Leaders,” explaining how to recognize and motivate different generations to enhance individual engagement; ways to promote healthy discord and spark innovation; and techniques to unify customer focus and build bridges across functions, cultures, and distance. They’ve shared these disciplines with their corporate clients and have now distilled their breakthrough findings into a succinct, engaging guide for business leaders everywhere. Gostick and Elton offer practical ways to address the real challenges today’s managers are facing, such as the rise of the Millennials, the increasing speed of change, the growing number of global and virtual teams, and the friction created by working cross-functionally. This is a must-read for anyone looking to maximize performance at work, from two of the most successful corporate consultants of their generation, whom The New York Times called “creative and refreshing.”

The Best Team Wins - Adrian Gostick - 2018-02-13
The New York Times bestselling authors of The Carrot Principle and All In deliver a breakthrough, groundbreaking guide for building today’s most collaborative teams—so any organization can operate at peak performance. A massive shift is taking place in the business world. In today’s average company, up to eighty percent of employees’ days are now spent working in teams. And yet the teams most people find themselves in are nowhere near as effective as they could be. They’re often divided by tensions, if not outright dissension, and dysfunctional teams drain employees’ energy, enthusiasm, and creativity. Now Adrian Gostick and Chester Elton share the proven ways managers can build cohesive, productive teams, despite the distractions and challenges every business is facing. In The Best Team Wins, Gostick and Elton studied more than 850,000 employee engagement surveys to develop their “Five Disciplines of Team Leaders,” explaining how to recognize and motivate different generations to enhance individual engagement; ways to promote healthy discord and spark innovation; and techniques to unify customer focus and build bridges across functions, cultures, and distance. They’ve shared these disciplines with their corporate clients and have now distilled their breakthrough findings into a succinct, engaging guide for business leaders everywhere. Gostick and Elton offer practical ways to address the real challenges today’s managers are facing, such as the rise of the Millennials, the increasing speed of change, the growing number of global and virtual teams, and the friction created by working cross-functionally. This is a must-read for anyone looking to maximize performance at work, from two of the most successful corporate consultants of their generation, whom The New York Times called “creative and refreshing.”

Guide to Spark Partitioning - Naushad Ahamad - 2020-10-14
Partitioning is one of the basic building blocks on which the Spark framework has been built. Partitioning at various stages of your program plays a very important role in ensuring the reliability, scalability and efficiency of the programs. In fact, just setting the right partitioning across various stages, lot of spark programs can be optimized right away. This book would assist you to understand the various aspects of Spark Partitioning in depth. Armed with the knowledge gained from the book, you would be able to set right partitioning in your Spark Jobs for large Datasets.

Guide to Spark Partitioning - Naushad Ahamad - 2020-10-14
Partitioning is one of the basic building blocks on which the Spark framework has been built. Partitioning at various stages of your program plays a very important role in ensuring the reliability, scalability and efficiency of the programs. In fact, just setting the right partitioning across various stages, lot of spark programs can be optimized right away. This book would assist you to understand the various aspects of Spark Partitioning in depth. Armed with the knowledge gained from the book, you would be able to set right partitioning in your Spark Jobs for large Datasets.

Getting Started with Kudu - Jean-Marc Spaggiari - 2018-07-09
Fast data ingestion, serving, and analytics in the Hadoop ecosystem have forced developers and architects to choose solutions using the least
analytics, predictive modeling, and in combination with another storage ingestion or fast data ingestion at the cost of slow analytics. There is an answer to this problem. With the Apache Kudu column-oriented data store, you can easily perform fast analytics on fast data. This practical guide shows you how. Begun as an internal project at Cloudera, Kudu is an open source solution compatible with many data processing frameworks in the Hadoop environment. In this book, current and former solutions professionals from Cloudera provide use cases, examples, best practices, and sample code to help you get up to speed with Kudu. Explore Kudu’s high-level design, including how it spreads data across servers. Fully administer a Kudu cluster, enable security, and add or remove nodes. Learn Kudu’s client-side APIs, including how to integrate Apache Impala, Spark, and other frameworks for data manipulation. Examine Kudu’s schema design, including basic concepts and primitives necessary to make your project successful. Explore case studies for using Kudu for real-time IoT analytics, predictive modeling, and in combination with another storage engine.

