

Real Time Analytics Algorithms And Systems Vldb

This book presents new software engineering approaches and methods, discussing real-world problems and exploratory research that describes novel approaches, modern design techniques, hybrid algorithms and empirical methods. This book constitutes part of the refereed proceedings of the Software Engineering and Algorithms in Intelligent Systems Section of the 7th Computer Science On-line Conference 2018 (CSOC 2018), held in April 2018.

Master alternative Big Data technologies that can do what Hadoop can't: real-time analytics and iterative machine learning. When most technical professionals think of Big Data analytics today, they think of Hadoop. But there are many cutting-edge applications that Hadoop isn't well suited for, especially real-time analytics and contexts requiring the use of iterative machine learning algorithms. Fortunately, several powerful new technologies have been developed specifically for use cases such as these. Big Data Analytics Beyond Hadoop is the first guide specifically designed to help you take the next steps beyond Hadoop. Dr. Vijay Srinivas Agneeswaran introduces the breakthrough Berkeley Data Analysis Stack (BDAS) in detail, including its motivation, design, architecture, Mesos cluster management, performance, and more. He presents realistic use cases and up-to-date example code for: Spark, the next generation in-memory computing technology from UC Berkeley Storm, the parallel real-time Big Data analytics technology from Twitter GraphLab, the next-generation graph processing paradigm from CMU and the University of Washington (with comparisons to alternatives such as Pregel and Piccolo) Halo also offers architectural and design guidance and code sketches for scaling machine learning algorithms to Big Data, and then realizing them in real-time. He concludes by previewing emerging trends, including real-time video analytics, SDNs, and even Big Data governance, security, and privacy issues. He identifies intriguing startups and new research possibilities, including BDAS extensions and cutting-edge model-driven analytics. Big Data Analytics Beyond Hadoop is an indispensable resource for everyone who wants to reach the cutting edge of Big Data analytics, and stay there: practitioners, architects, programmers, data scientists, researchers, startup entrepreneurs, and advanced students.

Provides comprehensive coverage of the current state of IoT, focusing on data processing infrastructure and techniques Written by experts in the field, this book addresses the IoT technology stack, from connectivity through data platforms to end-user case studies, and considers the tradeoffs between business needs and data security and privacy throughout. There is a particular emphasis on data processing technologies that enable the extraction of actionable insights from data to inform improved decision making. These include artificial intelligence techniques such as stream processing, deep learning and knowledge graphs, as well as data interoperability and the key aspects of privacy, security and trust. Additional aspects covered include: creating and supporting IoT ecosystems; edge computing; data mining of sensor datasets; and crowd-sourcing, amongst others. The book also presents several sections featuring use cases across a range of application areas such as smart energy, transportation, smart factories, and more. The book concludes with a chapter on key considerations when deploying IoT technologies in the enterprise, followed by a brief review of future research directions and challenges. The Internet of Things: From Data to Insight Provides a comprehensive overview of the Internet of Things technology stack with focus on data driven aspects from data modelling and processing to presentation for decision making Explains how IoT technology is applied in practice and the benefits being delivered. Acquaints readers that are new to the area with concepts, components, technologies, and verticals related to and enabled by IoT Gives IoT specialists a deeper insight into data and decision-making aspects as well as novel technologies and application areas Analyzes and presents important emerging technologies for the IoT arena Shows how different objects and devices can be connected to decision making processes at various levels of abstraction The Internet of Things: From Data to Insight will appeal to a wide audience, including IT and network specialists seeking a broad and complete understanding of IoT, CIOs and CIO teams, researchers in IoT and related fields, final year undergraduates, graduate students, post-graduates, and IT and science media professionals.

This book covers the topic of data science in a comprehensive manner and synthesizes both fundamental and advanced topics of a research area that has now reached its maturity. The book starts with the basic concepts of data science. It highlights the types of data and their use and importance, followed by a discussion on a wide range of applications of data science and widely used techniques in data science. Key Features • Provides an internationally respected collection of scientific research methods, technologies and applications in the area of data science. • Presents predictive outcomes by applying data science techniques to real-life applications. • Provides readers with the tools, techniques and cases required to excel with modern artificial intelligence methods. • Gives the reader a variety of intelligent applications that can be designed using data science and its allied fields. The book is aimed primarily at advanced undergraduates and graduates studying machine learning and data science. Researchers and professionals will also find this book useful.

Scala for Machine Learning

Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science

The algorithm of the universe

Digital Transformation for the Process Industries

Real-Time Data Analytics for Large Scale Sensor Data

Proceedings of the 5th ICACNI 2017, Volume 2

Algorithms and Applications

Leverage Scala and Machine Learning to study and construct systems that can learn from data About This Book Explore a broad variety of data processing, machine learning, and genetic algorithms through diagrams, mathematical formulation, and updated source code in Scala Take your expertise in Scala programming to the next level by creating and customizing AI applications Experiment with different techniques and evaluate their benefits and limitations using real-world applications in a tutorial style Who This Book Is For If you're a data scientist or a data analyst with a fundamental knowledge of Scala who wants to learn and implement various Machine learning techniques, this book is for you. All you need is a good understanding of the Scala programming language, a basic knowledge of statistics, a keen interest in Big Data processing, and this book! What You Will Learn Build dynamic workflows for scientific computing Leverage open source libraries to extract patterns from time series Write your own classification, clustering, or evolutionary algorithm Perform relative performance tuning and evaluation of Spark Master probabilistic models for sequential data Experiment with advanced techniques such as regularization and kernelization Dive into neural networks and some deep learning architecture Apply some basic multiarm-bandit algorithms Solve big data problems with Scala parallel collections, Akka actors, and Apache Spark clusters Apply key learning strategies to a technical analysis of financial markets In Detail The discovery of information through data clustering and classification is becoming a key differentiator for competitive organizations. Machine learning applications are everywhere, from self-driving cars, engineering design, logistics, manufacturing, and trading strategies, to detection of genetic anomalies. The book is your one stop guide that introduces you to the functional capabilities of the Scala programming language that are critical to the creation of machine learning algorithms such as dependency injection and implicits. You start by learning data preprocessing and filtering techniques. Following this, you'll move on to unsupervised learning techniques such as clustering and dimension reduction, followed by probabilistic graphical models such as Naive Bayes, hidden Markov models and Monte Carlo inference. Further, it covers the discriminative algorithms such as linear, logistic regression with regularization, kernelization, support vector machines, neural networks, and deep learning. You'll move on to evolutionary computing, multibandit algorithms, and reinforcement learning. Finally, the book includes a comprehensive overview of parallel computing in Scala and Akka followed by a description of Apache Spark and its ML library. With updated codes based on the latest version of Scala and comprehensive examples, this book will ensure that you have more than just a solid fundamental knowledge in machine learning with Scala. Style and approach This book is designed as a tutorial with hands-on exercises using technical analysis of financial markets and corporate data. The approach of each chapter is such that it allows you to understand key concepts easily.

