

Stream Processing With Apache Flink

This book constitutes the refereed proceedings of the 6th International Conference on Big Data analytics, BDA 2018, held in Warangal, India, in December 2018. The 29 papers presented in this volume were carefully reviewed and selected from 93 submissions. The papers are organized in topical sections named: big data analytics: vision and perspectives; financial data analytics and data streams; web and social media data; big data systems and frameworks; predictive analytics in healthcare and agricultural domains; and machine learning and pattern mining.

From an engineering perspective, scalability is one of the most pressing challenges in data science. Apache Flink, the powerful and popular stream-processing platform, offers features and functionality that can help developers tackle this challenge. In this course, learn how to build a real-time stream processing pipeline with Apache Flink. Instructor Kumaran Ponnambalam begins by reviewing key streaming concepts and features of Apache Flink. He then takes a deeper look at the DataStream API and explores various capabilities available for real-time stream processing, including windowing and joins. After delving into the platform's event-time processing and state management features, he provides a use case project that allows you to put your new skills to the test.

C?C++????

Vast amounts of data are continually being generated by a wide variety of data producers. This data ranges from quantitative sensor observations produced by robot systems to complex unstructured human-generated texts on social media. With data being so abundant, the ability to make sense of these streams of data through reasoning is of great importance. Reasoning over streams is particularly relevant for autonomous robotic systems that operate in physical environments. They commonly observe this environment through incremental observations, gradually refining information about their surroundings. This makes robust management of streaming data and their refinement an important problem. Many contemporary approaches to stream reasoning focus on the issue of querying data streams in order to generate higher-level information by relying on well-known database approaches. Other approaches apply logic-based reasoning techniques, which rarely consider the provenance of their symbolic interpretations. In this work, we integrate techniques for logic-based stream reasoning with the adaptive generation of the state streams needed to do the reasoning over. This combination deals with both the challenge of reasoning over uncertain streaming data and the problem of robustly managing streaming data and their refinement. The main contributions of this work are (1) a logic-based temporal reasoning technique based on path checking under uncertainty that combines temporal reasoning with qualitative spatial reasoning; (2) an adaptive reconfiguration procedure for generating and maintaining a data stream required to perform spatio-temporal stream reasoning over; and (3) integration of these two techniques into a stream reasoning framework. The proposed spatio-temporal stream reasoning technique is able to reason with intertemporal spatial relations by leveraging landmarks. Adaptive state stream generation allows the framework to adapt to situations in which the set of available streaming resources changes. Management of streaming resources is formalised in the DyKnow model, which introduces a configuration life-cycle to adaptively generate state streams. The DyKnow-ROS stream reasoning framework is a concrete realisation of this model that extends the Robot Operating System (ROS). DyKnow-ROS has been deployed on the SoftBank Robotics NAO platform to demonstrate the system's capabilities in a case study on run-time adaptive reconfiguration. The results show that the proposed system - by combining reasoning over and reasoning about streams - can robustly

perform stream reasoning, even when the availability of streaming resources changes.

This book constitutes revised selected papers from the workshops held at 25th International Conference on Parallel and Distributed Computing, Euro-Par 2019, which took place in Göttingen, Germany, in August 2019. The 53 full papers and 10 poster papers presented in this volume were carefully reviewed and selected from 77 submissions. Euro-Par is an annual, international conference in Europe, covering all aspects of parallel and distributed processing. These range from theory to practice, from small to the largest parallel and distributed systems and infrastructures, from fundamental computational problems to full-edged applications, from architecture, compiler, language and interface design and implementation to tools, support infrastructures, and application performance aspects. Chapter "In Situ Visualization of Performance-Related Data in Parallel CFD Applications" is available open access under a Creative Commons Attribution 4.0 International License via link.springer.com.