**Getting Started with Kudu** - Jean-Marc Spaggiari - 2018-07-09
Fast data ingestion, serving, and analytics in the Hadoop ecosystem have forced developers and architects to choose solutions using the least common denominator—either fast analytics at the cost of slow data ingestion or fast data ingestion at the cost of slow analytics. There is an answer to this problem. With the Apache Kudu column-oriented data store, you can easily perform fast analytics on fast data. This practical guide shows you how. Begun as an internal project at Cloudera, Kudu is an open source solution compatible with many data processing frameworks in the Hadoop environment. In this book, current and former solutions professionals from Cloudera provide use cases, examples, best practices, and sample code to help you get up to speed with Kudu. Explore Kudu’s high-level design, including how it spreads data across servers. Fully administer a Kudu cluster, enable security, and add or remove nodes. Learn Kudu’s client-side APIs, including how to integrate Apache Impala, Spark, and other frameworks for data manipulation. Examine Kudu’s schema design, including basic concepts and primitives necessary to make your project successful. Explore case studies for using Kudu for real-time IoT analytics, predictive modeling, and in combination with another storage engine.

**PostgreSQL Configuration** - Baji Shaik - 2020-03-16
Obtain all the skills you need to configure and manage a PostgreSQL database. In this book, you will begin by installing and configuring PostgreSQL on a server by focusing on system-level parameter settings before installation. You will also look at key post-installation steps to avoid issues in the future. The basic configuration of PostgreSQL is tuned for compatibility rather than performance. Keeping this in mind, you will fine-tune your PostgreSQL parameters based on your environment and application behavior. You will then get tips to improve database monitoring and maintenance followed by database security for handling sensitive data in PostgreSQL. Every system containing valuable data needs to be backed-up regularly. PostgreSQL follows a simple back-up procedure and provides fundamental approaches to back up your data. You will go through these approaches and choose the right one based on your environment. Running your application with limited resources can be tricky. To achieve this, you will implement a pooling mechanism for your PostgreSQL instances to connect to other databases. Finally, you will take a look at some basic errors faced while working with PostgreSQL and learn to resolve them in the quickest manner. What You Will Learn Configure PostgreSQL for performance. Monitor and maintain PostgreSQL instances. Implement a backup strategy for your data. Resolve errors faced while using PostgreSQL.

**Who This Book Is For** Readers with basic knowledge of PostgreSQL who wish to implement key solutions based on their environment.
trying to pick bits and pieces from different sources. The book also provides a chapter on Scala, the hottest functional programming language, and the program that underlies Spark. You'll learn the basics of functional programming in Scala, so that you can write Spark applications in it. What's more, Big Data Analytics with Spark provides an introduction to other big data technologies that are commonly used along with Spark, like Hive, Avro, Kafka and so on. So the book is self-sufficient; all the technologies that you need to know to use Spark are covered. The only thing that you are expected to know is programming in any language. There is a critical shortage of people with big data expertise, so companies are willing to pay top dollar for people with skills in areas like Spark and Scala. So reading this book and absorbing its principles will provide a boost—possibly a big boost—to your career.