Develop and manage effective real-time streaming solutions by leveraging the power of Microsoft Azure About This Book* Analyze your data from various sources using Microsoft Azure Stream Analytics* Develop, manage and automate your stream analytics solution with Microsoft Azure* A practical guide to real-time event processing and performing analytics on the cloud Who This Book Is For If you are looking for a resource that teaches you how to process continuous streams of data in real-time, this book is what you need. A basic understanding of the concepts in analytics is all you need to get started with this book What You Will Learn* Perform real-time event processing with Azure Stream Analytics* Incorporate the features of Big Data Lambda architecture pattern in real-time data processing* Design a streaming pipeline for storage and batch analysis* Implement data transformation and computation activities over stream of events* Automate your streaming pipeline using Powershell and the .NET SDK* Integrate your streaming pipeline with popular Machine Learning and Predictive Analytics modelling algorithms* Monitor and troubleshoot your Azure Streaming jobs effectively In Detail Microsoft Azure is a very popular cloud computing service used by many organizations around the world. Its latest analytics offering, Stream Analytics, allows you to process and get actionable insights from different kinds of data in real-time. This book is your guide to understanding the basics of how Azure Stream Analytics works, and building your own analytics solution using its capabilities. You will start with understanding what Stream Analytics is, and why it is a popular choice for getting real-time insights from data. Then, you will be introduced to Azure Stream Analytics, and see how you can use the tools and functions in Azure to develop your own Streaming Analytics. Over the course of the book, you will be given comparative analytic guidance on using Azure Streaming with other Microsoft Data Platform resources such as Big Data Lambda Architecture integration for real time data analysis and differences of scenarios for architecture designing with Azure HD Insight Hadoop clusters with Storm or Stream Analytics. The book also shows you how you can manage, monitor, and scale your solution for optimal performance. By the end of this book, you will be well-versed in using Azure Stream Analytics to develop an efficient analytics solution that can work with any type of data. Style and approach A comprehensive guidance on developing real-time event processing with Azure Stream Analysis

Real-Time Analytics Techniques to Analyze and Visualize Streaming Data John Wiley & Sons

Learn to develop and deploy dashboards as web apps using the Python programming language, and how to integrate algorithms into web apps. Author Tshepo Chris Nokeri begins by introducing you to the basics of constructing and styling static and interactive charts and tables before exploring the basics of HTML, CSS, and Bootstrap, including an approach to building web pages with HTML. From there, he'll show you the key Python web frameworks and techniques for building web apps with them. You'll then see how to style web apps and incorporate themes, including interactive charts and tables to build dashboards, followed by a walkthrough of creating URL routes and securing web apps. You'll then progress to more advanced topics, like building machine learning algorithms and integrating them into a web app. The book concludes with a demonstration of how to deploy web apps in prevalent cloud platforms. Web App Development and Real-Time Web Analytics with Python is ideal for intermediate data scientists, machine learning engineers, and web developers, who have little or no knowledge about building web apps that implement bootstrap technologies. After completing this book, you will have the knowledge necessary to create added value for your organization, as you will understand how to link front-end and back-end development, including machine learning. What You Will Learn Create interactive graphs and render static graphs into interactive ones Understand the essentials of HTML, CSS, and Bootstrap Gain insight into the key Python web frameworks, and how to develop web applications using them Develop machine learning algorithms and integrate them into web apps Secure web apps and deploy them to cloud platforms Who This Book Is For Intermediate data scientists, machine learning engineers, and web developers.

New Fundamental Technologies in Data Mining

Techniques and Intelligent Applications

Storm Blueprints: Patterns for Distributed Real-time Computation

Frontiers in Massive Data Analysis

Volume 2

Techniques to Analyze and Visualize Streaming Data

Enabling Real-time Analytics on IBM z Systems Platform

Edge computing is quickly becoming an important technology throughout a number of fields as businesses and industries alike embrace the benefits it can have in their companies. The streamlining of data is crucial for the development and evolution of businesses in order to keep up with competition and improve functions overall. In order to appropriately utilize edge computing to its full potential, further study is required to examine the potential pitfalls and opportunities of this innovative technology. The Research Anthology on Edge Computing Protocols, Applications, and Integration establishes critical research on the current uses, innovations, and challenges of edge computing across disciplines. The text highlights the history of edge computing and how it has been adapted over time to improve industries. Covering a range of topics such as bandwidth, data centers, and security, this major reference work is ideal for industry professionals, computer scientists, engineers, practitioners, researchers, academicians, scholars, instructors, and students.