Virtual, hands-on learning labs allow you to apply your technical skills in realistic environments. So Sybex has bundled AWS labs from XtremeLabs with our popular AWS Certified Data Analytics Study Guide to give you the same experience working in these labs as you prepare for the Certified Data Analytics Exam that you would face in a real-life application. These labs in addition to the book are a proven way to prepare for the certification and for work as an AWS Data Analyst. AWS Certified Data Analytics Study Guide: Specialty (DAS-C01) Exam is intended for individuals who perform in a data analytics-focused role. This UPDATED exam validates an examinee's comprehensive understanding of using AWS services to design, build, secure, and maintain analytics solutions that provide insight from data. It assesses an examinee's ability to define AWS data analytics services and understand how they integrate with each other; and explain how AWS data analytics services fit in the data lifecycle of collection, storage, processing, and visualization. The book focuses on the following domains: • Collection • Storage and Data Management • Processing • Analysis and Visualization • Data Security This is your opportunity to take the next step in your career by expanding and validating your skills on the AWS cloud. AWS is the frontrunner in cloud computing products and services, and the AWS Certified Data Analytics Study Guide: Specialty exam will get you fully prepared through expert content, and real-world knowledge, key exam essentials, chapter review questions, and much more. Written by an AWS subject-matter expert, this study guide covers exam concepts, and provides key review on exam topics. Readers will also have access to Sybex's superior online interactive learning environment and test bank, including chapter tests, practice exams, a glossary of key terms, and electronic flashcards. And included with this version of the book, XtremeLabs virtual labs that run from your browser. The registration code is included with the book and gives you 6 months of unlimited access to XtremeLabs AWS Certified Data Analytics Labs with 3 unique lab modules based on the book.

Definitive guide to lightning fast data processing for distributed systems with Apache Flink
About This Book* Build your expertise in processing realtime data with Apache Flink and its ecosystem* Gain insights into the working of all components of Apache Flink such as FlinkML, Gelly, and Table API
Filled with real world use cases,* Your guide to take advantage of Apache Flink for solving real world problems
Who This Book Is For
Big data developers who are looking to process batch and real-time data on distributed systems. Basic knowledge of Hadoop and big data is assumed. Reasonable knowledge of Java or Scala is expected.
What You Will Learn* Learn how to build end to end real time analytics projects* Integrate with existing big data stack and utilize existing infrastructure.* Build predictive analytics applications using FlinkML* Use graph library to perform graph querying and search.
In Detail
With the advent of massive computer systems, organizations in different domains generate large amounts of data at a realtime basis. The latest entrant to big data processing, Apache Flink, is designed to process continuous streams of data at a lightning fast pace. This book will be your definitive guide to batch and stream data processing with Apache

Read Online Stream Processing With Apache Flink

Flink. The book begins with introducing the Apache Flink ecosystem, setting it up and using the DataSet and DataStream API for processing batch and streaming datasets. Bringing the power of SQL to Flink, this book will then explore the Table API for querying and manipulating data. In the latter half of the book, readers will get to learn the remaining ecosystem of Apache Flink to achieve complex tasks such as event processing, machine learning, and graph processing. The final part of the book would consist of topics such as scaling Flink solutions, performance optimization and integrating Flink with other tools such as Elasticsearch. Whether you want to dive deeper into Apache Flink, or want to investigate how to get more out of this powerful technology, you'll find everything inside

Gain a practical introduction to DataOps, a new discipline for delivering data science at scale inspired by practices at companies such as Facebook, Uber, LinkedIn, Twitter, and eBay. Organizations need more than the latest AI algorithms, hottest tools, and best people to turn data into insight-driven action and useful analytical data products. Processes and thinking employed to manage and use data in the 20th century are a bottleneck for working effectively with the variety of data and advanced analytical use cases that organizations have today. This book provides the approach and methods to ensure continuous rapid use of data to create analytical data products and steer decision making. Practical DataOps shows you how to optimize the data supply chain from diverse raw data sources to the final data product, whether the goal is a machine learning model or other data-orientated output. The book provides an approach to eliminate wasted effort and improve collaboration between data producers, data consumers, and the rest of the organization through the adoption of lean thinking and agile software development principles. This book helps you to improve the speed and accuracy of analytical application development through data management and DevOps practices that securely expand data access, and rapidly increase the number of reproducible data products through automation, testing, and integration. The book also shows how to collect feedback and monitor performance to manage and continuously improve your processes and output.