Big Data Analytics with Spark - Mohammed Guller - 2015-12-29
Big Data Analytics with Spark is a step-by-step guide for learning Spark, which is an open-source fast and general-purpose cluster computing framework for large-scale data analysis. You will learn how to use Spark for different types of big data analytics projects, including batch, interactive, graph, and stream data analysis as well as machine learning. In addition, this book will help you become a much sought-after Spark expert. Spark is one of the hottest Big Data technologies. The amount of data generated today by devices, applications and users is exploding. Therefore, there is a critical need for tools that can analyze large-scale data and unlock value from it. Spark is a powerful technology that meets that need. You can, for example, use Spark to perform low latency computations through the use of efficient caching and iterative algorithms; leverage the features of its shell for easy and interactive Data analysis; employ its fast batch processing and low latency features to process your real time data streams and so on. As a result, adoption of Spark is rapidly growing and is replacing Hadoop MapReduce as the technology of choice for big data analytics. This book provides an introduction to Spark and related big-data technologies. It covers Spark core and its add-on libraries, including Spark SQL, Spark Streaming, GraphX, and MLlib. Big Data Analytics with Spark is therefore written for busy professionals who prefer learning a new technology from a consolidated source instead of spending countless hours on the Internet in PostgreSQL. Every system containing valuable data needs to be backed-up regularly. PostgreSQL follows a simple back-up procedure and provides fundamental approaches to back up your data. You will go through these approaches and choose the right one based on your environment. Running your application with limited resources can be tricky. To achieve this you will implement a pooling mechanism for your PostgreSQL instances to connect to other databases. Finally, you will take a look at some basic errors faced while working with PostgreSQL and learn to resolve them in the quickest manner. What You Will Learn Configure PostgreSQL for performance Monitor and maintain PostgreSQL instances Implement a backup strategy for your data Resolve errors faced while using PostgreSQL Who This Book Is For Readers with basic knowledge of PostgreSQL who wish to implement key solutions based on their environment.

Big Data Analytics with Spark - Mohammed Guller - 2015-12-29
Big Data Analytics with Spark is a step-by-step guide for learning Spark, which is an open-source fast and general-purpose cluster computing framework for large-scale data analysis. You will learn how to use Spark for different types of big data analytics projects, including batch, interactive, graph, and stream data analysis as well as machine learning. In addition, this book will help you become a much sought-after Spark expert. Spark is one of the hottest Big Data technologies. The amount of data generated today by devices, applications and users is exploding. Therefore, there is a critical need for tools that can analyze large-scale data and unlock value from it. Spark is a powerful technology that meets that need. You can, for example, use Spark to perform low latency computations through the use of efficient caching and iterative algorithms; leverage the features of its shell for easy and interactive Data analysis; employ its fast batch processing and low latency features to process your real time data streams and so on. As a result, adoption of Spark is rapidly growing and is replacing Hadoop MapReduce as the technology of choice for big data analytics. This book provides an introduction to Spark and related big-data technologies. It covers Spark core and its add-on libraries, including Spark SQL, Spark Streaming, GraphX, and MLlib. Big Data Analytics with Spark is therefore written for busy professionals who prefer learning a new technology from a
Kumar Sharma -
trying to pick bits and pieces from different sources. The book also provides
a chapter on Scala, the hottest functional programming language, and the
program that underlies Spark. You'll learn the basics of functional
programming in Scala, so that you can write Spark applications in it. What's
more, Big Data Analytics with Spark provides an introduction to other big
data technologies that are commonly used along with Spark, like Hive, Avro,
Kafka and so on. So the book is self-sufficient; all the technologies that you
need to know to use Spark are covered. The only thing that you are
expected to know is programming in any language. There is a critical
shortage of people with big data expertise, so companies are willing to pay
top dollar for people with skills in areas like Spark and Scala. So reading
this book and absorbing its principles will provide a boost—possibly a big
boost—to your career.

**Big Data Analytics** - Satish Narayana Srirama - 2022-01-01
This book constitutes the proceedings of the 8th International Conference
on Big Data Analytics, BDA 2021, which took place during December 2021.
Due to COVID-19 pandemic the conference was held virtually. The 16 full
and 3 short papers included in this volume were carefully reviewed and
selected from 41 submissions. The contributions were organized in topical
sections named as follows: medical and health applications; machine/deep
learning; IoTs, sensors, and networks; fundamentation; pattern mining and
data analytics.