The field of analytics is rapidly evolving, making it difficult for professionals and students to keep up the most current and effective applications. Managerial Analytics will help readers sort through all these new options and identify the appropriate solution. In this reference, authors Watson, Nelson and Cacioppi accurately define and identify the components of analytics and big data, giving readers the knowledge needed to effectively assess new aspects and applications. Building on this foundation, they review tools and solutions, identify the offerings best aligned to one's requirements, and show how to tailor analytics applications to an organization's specific needs. Drawing on extensive experience implementing, planning, and researching advanced analytics for business, the authors clearly explain all this, and more: What analytics is and isn't: great examples of successful usage - and other examples where the term is being degraded into meaninglessness The difference between using analytics and "competing on analytics" How to get started with big data, by analyzing the most relevant data Components of analytics systems, from databases and Excel to BI systems and beyond Anticipating and overcoming "confirmation bias" and other pitfalls Understanding predictive analytics and getting the high-quality random samples necessary Applying game theory, Efficient Frontier, benchmarking, and revenue management models Implementing optimization at the small and large scale, and using it to make "automatic decisions"

This book includes a selection of papers from the 2018 World Conference on Information Systems and Technologies (WorldCIST'18), held in Naples, Italy on March 27-29, 2018. WorldCIST is a global forum for researchers and practitioners to present and discuss recent results and innovations, current trends, professional experiences and the challenges of modern information systems and technologies research together with their technological development and applications. The main topics covered are: A) Information and Knowledge Management; B) Organizational Models and Information Systems; C) Software and Systems Modeling; D) Software Systems, Architectures, Applications and Tools; E) Multimedia Systems and Applications; F) Computer Networks, Mobility and Pervasive Systems; G) Intelligent and Decision Support Systems; H) Big Data Analytics and Applications; I) Human-Computer Interaction; J) Ethics, Computers & Security; K) Health Informatics; L) Information Technologies in Education; M) Information Technologies in Radiocommunications; N) Technologies for Biomedical Applications.

There are many attempts to explain the existence of the universe. Even the observable universe around us is filled with unknowns and numerous unexplainable phenomena, let alone the unobservable universe. Each and every theory proposed is just trying to find an explanation for the existence of the observable universe around us, ignoring the fact that it is potentially possible there is an invisible universe which reveals itself as we find newer ways to observe it. We need to first and foremost be able to abstract this universe as we observe it to be just a representation of something else underneath that is projected as this world around us by our descriptive thoughts. Then and only then can we find an explanation that is inclusive of everything that the universe is. As far as we search for the origin of just these appearances of descriptive thoughts, we will only dabble with physics, chemistry, mathematics and computations. It should be recognised that Wolfram introduces a very good concept. Simple things recurse to form complex worlds. But this without the correct transformation remains simply a recognisable pattern. How does this recognisable pattern go on to become this complex appearance of reality that we observe around us is up in the air. The ancient Sanskrit literature has an in-depth explanation of how this reality around us works and how it has come to be this. It just needs the correct context around which to translate these texts to understand it. For a long time I wondered what would be the best way to explain the algorithm of the universe as seen by the ancients and explained in the Sanskrit literature. I could explain it by using translation of the Sanskrit verses in a certain order. But I find that they do not take the reader on the journey that is necessary to understand and form the correct thoughts to understand the working of and creation of this reality around us. I thought I would write it as a description of astrology i.e., jyothiSha which, if we look at from the ancient view, is actually a science that tells us how to control the distribution of seeking which drives the formation of this reality around us and hence control this reality itself. But, I find that the term has been highly mangled with all sorts of meaning associated with it. I find, even if there was a very beautiful science explained it becomes very difficult to break these pre-conceived perceptions to get the reader to read objectively and understand. Finally, I have decided the best way is to use the computers that we have created and the AI algorithms that we are trying to code as the best way to describe it, because this was my primary driver to want to understand the Sanskrit literature and to debug the working of the universe. The underlying curiosity to understand why and how the AI that we write does not even seem to be a "fake imitation" of the working of the universe around us. The underlying curiosity to understand how this reality around me is so completely immersive while the virtual reality that we create does not even come near this immersive experience that we have with reality. The curiosity to understand if it is possible for us to create a hardware and software that comes even a little bit close to the kind of intelligence, knowledge and beauty in the algorithm of this reality around us. This has been my driver to decode and understand the Sanskrit literature. So, I find that is the best way to describe the algorithm. In this book I explain the algorithms proposed in the Sanskrit literatures for the working of this reality around us. I also compare these concepts to what we have currently present in computers and explain how they are different from what is explained in these books and why what we have in computers are not sufficient to create the AI that we want to create. Further, I try to put down my thoughts on how or what can be done to create a system that can be used to implement a AI system.

A Roadmap

5th International Workshop, BIRTE 2011, Held at the 37th International Conference on Very Large Databases, VLDB 2011, Seattle, WA, USA, September 2, 2011, Revised Selected Papers

Trends and Advances in Information Systems and Technologies

Thingalytics

Proceedings of 7th Computer Science On-line Conference 2018, Volume 1

Big Data and Artificial Intelligence in Digital Finance

Fog computing is quickly increasing its applications and uses to the next level. As it continues to grow, different types of virtualization technologies can thrust this branch of computing further into mainstream use. The Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science is a key reference volume on the latest research on the role of next-generation systems and devices that are capable of self-learning and how those devices will impact society. Featuring wide-ranging coverage across a variety of relevant views and themes such as cognitive analytics, data mining algorithms, and the internet of things, this publication is ideally designed for programmers, IT professionals, students, researchers, and engineers looking for innovative research on software-defined cloud infrastructures and domain-specific analytics.

This book constitutes the thoroughly refereed conference proceedings of the 5th International Workshop on Business Intelligence for the Real-Time Enterprise, BIRTE 2011, held in Seattle, WA, USA, in September 2011, in conjunction with VLDB 2011, the International Conference on Very Large Data Bases. The series of BIRTE workshops aims to provide a forum for researchers to discuss and advance the foundational science and engineering required to enable real-time business intelligence as well as novel applications and solutions based on these foundational techniques. The volume contains 6 research papers, which have been carefully reviewed and selected from 12 submissions, plus the 3 keynotes presented at the workshop. The topics cover all stages of the business intelligence cycle, including capturing of real-time data, handling of temporal or uncertain data, performance issues, event management, and the optimization of complex ETL workflows. The volume contains 6 research papers, which have been carefully reviewed and selected from 12 submissions, plus the 3 keynotes presented at the workshop. The topics cover all stages of the business intelligence cycle, including capturing of real-time data, handling of temporal or uncertain data, performance issues, event management, and the optimization of complex ETL workflows.