What You Will Learn

- Develop a data strategy for your organization to help it reach its long-term goals
- Recognize and eliminate barriers to delivering data to users at scale
- Work on the right things for the right stakeholders through agile collaboration
- Create trust in data via rigorous testing and effective data management
- Build a culture of learning and continuous improvement through monitoring deployments and measuring outcomes
- Create cross-functional self-organizing teams focused on goals not reporting lines
- Build robust, trustworthy, data pipelines in support of AI, machine learning, and other analytical data products

Who This Book Is For Data science and advanced analytics experts, CIOs, CDOs (chief data officers), chief analytics officers, business analysts, business team leaders, and IT professionals (data engineers, developers, architects, and DBAs) supporting data teams who want to dramatically increase the value their organization derives from data. The book is ideal for data professionals who want to overcome challenges of long delivery time, poor data quality, high maintenance costs, and scaling difficulties in getting data science output and machine learning into customer-facing production.

Today, cloud computing, big data, and the internet of things (IoT) are becoming indubitable parts of modern information and communication systems. They cover not only information and communication technology but also all types of systems in society including within the realms of business, finance, industry, manufacturing, and management. Therefore, it is critical to remain up-to-date on the latest advancements and applications, as well as current issues and challenges. The Handbook of Research on Cloud Computing and Big Data Applications in IoT is a pivotal reference source that provides relevant theoretical frameworks and the latest empirical research findings on principles, challenges,

and applications of cloud computing, big data, and IoT. While highlighting topics such as fog computing, language interaction, and scheduling algorithms, this publication is ideally designed for software developers, computer engineers, scientists, professionals, academicians, researchers, and students.

This book constitutes the workshop proceedings of the 18th International Conference on Algorithms and Architectures for Parallel Processing, ICA3PP 2018, held in Guangzhou, China, in November 2018. The 24 full papers presented were carefully selected and reviewed from numerous submissions to the two following workshops: - ICA3PP 2018 Workshop on Intelligent Algorithms for Large-scale Complex Optimization Problems - ICA3PP 2018 Workshop on Security and Privacy in Data Processing

Stream processing is becoming something like a "grand unifying paradigm" for data processing. Outgrowing its original space of real-time data processing, stream processing is becoming a technology that offers new approaches to data processing (including batch processing), real-time applications, and even distributed transactions. Stephan Ewen (Ververica) dives into these developments from the view of Apache Flink and presents some of the major efforts in the Flink community to build a unified stream processor data processing and data-driven applications. Flink already powers many of the world's most demanding stream processing applications. He explores the approach of Flink's next-generation streaming runtime that offers a state-of-the-art batch processing experience and performance. A new machine learning library, built on top of a unique new API, supports many algorithms to train dynamically across static and real-time data. And he examines the new building blocks stream processing offers for data-driven applications that open a new direction to solve application consistency. You'll see use cases from different users, showing how companies apply this broader streaming paradigm in practice. With use cases from different users, we show how companies apply this broader streaming paradigm in practice. Prerequisite knowledge A basic understanding of stream processing (useful but not required) What you'll learn Understand the breadth of changes that stream processing is introducing to data infrastructures and to the way we build applications and which areas are most likely to be affected next This session is from the 2019 O'Reilly Strata Conference in New York, NY.

This book constitutes the proceedings of the 8th International Conference on Big Data Analytics, BDA 2020, which took place during December 15-18, 2020, in Sonapat, India. The 11 full and 3 short papers included in this volume were carefully reviewed and selected from 48 submissions; the book also contains 4 invited and 3 tutorial papers. The contributions were organized in topical sections named as follows: data science systems; data science architectures; big data analytics in healthcare; information interchange of Web data resources; and business analytics.