**Big Data Analytics** - Satish Narayana Srirama - 2022-01-01
This book constitutes the proceedings of the 8th International Conference
on Big Data Analytics, BDA 2021, which took place during December 2021.
Due to COVID-19 pandemic the conference was held virtually. The 16 full
and 3 short papers included in this volume were carefully reviewed and
selected from 41 submissions. The contributions were organized in topical
sections named as follows: medical and health applications; machine/deep
learning; IoTs, sensors, and networks; fundamentation; pattern mining and
data analytics.

**Micro-Electronics and Telecommunication Engineering** - Devendra

**IBM Reference Architecture for High Performance Data and AI in
Healthcare and Life Sciences** - Dino Quintero - 2019-09-08
This IBM® Redpaper publication provides an update to the original
description of IBM Reference Architecture for Genomics. This paper
expands the reference architecture to cover all of the major vertical areas of
healthcare and life sciences industries, such as genomics, imaging, and
clinical and translational research. The architecture was renamed IBM
Reference Architecture for High Performance Data and AI in Healthcare
and Life Sciences to reflect the fact that it incorporates key building blocks
for high-performance computing (HPC) and software-defined storage, and
that it supports an expanding infrastructure of leading industry partners,
platforms, and frameworks. The reference architecture defines a highly
flexible, scalable, and cost-effective platform for accessing, managing,
storing, sharing, integrating, and analyzing big data, which can be deployed
on-premises, in the cloud, or as a hybrid of the two. IT organizations can use
the reference architecture as a high-level guide for overcoming data
management challenges and processing bottlenecks that are frequently
encountered in personalized healthcare initiatives, and in compute-intensive
and data-intensive biomedical workloads. This reference architecture also
provides a framework and context for modern healthcare and life sciences
institutions to adopt cutting-edge technologies, such as cognitive life
sciences solutions, machine learning and deep learning, Spark for analytics,
and cloud computing. To illustrate these points, this paper includes case
studies describing how clients and IBM Business Partners alike used the
reference architecture in the deployments of demanding infrastructures for
precision medicine. This publication targets technical professionals
(consultants, technical support staff, IT Architects, and IT Specialists) who
are responsible for providing life sciences solutions and support.

**IBM Reference Architecture for High Performance Data and AI in
Healthcare and Life Sciences** - Dino Quintero - 2019-09-08
should interest even the most advanced users. Analyze, explore, transform, description of IBM Reference Architecture for Genomics. This paper expands the reference architecture to cover all of the major vertical areas of healthcare and life sciences industries, such as genomics, imaging, and clinical and translational research. The architecture was renamed IBM Reference Architecture for High Performance Data and AI in Healthcare and Life Sciences to reflect the fact that it incorporates key building blocks for high-performance computing (HPC) and software-defined storage, and that it supports an expanding infrastructure of leading industry partners, platforms, and frameworks. The reference architecture defines a highly flexible, scalable, and cost-effective platform for accessing, managing, storing, sharing, integrating, and analyzing big data, which can be deployed on-premises, in the cloud, or as a hybrid of the two. IT organizations can use the reference architecture as a high-level guide for overcoming data management challenges and processing bottlenecks that are frequently encountered in personalized healthcare initiatives, and in compute-intensive and data-intensive biomedical workloads. This reference architecture also provides a framework and context for modern healthcare and life sciences institutions to adopt cutting-edge technologies, such as cognitive life sciences solutions, machine learning and deep learning, Spark for analytics, and cloud computing. To illustrate these points, this paper includes case studies describing how clients and IBM Business Partners alike used the reference architecture in the deployments of demanding infrastructures for precision medicine. This publication targets technical professionals (consultants, technical support staff, IT Architects, and IT Specialists) who are responsible for providing life sciences solutions and support.

**Mastering Spark with R** - Javier Luraschi - 2019-10-07
If you’re like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes 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; automate the process in production-ready workflows. Perform analysis and modeling across many machines using distributed computing techniques. Use large-scale data from multiple sources and different formats with ease from within Spark. Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale. Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions.