Massive modern datasets make traditional data structures and algorithms grind to a halt. This fun and practical guide introduces cutting-edge techniques that can reliably handle even the largest distributed datasets. In Algorithms and Data Structures for Massive Datasets you will learn: Probabilistic sketching data structures for practical problems Choosing the right database engine for your application Evaluating and designing efficient on-disk data structures and algorithms Understanding the algorithmic trade-offs involved in massive-scale systems Deriving basic statistics from streaming data Correctly sampling streaming data Computing percentiles with limited space resources Algorithms and Data Structures for Massive Datasets reveals a toolbox of new methods that are perfect for handling modern big data applications. You'll explore the novel data structures and algorithms that underpin Google, Facebook, and other enterprise applications that work with truly massive amounts of data. These effective techniques can be applied to any discipline, from finance to text analysis. Graphics, illustrations, and hands-on industry examples make complex ideas practical to implement in your projects—and there's no mathematical proofs to puzzle over. Work through this one-of-a-kind guide, and you'll find the sweet spot of saving space without sacrificing your data's accuracy. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Standard algorithms and data structures may become slow—or fail altogether—when applied to large distributed datasets. Choosing algorithms designed for big data saves time, increases accuracy, and reduces processing cost. This unique book distills cutting-edge research papers into practical techniques for sketching, streaming, and organizing massive datasets on-disk and in the cloud. About the book Algorithms and Data Structures for Massive Datasets introduces processing and analytics techniques for large distributed data. Packed with industry stories and entertaining illustrations, this friendly guide makes even complex concepts easy to understand. You'll explore real-world examples as you learn to map powerful algorithms like Bloom filters, Count-min sketch, HyperLogLog, and LSM-trees to your own use cases. What's inside Probabilistic sketching data structures Choosing the right database engine Designing efficient on-disk data structures and algorithms Algorithmic tradeoffs in massive-scale systems Computing percentiles with limited space resources About the reader Examples in Python, R, and pseudocode. About the author Dzejla Medjedovic earned her PhD in the Applied Algorithms Lab at Stony Brook University, New York. Emin Tahirovic earned his PhD in biostatistics from University of Pennsylvania. Illustrator Ines Dedovic earned her PhD at the Institute for Imaging and Computer Vision at RWTH Aachen University, Germany. Table of Contents 1 Introduction PART 1 HASH-BASED SKETCHES 2 Review of hash tables and modern hashing 3 Approximate membership: Bloom and quotient filters 4 Frequency estimation and count-min sketch 5 Cardinality estimation and HyperLogLog PART 2 REAL-TIME ANALYTICS 6 Streaming data: Bringing everything together 7 Sampling from data streams 8 Approximate quantiles on data streams PART 3 DATA STRUCTURES FOR DATABASES AND EXTERNAL MEMORY ALGORITHMS 9 Introducing the external memory model 10 Data structures for databases: B-trees, B+ trees, and LSM-trees 11 External memory sorting

A hands-on approach to tasks and techniques in data stream mining and real-time analytics, with examples in MOA, a popular freely available open-source software framework. Today many information sources—including sensor networks, financial markets,

social networks, and healthcare monitoring—are so-called data streams, arriving sequentially and at high speed. Analysis must take place in real time, with partial data and without the capacity to store the entire data set. This book presents algorithms and techniques used in data stream mining and real-time analytics. Taking a hands-on approach, the book demonstrates the techniques using MOA (Massive Online Analysis), a popular, freely available open-source software framework, allowing readers to try out the techniques after reading the explanations. The book first offers a brief introduction to the topic, covering big data mining, basic methodologies for mining data streams, and a simple example of MOA. More detailed discussions follow, with chapters on sketching techniques, change, classification, ensemble methods, regression, clustering, and frequent pattern mining. Most of these chapters include exercises, an MOA-based lab session, or both. Finally, the book discusses the MOA software, covering the MOA graphical user interface, the command line, use of its API, and the development of new methods within MOA. The book will be an essential reference for readers who want to use data stream mining as a tool, researchers in innovation or data stream mining, and programmers who want to create new algorithms for MOA.

Master complex big data processing, stream analytics, and machine learning with Apache Spark Streaming Data

Develop and Integrate Machine Learning Algorithms into Web Apps Managerial Analytics

15th EAI International Conference, BODYNETS 2020, Tallinn, Estonia, October 21, 2020, Proceedings Software Engineering and Algorithms in Intelligent Systems

Concepts, Methodologies, Tools, and Applications

Build efficient data flow and machine learning programs with this flexible, multi-functional open-source cluster-computing framework Key Features Master the art of real-time big data processing and machine learning Explore a wide range of use-cases to analyze large data Discover ways to optimize your work by using many features of Spark 2.x and Scala Book Description Apache Spark is an in-memory, cluster-based data processing system that provides a wide range of functionalities such as big data processing, analytics, machine learning, and more. With this Learning Path, you can take your knowledge of Apache Spark to the next level by learning how to expand Spark's functionality and building your own data flow and machine learning programs on this platform. You will work with the different modules in Apache Spark, such as interactive querying with Spark SQL, using DataFrames and datasets, implementing streaming analytics with Spark Streaming, and applying machine learning and deep learning techniques on Spark using MLlib and various external tools. By the end of this elaborately designed Learning Path, you will have all the knowledge you need to master Apache Spark, and build your own big data processing and analytics pipeline quickly and without any hassle. This Learning Path includes content from the following Packt products: Mastering Apache Spark 2.x by Romeo Kienzler Scala and Spark for Big Data Analytics by Md. Rezaul Karim, Sridhar Alla Apache Spark 2.x Machine Learning Cookbook by Siamak Amirghodsi, Meenakshi Rajendran, Broderick Hall, Shuen Mei Cookbook What you will learn Get to grips with all the features of Apache Spark 2.x Perform highly optimized real-time big data processing Use ML and DL techniques with Spark MLlib and third-party tools Analyze structured and unstructured data using SparkSQL and GraphX Understand tuning, debugging, and monitoring of big data applications Build scalable and fault-tolerant streaming applications Develop scalable recommendation engines Who this book is for If you are an intermediate-level Spark developer looking to master the advanced capabilities and use-cases of Apache Spark 2.x, this Learning Path is ideal for you. Big data professionals who want to learn how to integrate and use the features of Apache Spark and build a strong big data pipeline will also find this Learning Path useful. To grasp the concepts explained in this Learning Path, you must know the fundamentals of Apache Spark and Scala.