Learn how today's businesses can transform themselves by leveraging real-time data and advanced machine learning

analytics. This book provides prescriptive guidance for architects and developers on the design and development of modern Internet of Things (IoT) and Advanced Analytics solutions. In addition, *Business in Real-Time Using Azure IoT and Cortana Intelligence Suite* offers patterns and practices for those looking to engage their customers and partners through Software-as-a-Service solutions that work on any device. Whether you're working in Health & Life Sciences, Manufacturing, Retail, Smart Cities and Buildings or Process Control, there exists a common platform from which you can create your targeted vertical solutions. *Business in Real-Time Using Azure IoT and Cortana Intelligence Suite* uses a reference architecture as a road map. Building on Azure's PaaS services, you'll see how a solution architecture unfolds that demonstrates a complete end-to-end IoT and Advanced Analytics scenario. What You'll Learn: Automate your software product life cycle using PowerShell, Azure Resource Manager Templates, and Visual Studio Team Services Implement smart devices using Node.JS and C# Use Azure Streaming Analytics to ingest millions of events Provide both "Hot" and "Cold" path outputs for real-time alerts, data transformations, and aggregation analytics Implement batch processing using Azure Data Factory Create a new form of Actionable Intelligence (AI) to drive mission critical business processes Provide rich Data Visualizations across a wide variety of mobile and web devices Who This Book is For: Solution Architects, Software Developers, Data Architects, Data Scientists, and CIO/CTA Technical Leadership Professionals

Data has cemented itself as a building block of daily life. However, surrounding oneself with great quantities of information heightens risks to one's personal privacy. Additionally, the presence of massive amounts of information prompts researchers into how best to handle and disseminate it. Research is necessary to understand how to cope with the current technological requirements. *Large-Scale Data Streaming, Processing, and Blockchain Security* is a collection of innovative research that explores the latest methodologies, modeling, and simulations for coping with the generation and management of large-scale data in both scientific and individual applications. Featuring coverage on a wide range of topics including security models, internet of things, and collaborative filtering, this book is ideally designed for entrepreneurs, security analysts, IT consultants, security professionals, programmers, computer technicians, data scientists, technology developers, engineers, researchers, academicians, and students.

A comprehensive guide to mastering the most advanced Hadoop 3 concepts Key Features Get to grips with the newly introduced features and capabilities of Hadoop 3 Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystem Sharpen your Hadoop skills with real-world case studies and code Book Description Apache Hadoop is one of the most popular big data solutions for distributed storage and for processing large chunks of data. With Hadoop 3, Apache promises to provide a high-performance, more fault-tolerant, and highly efficient big data

processing platform, with a focus on improved scalability and increased efficiency. With this guide, you'll understand advanced concepts of the Hadoop ecosystem tool. You'll learn how Hadoop works internally, study advanced concepts of different ecosystem tools, discover solutions to real-world use cases, and understand how to secure your cluster. It will then walk you through HDFS, YARN, MapReduce, and Hadoop 3 concepts. You'll be able to address common challenges like using Kafka efficiently, designing low latency, reliable message delivery Kafka systems, and handling high data volumes. As you advance, you'll discover how to address major challenges when building an enterprise-grade messaging system, and how to use different stream processing systems along with Kafka to fulfil your enterprise goals. By the end of this book, you'll have a complete understanding of how components in the Hadoop ecosystem are effectively integrated to implement a fast and reliable data pipeline, and you'll be equipped to tackle a range of real-world problems in data pipelines. What you will learn

- Gain an in-depth understanding of distributed computing using Hadoop 3
- Develop enterprise-grade applications using Apache Spark, Flink, and more
- Build scalable and high-performance Hadoop data pipelines with security, monitoring, and data governance
- Explore batch data processing patterns and how to model data in Hadoop Master best practices for enterprises using, or planning to use, Hadoop 3 as a data platform
- Understand security aspects of Hadoop, including authorization and authentication

Who this book is for If you want to become a big data professional by mastering the advanced concepts of Hadoop, this book is for you. You'll also find this book useful if you're a Hadoop professional looking to strengthen your knowledge of the Hadoop ecosystem. Fundamental knowledge of the Java programming language and basics of Hadoop is necessary to get started with this book.