**Accelerate** - Nicole Forsgren PhD - 2018-03-27
Winner of the Shingo Publication Award. Accelerate your organization to win in the marketplace. How can we apply technology to drive business value? For years, we’ve been told that the performance of software delivery teams doesn’t matter—that it can’t provide a competitive advantage to our
for graph analytics. Also included: sample code and tips for over 20 collected from the State of DevOps reports conducted with Puppet, Dr. Nicole Forsgren, Jez Humble, and Gene Kim set out to find a way to measure software delivery performance—and what drives it—using rigorous statistical methods. This book presents both the findings and the science behind that research, making the information accessible for readers to apply in their own organizations. Readers will discover how to measure the performance of their teams, and what capabilities they should invest in to drive higher performance. This book is ideal for management at every level.

**Accelerate** - Nicole Forsgren PhD - 2018-03-27
Winner of the Shingo Publication Award Accelerate your organization to win in the marketplace. How can we apply technology to drive business value? For years, we’ve been told that the performance of software delivery teams doesn’t matter—that it can't provide a competitive advantage to our companies. Through four years of groundbreaking research to include data collected from the State of DevOps reports conducted with Puppet, Dr. Nicole Forsgren, Jez Humble, and Gene Kim set out to find a way to measure software delivery performance—and what drives it—using rigorous statistical methods. This book presents both the findings and the science behind that research, making the information accessible for readers to apply in their own organizations. Readers will discover how to measure the performance of their teams, and what capabilities they should invest in to drive higher performance. This book is ideal for management at every level.

**Graph Algorithms** - Mark Needham - 2019-05-16
Discover how graph algorithms can help you leverage the relationships within your data to develop more intelligent solutions and enhance your machine learning models. You'll learn how graph analytics are uniquely suited to unfold complex structures and reveal difficult-to-find patterns lurking in your data. Whether you are trying to build dynamic network models or forecast real-world behavior, this book illustrates how graph algorithms deliver value—from finding vulnerabilities and bottlenecks to detecting communities and improving machine learning predictions. This practical book walks you through hands-on examples of how to use graph algorithms in Apache Spark and Neo4j—two of the most common choices for graph analytics. Also included: sample code and tips for over 20 practical graph algorithms that cover optimal pathfinding, importance through centrality, and community detection. Learn how graph analytics vary from conventional statistical analysis. Understand how classic graph algorithms work, and how they are applied. Get guidance on which algorithms to use for different types of questions. Explore algorithm examples with working code and sample datasets from Spark and Neo4j. See how connected feature extraction can increase machine learning accuracy and precision. Walk through creating an ML workflow for link prediction combining Neo4j and Spark.

**Nineteen Eighty-Four** - George Orwell - 2022-01-20
WAR IS PEACE FREEDOM IS SLAVERY IGNORANCE IS STRENGTH
Winston Smith is a good worker. He supports the Party. He is good at his...
individuals that comprise them) achieve their B State, enabling them to but there is nothing to see. Winston's struggle against the totalitarian world he inhabits is a closely guarded secret. It exists only in his mind until he begins a secret love affair with Julia, a fellow worker. Is this enough to push him to revolution? Or is it the beginning of his downfall? A masterwork of dystopian fiction, Nineteen Eighty-Four is harrowingly prescient, and its impact has stretched around the globe. With a new introduction by political editor and writer Ian Dunt, this brand new edition of a science fiction classic is a must-have for any collector.

**Nineteen Eighty-Four** - George Orwell - 2022-01-20

**WAR IS PEACE FREEDOM IS SLAVERY IGNORANCE IS STRENGTH**

Winston Smith is a good worker. He supports the Party. He is good at his job rewriting history to Government specification. Big Brother watches him, but there is nothing to see. Winston's struggle against the totalitarian world he inhabits is a closely guarded secret. It exists only in his mind until he begins a secret love affair with Julia, a fellow worker. Is this enough to push him to revolution? Or is it the beginning of his downfall? A masterwork of dystopian fiction, Nineteen Eighty-Four is harrowingly prescient, and its impact has stretched around the globe. With a new introduction by political editor and writer Ian Dunt, this brand new edition of a science fiction classic is a must-have for any collector.