Learn the right cutting-edge skills and knowledge to leverage Spark Streaming to implement a wide array of real-time, streaming applications. This book walks you through end-to-end real-time application development using real-world applications, data, and code. Taking an application-first approach, each chapter introduces use cases from a specific industry and uses publicly available datasets from that domain to unravel the intricacies of production-grade design and implementation. The domains covered in Pro Spark Streaming include social media, the sharing economy, finance, online advertising, telecommunication, and IoT. In the last few years, Spark has become synonymous with big data processing. DStreams enhance the underlying Spark processing engine to support streaming analysis with a novel micro-batch processing model. Pro Spark Streaming by Zubair Nabi will enable you to become a specialist of latency sensitive applications by leveraging the key features of DStreams, micro-batch processing, and functional programming. To this end, the book includes ready-to-deploy examples and actual code. Pro Spark Streaming will act as the bible of Spark Streaming. What You'll Learn Discover Spark Streaming application development and best practices Work with the low-level details of discretized streams Optimize production-grade deployments of Spark Streaming via configuration recipes and instrumentation using Graphite, collectd, and Nagios Ingest data from disparate sources including MQTT, Flume, Kafka, Twitter, and a custom HTTP receiver Integrate and couple with HBase, Cassandra, and Redis Take advantage of design patterns for side-effects and maintaining state across the Spark Streaming micro-batch model Implement real-time and scalable ETL using data frames, SparkSQL, Hive, and SparkR Use streaming machine learning, predictive analytics, and recommendations Mesh batch processing with stream processing via the Lambda architecture Who This Book Is For Data scientists, big data experts, BI analysts, and data architects.

Summary Streaming Data introduces the concepts and requirements of streaming and real-time data systems. The book is an idea-rich tutorial that teaches you to think about how to efficiently interact with fast-flowing data. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology As humans, we're constantly filtering and deciphering the information streaming toward us. In the same way, streaming data applications can accomplish amazing tasks like reading live location data to recommend nearby services, tracking faults with machinery in real time, and sending digital receipts before your customers leave the shop. Recent advances in streaming data technology and techniques make it possible for any developer to build these applications if they have the right mindset. This book will let you join them. About the Book Streaming Data is an idea-rich tutorial that teaches you to think about efficiently interacting with fast-flowing data. Through relevant examples and illustrated use cases, you'll explore designs for applications that read, analyze, share, and store streaming data. Along the way, you'll discover the roles of key technologies like Spark, Storm, Kafka, Flink, RabbitMQ, and more. This book offers the perfect balance between big-picture thinking and implementation details. What's Inside The right way to collect real-time data Architecting a streaming pipeline Analyzing the data Which technologies to use and when About the Reader Written for developers familiar with relational database concepts. No experience with streaming or real-time applications required. About the Author Andrew Psaltis is a software engineer focused on massively scalable real-time analytics. Table of Contents PART 1 – A NEW HOLISTIC APPROACH Introducing streaming data Getting data from clients: data ingestion Transporting the data from collection tier: decoupling the data pipeline Analyzing streaming data Algorithms for data analysis Storing the analyzed or collected data Making the data available Consumer device capabilities and limitations accessing the data PART 2 – TAKING IT REAL WORLD Analyzing Meetup RSVPs in real time

Regarding online transaction processing (OLTP) workloads, IBM® z Systems™ platform, with IBM DB2®, data sharing, Workload Manager (WLM), geoplex, and other high-end features, is the widely acknowledged leader. Most customers now integrate business analytics with OLTP by running, for example, scoring functions from transactional context for real-time analytics or by applying machine-learning algorithms on enterprise data that is kept on the mainframe. As a result, IBM adds investment so clients can keep the complete lifecycle for data analysis, modeling, and scoring on z Systems control in a cost-efficient way, keeping the qualities of services in availability, security, reliability that z Systems solutions offer. Because of the changed architecture and tighter integration, IBM has shown, in a customer proof-of-concept, that a particular client was able to achieve an orders-of-magnitude improvement in performance, allowing that client's data scientist to investigate the data in a more interactive process. Open technologies, such as Predictive Model Markup Language (PMML) can help customers update single components instead of being forced to replace everything at once. As a result, you have the possibility to combine your preferred tool for model generation (such as SAS Enterprise Miner or IBM SPSS® Modeler) with a different technology for model scoring (such as Zementis, a company focused on PMML scoring). IBM SPSS Modeler is a leading data mining workbench that can apply various algorithms in data preparation, cleansing, statistics, visualization, machine learning, and predictive analytics. It has over 20 years of experience and continued development, and is integrated with z Systems. With IBM DB2 Analytics Accelerator 5.1 and SPSS Modeler 17.1, the possibility exists to do the complete predictive model creation including data transformation within DB2 Analytics Accelerator. So, instead of moving the data to a distributed environment, algorithms can be pushed to the data, using cost-efficient DB2 Accelerator for the required resource-intensive operations. This IBM Redbooks® publication explains the overall z Systems architecture, how the components can be installed and customized, how the new IBM DB2 Analytics Accelerator loader can help efficient data loading for z Systems data and external data, how in-database transformation, in-database modeling, and in-transactional real-time scoring can be used, and what other related technologies are available. This book is intended for technical specialists and architects, and data scientists who want to use the technology on the z Systems platform. Most of the technologies described in this book require IBM DB2 for z/OS®. For acceleration of the data investigation, data transformation, and data modeling process, DB2 Analytics Accelerator is required. Most value can be achieved if most of the data already resides on z Systems platforms, although adding external data (like from social sources) poses no problem at all.