This book constitutes revised selected papers from the workshops held at 24th International Conference on Parallel and Distributed Computing, Euro-Par 2018, which took place in Turin, Italy, in August 2018. The 64 full papers presented in this volume were carefully reviewed and selected from 109 submissions. Euro-Par is an annual, international conference in Europe, covering all aspects of parallel and distributed processing. These range from theory to practice, from small to the largest parallel and distributed systems and infrastructures, from fundamental computational problems to full-edged applications, from architecture, compiler, language and interface design and implementation to tools, support infrastructures, and application performance aspects.

There are many frameworks that try to process Big Data in real time such as Apache Spark, Apache Flink, and Apache Beam. The main purpose of this research is to give a clear and fair comparison among the mentioned frameworks from different perspectives such as the latency, processing guarantees, the accuracy of results, fault tolerance, and the available functionalities of each framework. Keywords: Big Data, Stream processing frameworks, Real-time analytic, Apache Spark, Apache Flink, Google Cloud Dataflow, Apache Beam, Lambda

Read Online Stream Processing With Apache Flink

Architecture

Data engineering is the foundation for enabling analytics and data science applications in the world of big data. It requires building scalable data processing pipelines and delivering them in short time frames. Apache Flink, the powerful and popular stream-processing platform, was designed to help you achieve these goals. In this course, join Kumaran Ponnambalam as he focuses on how to build batch mode data pipelines with Apache Flink. Kumaran kicks off the course by reviewing the features and architecture of Apache Flink. He then takes a deeper look at the DataSet API and explores various capabilities available for transforming, aggregating, and combining data. To wrap up the course, he presents a use case project that allows you to leverage your new skills.

With the advancements of semantic web, ontology has become the crucial mechanism for representing concepts in various domains. For research and dispersal of customized healthcare services, a major challenge is to efficiently retrieve and analyze individual patient data from a large volume of heterogeneous data over a long time span. This requirement demands effective ontology-based information retrieval approaches for clinical information systems so that the pertinent information can be mined from large amount of distributed data. This unique and groundbreaking book highlights the key advances in ontology-based information retrieval techniques being applied in the healthcare domain and covers the following areas: Semantic data integration in e-health care systems Keyword-based medical information retrieval Ontology-based query retrieval support for e-health implementation Ontologies as a database management system technology for medical information retrieval Information integration using contextual knowledge and ontology merging Collaborative ontology-based information indexing and retrieval in health informatics An ontology-based text mining framework for vulnerability assessment in health and social care An ontology-based multi-agent system for matchmaking patient healthcare monitoring A multi-agent system for querying heterogeneous data sources with ontologies for reducing cost of customized healthcare systems A methodology for ontology based multi agent systems development Ontology based systems for clinical systems: validity, ethics and regulation

There's growing interest in learning how to analyze streaming data in large-scale systems such as web traffic, financial transactions, machine logs, industrial sensors, and many others. But analyzing data streams at scale has been difficult to do well—until now. This practical book delivers a deep introduction to Apache Flink, a highly innovative open source stream processor with a surprising range of capabilities. Authors Ellen Friedman and Kostas Tzoumas show technical and nontechnical readers alike how Flink is engineered to overcome significant tradeoffs that have limited the effectiveness of other approaches to stream processing. You'll also learn how Flink has the ability to handle both stream and batch data processing with one technology. Learn the consequences of not doing streaming well—in retail and marketing, IoT, telecom, and banking and finance Explore how to design data architecture to gain the best advantage from stream processing Get an overview of Flink's capabilities and features, along with examples of how companies use Flink, including in production Take a technical dive into Flink, and learn how it handles time and stateful computation Examine how Flink processes both streaming (unbounded) and batch (bounded) data without sacrificing performance