**B State** - Mark Samuel - 2018-10-16

Transforming Business, Organizational Culture, and Self In business and life, there are often moments when one simply can't seem to find a way forward. Searching in the past for solutions to persistent problems results in frustration and confusion. Issues in corporate teamwork and individual relationships can feel overwhelming and even insurmountable. There's a lack of control and a sense of being stuck. B State provides a clear roadmap from point A to point B to rapidly achieve measurable, breakthrough results. It's about a true transformation that removes old mindsets and silos, while replacing inefficient behaviors with desired habits to quickly create the highest performing culture for groundbreaking business outcomes. Equipped with over 30 years of professional and academic expertise, author, speaker, and change agent Mark Samuel helps companies (and the individuals that comprise them) achieve their B State, enabling them to make the necessary changes they didn't think were possible. His strategies for finding and enacting solutions to complex challenges use real life examples to help readers embrace accountability and envision their success in order to achieve the transformation they need. This book focuses readers on where they want to go, and it helps them get there fast. Written for business executives, managers, supervisors, and leaders at all levels, this is a book about how to not just do business but also live life. It brings about the dynamic forward launch readers are looking for, creating results that are both unprecedented and sustainable.

**Data Science on AWS** - Chris Fregly - 2021-04-07
With this practical book, AI and machine learning practitioners will learn how to successfully build and deploy data science projects on Amazon Web Services. The Amazon AI and machine learning stack unifies data science, data engineering, and application development to help level up your skills. This guide shows you how to build and run pipelines in the cloud, then integrate the results into applications in minutes instead of days. Throughout the book, authors Chris Fregly and Antje Barth demonstrate how to reduce cost and improve performance. Apply the Amazon AI and ML stack to real-world use cases for natural language processing, computer vision, fraud detection, conversational devices, and more. Use automated machine learning to implement a specific subset of use cases with SageMaker Autopilot. Dive deep into the complete model development lifecycle for a BERT-based NLP use case including data ingestion, analysis, model training, and deployment. Tie everything together into a repeatable machine learning operations pipeline. Explore real-time ML, anomaly detection, and streaming analytics on data streams with Amazon Kinesis and Managed Streaming for Apache Kafka. Learn security best practices for data science projects and workflows including identity and access management, authentication, authorization, and more.

**Data Science on AWS** - Chris Fregly - 2021-04-07

This book presents the proceedings of the International Conference on Cyber-Physical Systems and Control (CPS&C'2019), held in Peter the Great St. Petersburg Polytechnic University, which is celebrating its 120th anniversary in 2019. The CPS&C'2019 was dedicated to the 35th anniversary of the partnership between Peter the Great St. Petersburg Polytechnic University and Leibniz University of Hannover. Cyber-physical systems (CPSs) are a new generation of control systems and techniques that help promote prospective interdisciplinary research. A wide range of theories and methodologies are currently being investigated and developed in this area to tackle various complex and challenging problems. Accordingly, CPSs represent a scientific and engineering discipline that is set to make an impact on future systems of industrial and social scale that are characterized by the deep integration of real-time processing, sensing, and actuation into logical and physical heterogeneous domains. The CPS&C'2019 brought together researchers and practitioners from all over the world and to discuss cross-cutting fundamental scientific and engineering principles that underline the integration of cyber and physical elements across all application fields. The participants represented research institutions and universities from Austria, Belgium, Bulgaria, China, Finland, Germany, the Netherlands, Russia, Syria, Ukraine, the USA, and Vietnam. These proceedings include 75 papers arranged into five sections, namely keynote papers, fundamentals, applications, technologies, and education and social aspects.