An Applied Guide to Principles, Methods, Tools, and Best Practices

Data Streams

Data Science

Big Data Analytics Beyond Hadoop

Body Area Networks. Smart IoT and Big Data for Intelligent Health

Automotive, Mechanical and Electrical Engineering

Critical Heart Disease in Infants and Children E-Book

Master alternative Big Data technologies that can do what Hadoop can't: real-time analytics and iterative machine learning. When most technical professionals think of Big Data analytics today, they think of Hadoop. But there are many cutting-edge applications that Hadoop isn't well suited for, especially real-time analytics and contexts requiring the use of iterative machine learning algorithms. Fortunately, several powerful new technologies have been developed specifically for use cases such as these. Big Data Analytics Beyond Hadoop is the first guide specifically designed to help you take the next steps beyond Hadoop. Dr. Vijay Srinivas Agneeswaran introduces the breakthrough Berkeley Data Analytics Stack (BDAS) in detail, including its motivation, design, architecture, Mesos cluster management, performance, and more. He presents realistic use cases and up-to-date example code for: Spark, the next generation in-memory computing technology from UC Berkeley Storm, the parallel real-time Big Data analytics technology from Twitter GraphLab, the next-generation graph processing paradigm from CMU and the University of Washington (with comparisons to alternatives such as Pregel and Piccolo) Halo also offers architectural and design guidance and code sketches for scaling machine learning algorithms to Big Data, and then realizing them in real-time. He concludes by previewing emerging trends, including real-time video analytics, SDNs, and even Big Data governance, security, and privacy issues. He identifies intriguing startups and new research possibilities, including BDAS extensions and cutting-edge model-driven analytics. Big Data Analytics Beyond Hadoop is an indispensable resource for everyone who wants to reach the cutting edge of Big Data analytics, and stay there: practitioners, architects, programmers, data scientists, researchers, startup entrepreneurs, and advanced students. This book aims to explain Data Analytics towards decision making in terms of models and algorithms, theoretical concepts, applications, experiments in relevant domains or focused on specific issues. It explores the concepts of database technology, machine learning, knowledge-based system, high performance computing, information retrieval, finding patterns hidden in large datasets and data visualization. Also, it presents various paradigms including pattern mining, clustering, classification, and data analysis. Overall aim is to provide technical solutions in the field of data analytics and data mining. Features: Covers descriptive statistics with respect to predictive analytics and business analytics. Discusses different data analytics platforms for real-time applications. Explain SMART business models. Includes algorithms in data sciences alongwith automated methods and models. Explores varied challenges encountered by researchers and businesses in the realm of real-time analytics. This book aims at researchers and graduate students in data analytics, data sciences, data mining, and signal processing.

This book constitutes the refereed post-conference proceedings of the 15th International Conference on Body Area Networks, BodyNets 2020, held in Tallinn, Estonia, in October 2020. The conference was held virtually due to the COVID-19 pandemic. The 15 papers presented were selected from 30 submissions and issue new technologies to provide trustable measuring and communications mechanisms from the data source to medical health databases. Wireless body area networks (WBAN) are one major element in this process. Not only on-body devices but also technologies providing information from inside a body are in the focus of this conference. Dependable communications combined with accurate localization and behavior analysis will benefit WBAN technology and make the healthcare processes more effective. Five or six years ago, analysts working with big datasets made queries and got the results back overnight. The data world was revolutionized a few years ago when Hadoop and other tools made it possible to get the results from queries in minutes. But the revolution continues. Analysts now demand sub-second, near real-time query results. Fortunately, we have the tools to deliver them. This report examines tools and technologies that are driving real-time big data analytics.

Data processing, ML algorithms, smart analytics, and more

Recent Findings in Intelligent Computing Techniques

The Internet of Things

Apache Spark 2

with Practical Examples in MOA

First International Conference, DMBD 2016, Bali, Indonesia, June 25-30, 2016. Proceedings

Stream Analytics with Microsoft Azure

In the data stream scenario, input arrives very rapidly and there is limited memory to store the input. Algorithms have to work with one or few passes over the data, space less than linear in the input size or time significantly less than the input size. In the past few years, a new theory has emerged for reasoning about algorithms that work within these constraints on space, time, and number of passes. Some of the methods rely on metric embeddings, pseudo-random computations, sparse approximation theory and communication complexity. The applications for this scenario include IP network traffic analysis, mining text message streams and processing massive data sets in general. Researchers in Theoretical Computer Science, Databases, IP Networking and Computer Systems are working on the data stream challenges.

Construct a robust end-to-end solution for analyzing and visualizing streaming data Real-time analytics is the hottest topic in data analytics today. In Real-Time Analytics: Techniques to Analyze and Visualize Streaming Data, expert Byron Ellis teaches data analysts technologies to build an effective real-time analytics platform. This platform can then be used to make sense of the constantly changing data that is beginning to outpace traditional batch-based analysis platforms. The author is among a very few leading experts in the field. He has a prestigious background in research, development, analytics, real-time visualization, and Big Data streaming and is uniquely qualified to help you explore this revolutionary field. Moving from a description of the overall analytic architecture of real-time analytics to using specific tools to obtain targeted results, Real-Time Analytics leverages open source and modern commercial tools to construct robust, efficient systems that can provide real-time analysis in a cost-effective manner. The book includes: A deep discussion of streaming data systems and architectures Instructions for analyzing, storing, and delivering streaming data Tips on aggregating data and working with sets Information on data warehousing options and techniques Real-Time Analytics includes in-depth case studies for website analytics, Big Data, visualizing streaming and mobile data, and mining and visualizing operational data flows. The book's "recipe" layout lets readers quickly learn and implement different techniques. All of the code examples presented in the book, along with their related data sets, are available on the companion website.

The Internet of Things is changing the world. Thingalytics by Dr. John Bates is the most powerful book written to date about the Internet of Things (IoT), showing businesses how to take advantage of the fast Big Data that flows across the digital planet. Pulling from exciting examples of real-life innovation and invention, John makes IoT come alive. From digitally enriching exotic shops in Istanbul, Turkey, to crossing the USA on a sensor-enabled Greyhound Bus to finding new ways to mend people in hospital smart operating rooms, Thingalytics depicts how IoT can make our lives happier, easier, more productive and even safer. Thingalytics, a composite of "Things" and "Analytics," shows businesses how to use real-time analytics and algorithms in order to seize the opportunities that flow from IoT, while simultaneously spotting and navigating around threats. As each real world object - from people to refrigerators, to tractors and ships or cans of fizzy pop - is digitized and connected to the Internet, it presents a unique opportunity for innovative businesses to learn from, and take advantage of, the digital vibrations it creates. Illustrated by case studies from global, visionary organizations such as Coca Cola, Greyhound Bus and Medtronic, Thingalytics highlights how the alchemy of real-time analytics and smart algorithms can help turn fast Big Data into actionable gold nuggets for any business, anywhere. Digital disruption to traditional "bricks-and-mortar" businesses is happening now. Organizations must transform themselves using digital technologies. Time does not stand still in this brave, new digital world. "Digital Darwinism is unkind to those who wait," says R "Ray" Wang, a leading industry analyst who has written the Foreword to Thingalytics. John Bates personally interviewed each of the people in this book. His deep knowledge of their vision, their businesses and their goals gives him the insight and the gravitas to explain how each organization is conquering the digital world. Winners in the IoT race will not only profit but could - just possibly - avert disaster. Thingalytics becomes very exciting when we see how lives can be saved, fraud avoided, customers delighted and carbon emissions reduced.

Features comprehensive updates throughout the text, including indications, techniques, potential complications in perioperative management of patients, and surgical techniques for congenital heart disease. Covers recent advances in the treatment of pulmonary hypertension, developments in mechanical assist devices, heart and lung transplantation, and interventional cardiac catheterization. Features an all-new, full-color format that speeds navigation and helps clarify complex concepts. Contains 27 new chapters with an emphasis on the team approach to patient care in the ICU including creating multidisciplinary teams, quality and performance improvement, training, and challenges and solutions to developing a cohesive team environment. Includes a detailed chapter on bedside ultrasound, walking you through the techniques you're most likely to encounter in the ICU. Employs well-documented tables, text boxes, and algorithms to make clinical information easy to access, and to provide a more complete understanding of echocardiography, imaging modalities, pulmonary hypertension, and more. Describes the basic pharmacology and clinical applications of new pharmacologic agents. Examines issues affecting adults with congenital heart disease.

Research Anthology on Edge Computing Protocols, Applications, and Integration

Web App Development and Real-Time Web Analytics with Python

The Zen of Real-Time Analytics Using Apache Spark

Pro Spark Streaming

Real-Time Analytics

Data Processing and Real-Time Analytics

Algorithms and Data Structures for Massive Datasets

Data mining of massive data sets is transforming the way we think about crisis response, marketing, entertainment, cybersecurity and national intelligence. Collections of documents, images, videos, and networks are being thought of not merely as bit strings to be stored, indexed, and retrieved, but as potential sources of discovery and knowledge, requiring sophisticated analysis techniques that go far beyond classical indexing and keyword counting, aiming to find relational and semantic interpretations of the phenomena underlying the data. Frontiers in Massive Data Analysis examines the frontier of analyzing massive amounts of data, whether in a static database or streaming through a system. Data at that scale--terabytes and petabytes--is increasingly common in science (e.g., particle physics, remote sensing, genomics), Internet commerce, business analytics, national security, communications, and elsewhere. The tools that work to infer knowledge from data at smaller scales do not necessarily work, or work well, at such massive scale. New tools, skills, and approaches are necessary, and this report identifies many of them, plus promising research directions to explore. Frontiers in Massive Data Analysis discusses pitfalls in trying to infer knowledge from massive data, and it characterizes seven major classes of computation that are common in the analysis of massive data. Overall, this report illustrates the cross-disciplinary knowledge--from computer science, statistics, machine learning, and application disciplines--that must be brought to bear to make useful inferences from massive data.

This three volume book contains the Proceedings of 5th International Conference on Advanced Computing, Networking and Informatics (ICACNI 2017). The book focuses on the recent advancement of the broad areas of advanced computing, networking and informatics. It also includes novel approaches devised by researchers from across the globe. This book brings together academic scientists, professors, research scholars and students to share and disseminate information on knowledge and scientific research works related to computing, networking, and informatics to discuss the practical challenges encountered and the solutions adopted. The book also promotes translation of basic research into applied investigation and convert applied investigation into practice.

Apply machine learning to streaming data with the help of practical examples, and deal with challenges that surround streaming Key FeaturesWork on streaming use cases that are not taught in most data science coursesGain experience with state-of-the-art tools for streaming dataMitigate various challenges while handling streaming dataBook Description Streaming data is the new top technology to watch out for in the field of data science and machine learning. As business needs become more demanding, many use cases require real-time analysis as well as real-time machine learning. This book will help you to get up to speed with data analytics for streaming data and focus strongly on adapting machine learning and other analytics to the case of streaming data. You will first learn about the architecture for streaming and real-time machine learning. Next, you will look at the state-of-the-art frameworks for streaming data like River. Later chapters will focus on various industrial use cases for streaming data like Online Anomaly Detection and others. As you progress, you will discover various challenges and learn how to mitigate them. In addition to this, you will learn best practices that will help you use streaming data to generate real-time insights. By the end of this book, you will have gained the confidence you need to stream data in your machine learning models. What you will learnUnderstand the challenges and advantages of working with streaming dataDevelop real-time insights from streaming dataUnderstand the implementation of streaming data with various use cases to boost your knowledgeDevelop a PCA alternative that can work on real-time dataExplore best practices for handling streaming data that you absolutely need to rememberDevelop an API for real-time machine learning inferenceWho this book is for This book is for data scientists and machine learning engineers who have a background in machine learning, are practice and technology-oriented, and want to learn how to apply machine learning to streaming data through practical examples with modern technologies. Although an understanding of basic Python and machine learning concepts is a must, no prior knowledge of streaming is required.