This book constitutes the refereed proceedings of the 19th International Conference on Big Data Analytics and Knowledge Discovery, DaWaK 2017, held in Lyon, France, in August 2017. The 24 revised full papers and 11 short papers presented were carefully reviewed and selected from 97 submissions. The papers are organized in the following topical sections: new generation data warehouses design; cloud and NoSQL databases; advanced programming paradigms; non-functional requirements satisfaction; machine learning; social media and twitter analysis; sentiment analysis and user influence; knowledge discovery; and data flow management and optimization.

how this technology differs from traditional batch data processing. Longtime Apache Flink committers Fabian Hueske and Vasia Kalavri show you how to implement scalable streaming applications with Flink's DataStream API and continuously run and maintain these applications in operational environments. Stream processing is ideal for many use cases, including low-latency ETL, streaming analytics, and real-time dashboards as well as fraud detection, anomaly detection, and alerting. You can process continuous data of any kind, including user interactions, financial transactions, and IoT data, as soon as you generate them. Learn concepts and challenges of distributed stateful stream processing Explore Flink's system architecture, including its event-time processing mode and fault-tolerance model Understand the fundamentals and building blocks of the DataStream API, including its time-based and stateful operators Read data from and write data to external systems with exactly-once consistency Deploy and configure Flink clusters Operate continuously running streaming applications

This book covers the theory, design and applications of computer networks, distributed computing and information systems. Networks of today are going through a rapid evolution, and there are many emerging areas of information networking and their applications. Heterogeneous networking supported by recent technological advances in low-power wireless communications along with silicon integration of various functionalities such as sensing, communications, intelligence and actuations is emerging as a critically important disruptive computer class based on a new platform, networking structure and interface that enable novel, low-cost and high-volume applications. Several of such applications have been difficult to realize because of many interconnections problems. To fulfill their large range of applications, different kinds of networks need to collaborate, and wired and next-generation wireless systems should be integrated in order to develop high-performance computing solutions to problems arising from the complexities of these networks. The aim of the book "Advanced Information Networking and Applications" is to provide latest research findings, innovative research results, methods and development techniques from both theoretical and practical perspectives related to the emerging areas of information networking and applications.

Stream Processing with Apache Flink Fundamentals, Implementation, and Operation of Streaming Applications O'Reilly Media

More and more data-driven companies are looking to adopt stream processing and streaming analytics. With this concise ebook, you'll learn best practices for designing a reliable architecture that supports this emerging big-data paradigm. Authors Ted Dunning and Ellen Friedman (Real World Hadoop) help you explore some of the best technologies to handle stream processing and analytics, with a focus on the upstream queuing or message-passing layer. To illustrate the effectiveness of these technologies, this book also includes specific use cases. Ideal for developers and non-technical

people alike, this book describes: Key elements in good design for streaming analytics, focusing on the essential characteristics of the messaging layer New messaging technologies, including Apache Kafka and MapR Streams, with links to sample code Technology choices for streaming analytics: Apache Spark Streaming, Apache Flink, Apache Storm, and Apache Apex How stream-based architectures are helpful to support microservices Specific use cases such as fraud detection and geo-distributed data streams Ted Dunning is Chief Applications Architect at MapR Technologies, and active in the open source community. He currently serves as VP for Incubator at the Apache Foundation, as a champion and mentor for a large number of projects, and as committer and PMC member of the Apache ZooKeeper and Drill projects. Ted is on Twitter as @ted_dunning. Ellen Friedman, a committer for the Apache Drill and Apache Mahout projects, is a solutions consultant and well-known speaker and author, currently writing mainly about big data topics. With a PhD in Biochemistry, she has years of experience as a research scientist and has written about a variety of technical topics. Ellen is on Twitter as @Ellen_Friedman.

Before you can build analytics tools to gain quick insights, you first need to know how to process data in real time. With this practical guide, developers familiar with Apache Spark will learn how to put this in-memory framework to use for streaming data. You'll discover how Spark enables you to write 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

This book provides detailed descriptions of big data solutions for activity detection and forecasting of very large numbers of moving entities spread across large geographical areas. It presents state-of-the-art methods for processing, managing, detecting and predicting trajectories and important events related to moving entities, together with advanced visual analytics methods, over multiple heterogeneous, voluminous, fluctuating and noisy data streams from moving entities, correlating them with data from archived data sources expressing e.g. entities' characteristics, geographical information, mobility patterns, mobility regulations and intentional data. The book is divided into six parts: Part I discusses the motivation and background of mobility forecasting supported by trajectory-oriented analytics, and includes specific

problems and challenges in the aviation (air-traffic management) and the maritime domains. Part II focuses on big data quality assessment and processing, and presents novel technologies suitable for mobility analytics components. Next, Part III describes solutions toward processing and managing big spatio-temporal data, particularly enriching data streams and integrating streamed and archival data to provide coherent views of mobility, and storing of integrated mobility data in large distributed knowledge graphs for efficient query-answering. Part IV focuses on mobility analytics methods exploiting (online) processed, synopsisized and enriched data streams as well as (offline) integrated, archived mobility data, and highlights future location and trajectory prediction methods, distinguishing between short-term and more challenging long-term predictions. Part V examines how methods addressing data management, data processing and mobility analytics are integrated in big data architectures with distinctive characteristics compared to other known big data paradigmatic architectures. Lastly, Part VI covers important ethical issues that research on mobility analytics should address. Providing novel approaches and methodologies related to mobility detection and forecasting needs based on big data exploration, processing, storage, and analysis, this book will appeal to computer scientists and stakeholders in various application domains.

This book constitutes refereed proceedings of the 4th International Workshop on Software Foundations for Data Interoperability, SFDI 2020, and 2nd International Workshop on Large Scale Graph Data Analytics, LSGDA 2020, held in Conjunction with VLDB 2020, in September 2020. Due to the COVID-19 pandemic the conference was held online. The 11 full papers and 4 short papers were thoroughly reviewed and selected from 38 submissions. The volume presents original research and application papers on the development of novel graph analytics models, scalable graph analytics techniques and systems, data integration, and data exchange.

This book constitutes the proceedings of the 22nd European Conference on Advances in Databases and Information Systems, ADBIS 2018, held in Budapest, Hungary, in September 2018. The 17 regular papers presented together with two invited papers were carefully selected and reviewed from numerous submissions. The papers are organized in topical sections such as information extraction and integration; data mining and knowledge discovery; indexing, query processing and optimization; data quality and data cleansing; distributed data platforms, including cloud data systems, key-value stores, and big data systems; and streaming data analysis; web, XML and semi-structured databases.

The four-volume set LNCS 11334-11337 constitutes the proceedings of the 18th International Conference on Algorithms and Architectures for Parallel Processing, ICA3PP 2018, held in Guangzhou, China, in November 2018. The 141 full and 50 short papers presented were carefully reviewed and selected from numerous submissions. The papers are organized in topical sections on Distributed and Parallel Computing; High Performance Computing; Big Data and Information

Processing; Internet of Things and Cloud Computing; and Security and Privacy in Computing.

This book constitutes the refereed proceedings of the 21st IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems, DAIS 2021, held in Valletta, Malta, in June 2021, as part of the 16th International Federated Conference on Distributed Computing Techniques, DisCoTec 2021. The 7 regular papers and 3 short papers presented in this book were carefully reviewed and selected from 15 submissions. DAIS addresses all practical and conceptual aspects of distributed applications, including their design, modeling, implementation and operation, the supporting middleware, appropriate software engineering methodologies and tools, as well as experimental studies and applications.

[Copyright: 1364117a728df23faead8618d226195f](https://doi.org/10.1007/978-1-364-117a-7_28)