**Cyber-Physical Systems and Control** - Dmitry G. Arseniev - 2019-11-29

This book presents the proceedings of the International Conference on Cyber-Physical Systems and Control (CPS&C'2019), held in Peter the Great St. Petersburg Polytechnic University, which is celebrating its 120th anniversary in 2019. The CPS&C'2019 was dedicated to the 35th anniversary of the partnership between Peter the Great St. Petersburg Polytechnic University and Leibniz University of Hannover. Cyber-physical systems (CPSs) are a new generation of control systems and techniques that help promote prospective interdisciplinary research. A wide range of theories and methodologies are currently being investigated and developed in this area to tackle various complex and challenging problems. Accordingly, CPSs represent a scientific and engineering discipline that is set to make an impact on future systems of industrial and social scale that are characterized by the deep integration of real-time processing, sensing, and actuation into logical and physical heterogeneous domains. The CPS&C'2019 brought together researchers and practitioners from all over the world and to discuss cross-cutting fundamental scientific and engineering principles that underline the integration of cyber and physical elements across all application fields. The participants represented research institutions and universities from Austria, Belgium, Bulgaria, China, Finland, Germany, the Netherlands, Russia, Syria, Ukraine, the USA, and Vietnam. These proceedings include 75 papers arranged into five sections, namely keynote papers, fundamentals, applications, technologies, and education and social aspects.
systems (CPSs) are a new generation of control systems and techniques that help promote prospective interdisciplinary research. A wide range of theories and methodologies are currently being investigated and developed in this area to tackle various complex and challenging problems. Accordingly, CPSs represent a scientific and engineering discipline that is set to make an impact on future systems of industrial and social scale that are characterized by the deep integration of real-time processing, sensing, and actuation into logical and physical heterogeneous domains. The CPS&C'2019 brought together researchers and practitioners from all over the world and to discuss cross-cutting fundamental scientific and engineering principles that underline the integration of cyber and physical elements across all application fields. The participants represented research institutions and universities from Austria, Belgium, Bulgaria, China, Finland, Germany, the Netherlands, Russia, Syria, Ukraine, the USA, and Vietnam. These proceedings include 75 papers arranged into five sections, namely keynote papers, fundamentals, applications, technologies, and education and social aspects.

**Innovations in Smart Cities Applications Volume 4** - Mohamed Ben Ahmed - 2021-02-12

This proceedings book is the fourth edition of a series of works which features emergent research trends and recent innovations related to smart city presented at the 5th International Conference on Smart City Applications SCA20 held in Saffronboul, Turkey. This book is composed of peer-reviewed chapters written by leading international scholars in the field of smart cities from around the world. This book covers all the smart city topics including Smart Citizenship, Smart Education, Smart Mobility, Smart Healthcare, Smart Mobility, Smart Security, Smart Earth Environment & Agriculture, Smart Economy, Smart Factory and Smart Recognition Systems. This book contains a special section intended for Covid-19 pandemic research. This book edition is an invaluable resource for courses in computer science, electrical engineering and urban sciences for sustainable development.

**High Performance JavaScript** - Nicholas C. Zakas - 2010-03-11

If you're like most developers, you rely heavily on JavaScript to build interactive and quick-responding web applications. The problem is that all of those lines of JavaScript code can slow down your apps. This book reveals techniques and strategies to help you eliminate performance bottlenecks during development. You'll learn how to improve execution time, downloading, interaction with the DOM, page life cycle, and more. Yahoo! frontend engineer Nicholas C. Zakas and five other JavaScript experts—Ross Harmes, Julien Lecomte, Steven Levithan, Stoyan Stefanov, and Matt Sweeney—demonstrate optimal ways to load code onto a page, and offer programming tips to help your JavaScript run as efficiently and quickly as possible. You'll learn the best practices to build and deploy your files to a production environment, and tools that can help you find problems once your site goes live. Identify problem code and use faster alternatives to accomplish the same task. Improve scripts by learning how JavaScript stores and accesses data. Implement JavaScript code so that it doesn't slow down interaction with the DOM. Use optimization techniques to improve runtime performance. Learn ways to ensure the UI is responsive at all times. Achieve faster client-server communication. Use a build system to minify files, and HTTP compression to deliver them to the browser.
If you're like most developers, you rely heavily on JavaScript to build interactive and quick-responding web applications. The problem is that all of those lines of JavaScript code can slow down your apps. This book reveals techniques and strategies to help you eliminate performance bottlenecks during development. You'll learn how to improve execution time, downloading, interaction with the DOM, page life cycle, and more. Yahoo! frontend engineer Nicholas C. Zakas and five other JavaScript experts—Ross Harmes, Julien Lecomte, Steven Levithan, Stoyan Stefanov, and Matt Sweeney—demonstrate optimal ways to load code onto a page, and offer programming tips to help your JavaScript run as efficiently and quickly as possible. You'll learn the best practices to build and deploy your files to a production environment, and tools that can help you find problems once your site goes live. Identify problem code and use faster alternatives to accomplish the same task Improve scripts by learning how JavaScript stores and accesses data Implement JavaScript code so that it doesn't slow down interaction with the DOM Use optimization techniques to improve runtime performance Learn ways to ensure the UI is responsive at all times Achieve faster client-server communication Use a build system to minify files, and HTTP compression to deliver them to the browser