A blueprints book with 10 different projects built in 10 different chapters which demonstrate the various use cases of storm for both beginner and intermediate users, grounded in real-world example applications. Although the book focuses primarily on Java development with Storm, the patterns are more broadly applicable and the tips, techniques, and approaches described in the book apply to architects, developers, and operations. Additionally, the book should provoke and inspire applications of distributed computing to other industries and domains. Hadoop enthusiasts will also find this book a good introduction to Storm, providing a potential migration path from batch processing to the world of real-time analytics.

Machine Learning for Streaming Data with Python

From Data to Insight

Understanding the real-time pipeline

Apache Spark 2: Data Processing and Real-Time Analytics

A new perspective on cognitive AI

Real-time Applications with Storm, Spark, and More Hadoop Alternatives
Increasing Personalization and Trust in Digital Finance using Big Data and AI

The LNCS volume LNCS 9714 constitutes the refereed proceedings of the International Conference on Data Mining and Big Data, DMBD 2016, held in Bali, Indonesia, in June 2016. The 57 papers presented in this volume were carefully reviewed and selected from 115 submissions. The theme of DMBD 2016 is "Serving Life with Data Science". Data mining refers to the activity of going through big data sets to look for relevant or pertinent information. The papers are organized in 10 cohesive sections covering all major topics of the research and development of data mining and big data and one Workshop on Computational Aspects of Pattern Recognition and Computer Vision.

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

Real-Time Data Analytics for Large-Scale Sensor Data covers the theory and applications of hardware platforms and architectures, the development of software methods, techniques and tools, applications, governance and adoption strategies for the use of massive sensor data in real-time data analytics. It presents the leading-edge research in the field and identifies future challenges in this fledgling research area. The book captures the essence of real-time IoT based solutions that require a multidisciplinary approach for catering to on-the-fly processing, including methods for high performance stream processing, adaptively streaming adjustment, uncertainty handling, latency handling, and more. Examines IoT applications, the design of real-time intelligent systems, and how to manage the rapid growth of the large volume of sensor data. Discusses intelligent management systems for applications such as healthcare, robotics and environment modeling. Provides a focused approach towards the design and implementation of real-time intelligent systems for the management of sensor data in large-scale environments.

The 2016 International Conference on Automotive Engineering, Mechanical and Electrical Engineering (AEMEE 2016) was held December 9-11, 2016 in Hong Kong, China. AEMEE 2016 was a platform for presenting excellent results and new challenges facing the fields of automotive, mechanical and electrical engineering. Automotive, Mechanical and Electrical Engineering brings together a wide range of contributions from industry and governmental experts and academics, experienced in engineering, design and research. Papers have been categorized under the following headings: Automotive Engineering and Rail Transit Engineering.

Mechanical, Manufacturing, Process Engineering, Network, Communications and Applied Information Technologies. Technologies in Energy and Power, Cell, Engines, Generators, Electric Vehicles. System Test and Diagnosis, Monitoring and Identification, Video and Image Processing. Applied and Computational Mathematics, Methods, Algorithms and Optimization. Technologies in Electrical and Electronic, Control and Automation. Industrial Production, Manufacturing, Management and Logistics.

Rapidly build practical online machine learning solutions using River and other top key frameworks

Big Data: Concepts, Methodologies, Tools, and Applications

Data Mining and Big Data

Enabling Real-Time Business Intelligence

Data Driven Decision Making using Analytics

Proceedings of the 2016 International Conference on Automotive Engineering, Mechanical and Electrical Engineering (AEMEE 2016), Hong Kong, China, December 9-11, 2016

Real-Time Big Data Analytics: Emerging Architecture

Imagine if your process manufacturing plants were running so well that your production, safety, environmental, and profitability targets were being met so that your subject matter experts could focus on data-driven business improvements. Through proper use and analysis of your existing operations data, your company can become an industry leader and reward your stakeholders. Written in an engaging and easily understandable manner, this book demonstrates a step-by-step process of how an organization can effectively utilize technology and make the necessary culture changes to achieve operational excellence. You will see how several industry-leading companies have used an effective real-time data infrastructure for mission-critical business use cases. The book also addresses challenges involved, such as effectively integrating operational (OT) data with business (IT) systems to enable a more proactive, predictive management model for a fleet of process plants. Some of the things you will take away: Learn how a real-time data infrastructure enables transformation of raw sensor data into contextualized information for operational insights and business process improvement. Understand how reusing the same operational data for multiple use cases significantly impacts fleet management, profitability, and asset stewardship. See how a simple digital unit template representing production flows can be repeatedly used to identify critical inefficiencies in plant operations. Discover best practices of deploying real-time situational awareness alerts and predictive analytics. Realize how to transform your organization into a data-driven culture for continuous sustainable improvement. Find out how leading companies integrate operations data with business intelligence and predictive analytics tools in a corporate on-premises or cloud-enabled environment. Learn how industry-leading companies have imaginatively used a real-time data infrastructure to improve yields, reduce cycle times, and slash operating costs. This book is targeted for process industries production and operations leadership, senior engineers, IT management, CIOs, and service providers to those industries. Academics will benefit from latest data analysis strategies. This book guides readers to use the best, results-proven approaches to ensure operational excellence.

The digital age has presented an exponential growth in the amount of data available to individuals looking to draw conclusions based on given or collected information across industries. Challenges associated with the analysis, security, sharing, storage, and visualization of large and complex data sets continue to plague data scientists and analysts alike as traditional data processing applications struggle to adequately manage big data. Big Data: Concepts, Methodologies, Tools, and Applications is a multi-volume compendium of research-based perspectives and solutions within the realm of large-scale and complex data sets. Taking a multidisciplinary approach, this publication presents exhaustive coverage of crucial topics in the field of big data including diverse applications, storage solutions, analysis techniques, and methods for searching and transferring large data sets, in addition to security issues. Emphasizing essential research in the field of data science, this publication is an ideal reference source for data analysts, IT professionals, researchers, and academics.

Machine Learning for Data Streams