**Neural Approaches to Dynamics of Signal Exchanges** - Anna Esposito - 2019-09-18

The book presents research that contributes to the development of intelligent dialog systems to simplify diverse aspects of everyday life, such as medical diagnosis and entertainment. Covering major thematic areas: machine learning and artificial neural networks; algorithms and models; and social and biometric data for applications in human–computer interfaces, it discusses processing of audio-visual signals for the detection of user-perceived states, the latest scientific discoveries in processing verbal (lexicon, syntax, and pragmatics), auditory (voice, intonation, vocal expressions) and visual signals (gestures, body language, facial expressions), as well as algorithms for detecting communication disorders, remote health-status monitoring, sentiment and affect analysis, social behaviors and engagement. Further, it examines neural and machine learning algorithms for the implementation of advanced telecommunication systems, communication with people with special needs, emotion modulation by computer contents, advanced sensors for tracking changes in real-life and automatic systems, as well as the development of advanced human–computer interfaces. The book does not focus on solving a particular problem, but instead describes the results of research that has positive effects in different fields and applications.


Apache Ignite is one of the most widely used open source memory-centric distributed, caching, and processing platform. This allows the users to use the platform as an in-memory computing framework or a full functional persistence data stores with SQL and ACID transaction support. On the other hand, Apache Ignite can be used for accelerating existing Relational and NoSQL databases, processing events & streaming data or developing...
Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists interested in learning in-memory computing and distributed database. This book intends to provide someone with little to no experience of Apache Ignite with an opportunity to learn how to use this platform effectively from scratch taking a practical hands-on approach to learning. Please see the table of contents for more details.

**Apache Ignite Book** - Michael Zheludkov - 2019-02-25
Apache Ignite is one of the most widely used open source memory-centric distributed, caching, and processing platform. This allows the users to use the platform as an in-memory computing framework or a full functional persistence data stores with SQL and ACID transaction support. On the other hand, Apache Ignite can be used for accelerating existing Relational and NoSQL databases, processing events & streaming data or developing Microservices in fault-tolerant fashion. This book addressed anyone interested in learning in-memory computing and distributed database. This book intends to provide someone with little to no experience of Apache Ignite with an opportunity to learn how to use this platform effectively from scratch taking a practical hands-on approach to learning. Please see the table of contents for more details.

**CliftonStrengths for Students** - Gallup - 2017-07-25
Helps aspiring college students discover where their strengths truly lie and how to develop them to reach their full potential at school and later in the real world.

**CliftonStrengths for Students** - Gallup - 2017-07-25
Helps aspiring college students discover where their strengths truly lie and how to develop them to reach their full potential at school and later in the real world.

**Python Data Science Handbook** - Jake VanderPlas - 2016-11-21
For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you’ll learn how to use: IPython and Jupyter: provide computational environments for data scientists using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib: includes capabilities for a flexible range of data visualizations in Python Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms