Free download Download for pro hadoop by jason venner (Download Only)

Pro Hadoop Pro Hadoop Data Analytics Professional Hadoop Professional Hadoop Solutions Pro Apache Hadoop Professional Hadoop Pro Microsoft HDInsight Hadoop: The Definitive Guide Big Data Analytics with Hadoop 3 Pro Apache Phoenix Big Data Analytics with R and Hadoop The Impact of Digital Transformation and FinTech on the Finance Professional Apache Hadoop YARN Optimizing Hadoop for MapReduce Professional NoSQL Practical Ext JS 4 Modern Big Data Processing with Hadoop Hadoop in Action Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data Official Google Cloud Certified Professional Data Engineer Study Guide Data Analytics with Hadoop Proceedings of the International Conference on Systems, Science, Control, Communication, Engineering and Technology 2015 Practical Hadoop Ecosystem Hadoop 2 Quick-Start Guide Implementing Big Data Analytics Using Hadoop Architecting Modern Data Platforms Beginning Hadoop Hadoop in Practice Practical Data Science with Hadoop and Spark Pro Docker Hadoop Practice Guide Big Data Analytics Beyond Hadoop Information Technology - New Generations Field Guide to Hadoop Securing Hadoop Hadoop For Dummies Pro Tableau Apache Oozie BIG DATA AND HADOOP Beginning Apache Hadoop Administration

Pro Hadoop 2009-08-09 you ve heard the hype about hadoop it runs petabyte scale data mining tasks insanely fast it runs gigantic tasks on clouds for absurdly cheap it s been heavily committed to by tech giants like ibm yahoo and the apache project and it s completely open source thus free but what exactly is it and more importantly how do you even get a hadoop cluster up and running from apress the name you ve come to trust for hands on technical knowledge pro hadoop brings you up to speed on hadoop you learn the ins and outs of mapreduce how to structure a cluster design and implement the hadoop file system and how to build your first cloud computing tasks using hadoop learn how to let hadoop take care of distributing and parallelizing your software you just focus on the code hadoop takes care of the rest best of all you ll learn from a tech professional who s been in the hadoop scene since day one written from the perspective of a principal engineer with down in the trenches knowledge of what to do wrong with hadoop you learn how to avoid the common expensive first errors that everyone makes with creating their own hadoop system or inheriting someone else s skip the novice stage and the expensive hard to fix mistakes go straight to seasoned pro on the hottest cloud computing framework with pro hadoop your productivity will blow your managers away

Pro Hadoop Data Analytics 2016-12-29 learn advanced analytical techniques and leverage existing tool kits to make your analytic applications more powerful precise and efficient this book provides the right combination of architecture design and implementation information to create analytical systems that go beyond the basics of classification clustering and recommendation pro hadoop data analytics emphasizes best practices to ensure coherent efficient development a complete example system will be developed using standard third party components that consist of the tool kits libraries visualization and reporting code as well as support glue to provide a working and extensible end to end system the book also highlights the importance of end to end flexible configurable high performance data pipeline systems with analytical components as well as appropriate visualization results you ll discover the importance of mix and match or hybrid systems using different analytical components in one application this hybrid approach will be prominent in the examples what you ll learn build big data analytic systems with the hadoop ecosystem use libraries tool kits and algorithms to make development easier and more effective apply metrics to measure performance and efficiency of components and systems connect to standard relational databases nosql data sources and more follow case studies with example components to create your own systems who this book is for software engineers architects and data scientists with an interest in the design and implementation of big data analytical systems using hadoop the hadoop ecosystem and other associated technologies

Professional Hadoop 2016-05-23 the professional s one stop guide to this open source java based big data framework professional hadoop is the complete reference and resource for experienced developers looking to employ apache hadoop in real world settings written by an expert team of certified hadoop developers committers and summit speakers this book details every key aspect of hadoop technology to enable optimal processing of large data sets designed expressly for the professional developer this book skips over the basics of database development to get you acquainted with the framework s processes and capabilities right away the discussion covers each key hadoop component individually culminating in a sample application that brings all of the pieces together to illustrate the cooperation and interplay that make hadoop a major big data solution coverage includes everything from storage and security to computing and user experience with expert guidance on integrating other software and more hadoop is quickly reaching significant market usage and more and more developers are being called upon to develop big data solutions using the hadoop framework this book covers the process from beginning to end providing a crash course for professionals needing to learn and apply hadoop quickly configure storage ue and in memory computing integrate hadoop with other programs including kafka and storm master the fundamentals of apache big top and ignite build robust data security with expert tips and advice hadoop s popularity is largely due to its accessibility open source and written in java the framework offers almost no barrier to entry for experienced database

developers already familiar with the skills and requirements real world programming entails professional hadoop gives you the practical information and framework specific skills you need quickly

Professional Hadoop Solutions 2013-09-12 the go to guidebook for deploying big data solutions withhadoop today s enterprise architects need to understand how the hadoopframeworks and apis fit together and how they can be integrated todeliver real world solutions this book is a practical detailedguide to building and implementing those solutions with code levelinstruction in the popular wrox tradition it covers storing datawith hdfs and hbase processing data with mapreduce and automatingdata processing with oozie hadoop security running hadoop withamazon services best practices and automating hadoopprocesses in real time are also covered in depth with in depth code examples in java and xml and the latest onrecent additions to the hadoop ecosystem this complete resourcealso covers the use of apis exposing their inner workings andallowing architects and developers to better leverage and customizethem the ultimate guide for developers designers and architectswho need to build and deploy hadoop applications covers storing and processing data with various technologies automating data processing hadoop security and deliveringreal time solutions includes detailed real world examples and code levelguidelines explains when why and how to use these tools effectively written by a team of hadoop experts in theprogrammer to programmer wrox style professional hadoop solutions is the reference enterprisearchitects and developers need to maximize the power of hadoop

Pro Apache Hadoop 2014 pro apache hadoop second edition brings you up to speed on hadoop the framework of big data revised to cover hadoop 2 0 the book covers the very latest developments such as yarn aka mapreduce 2 0 new hdfs high availability features and increased scalability in the form of hdfs federations all the old content has been revised too giving the latest on the ins and outs of mapreduce cluster design the hadoop distributed file system and more this book covers everything you need to build your first hadoop cluster and begin analyzing and deriving value from your business and scientific data learn to solve big data problems the mapreduce way by breaking a big problem into chunks and creating small scale solutions that can be flung across thousands upon thousands of nodes to analyze large data volumes in a short amount of wall clock time learn how to let hadoop take care of distributing and parallelizing your softwareyou just focus on the code hadoop takes care of the rest covers all that is new in hadoop 2 0 written by a professional involved in hadoop since day one takes you quickly to the seasoned pro level on the hottest cloud computing framework

Professional Hadoop 2017-06-07 professional hadoop is the complete reference and resource for experienced developers looking to employ apache hadoop in real world settings written by an expert team of certified hadoop developers committers and summit speakers this book details every key aspect of hadoop technology to enable optimal processing of large data sets designed expressly for the professional developer this book skips over the basics of database development to get you acquainted with the framework s processes and capabilities right away the discussion covers each key hadoop component individually culminating in a sample application that brings all of the pieces together to illustrate the cooperation and interplay that make hadoop a major big data solution coverage includes everything from storage and security to computing and user experience with expert guidance on integrating other software and more

Pro Microsoft HDInsight 2014-02-18 the expert s voice in big data cover

Hadoop: The Definitive Guide 2012-05-10 ready to unlock the power of your data with this comprehensive guide you ll learn how to build and maintain reliable scalable distributed systems with apache hadoop this book is ideal for programmers looking to analyze datasets of any size and for administrators who want to set up and run hadoop clusters you ll find illuminating case studies that demonstrate how hadoop is used to solve specific problems this third edition covers recent changes to hadoop including material on the new mapreduce api as well as mapreduce 2 and its more flexible execution model yarn store large datasets with the hadoop distributed file system hdfs run distributed

computations with mapreduce use hadoop s data and i o building blocks for compression data integrity serialization including avro and persistence discover common pitfalls and advanced features for writing real world mapreduce programs design build and administer a dedicated hadoop cluster or run hadoop in the cloud load data from relational databases into hdfs using sqoop perform large scale data processing with the pig query language analyze datasets with hive hadoop s data warehousing system take advantage of hbase for structured and semi structured data and zookeeper for building distributed systems

Big Data Analytics with Hadoop 3 2018-05-31 explore big data concepts platforms analytics and their applications using the power of hadoop 3 key features learn hadoop 3 to build effective big data analytics solutions on premise and on cloud integrate hadoop with other big data tools such as r python apache spark and apache flink exploit big data using hadoop 3 with real world examples book description apache hadoop is the most popular platform for big data processing and can be combined with a host of other big data tools to build powerful analytics solutions big data analytics with hadoop 3 shows you how to do just that by providing insights into the software as well as its benefits with the help of practical examples once you have taken a tour of hadoop 3 s latest features you will get an overview of hdfs mapreduce and yarn and how they enable faster more efficient big data processing you will then move on to learning how to integrate hadoop with the open source tools such as python and r to analyze and visualize data and perform statistical computing on big data as you get acquainted with all this you will explore how to use hadoop 3 with apache spark and apache flink for real time data analytics and stream processing in addition to this you will understand how to use hadoop to build analytics solutions on the cloud and an end to end pipeline to perform big data analysis using practical use cases by the end of this book you will be well versed with the analytical capabilities of the hadoop ecosystem you will be able to build powerful solutions to perform big data analytics and get insight effortlessly what you will learn explore the new features of hadoop 3 along with hdfs yarn and mapreduce get well versed with the analytical capabilities of hadoop ecosystem using practical examples integrate hadoop with r and python for more efficient big data processing learn to use hadoop with apache spark and apache flink for real time data analytics set up a hadoop cluster on aws cloud perform big data analytics on aws using elastic map reduce who this book is for big data analytics with hadoop 3 is for you if you are looking to build high performance analytics solutions for your enterprise or business using hadoop 3 s powerful features or you re new to big data analytics a basic understanding of the java programming language is required

Pro Apache Phoenix 2016-12-29 leverage phoenix as an ansi sql engine built on top of the highly distributed and scalable nosql framework hbase learn the basics and best practices that are being adopted in phoenix to enable a high write and read throughput in a big data space this book includes real world cases such as internet of things devices that send continuous streams to phoenix and the book explains how key features such as joins indexes transactions and functions help you understand the simple flexible and powerful api that phoenix provides examples are provided using real time data and data driven businesses that show you how to collect analyze and act in seconds pro apache phoenix covers the nuances of setting up a distributed hbase cluster with phoenix libraries running performance benchmarks configuring parameters for production scenarios and viewing the results the book also shows how phoenix plays well with other key frameworks in the hadoop ecosystem such as apache spark pig flume and sqoop you will learn how to handle a petabyte data store by applying familiar sql techniques store analyze and manipulate data in a nosql hadoop echo system with hbase apply best practices while working with a scalable data store on hadoop and hbase integrate popular frameworks apache spark pig flume to simplify big data analysis demonstrate real time use cases and big data modeling techniques who this book is for data engineers big data administrators and architects

Big Data Analytics with R and Hadoop 2013 big data analytics with r and hadoop is a tutorial style book that focuses on all the powerful big data tasks that can be achieved by integrating r and hadoop this book is ideal for r developers

who are looking for a way to perform big data analytics with hadoop this book is also aimed at those who know hadoop and want to build some intelligent applications over big data with r packages it would be helpful if readers have basic knowledge of r

<u>The Impact of Digital Transformation and FinTech on the Finance Professional</u> 2019-10-02 this book demystifies the developments and defines the buzzwords in the wide open space of digitalization and finance exploring the space of fintech through the lens of the financial services professional and what they need to know to stay ahead with chapters focusing on the customer interface payments smart contracts workforce automation robotics crypto currencies and beyond this book aims to be the go to guide for professionals in financial services and banking on how to better understand the digitalization of their industry the book provides an outlook of the impact digitalization will have in the daily work of a cfo cro and a structural influence to the financial management including risk management department of a bank

Apache Hadoop YARN 2014-03-14 this book is a critically needed resource for the newly released apache hadoop 2 0 highlighting yarn as the significant breakthrough that broadens hadoop beyond the mapreduce paradigm from the foreword by raymie stata ceo of altiscale the insider s guide to building distributed big data applications with apache hadoop yarn apache hadoop is helping drive the big data revolution now its data processing has been completely overhauled apache hadoop yarn provides resource management at data center scale and easier ways to create distributed applications that process petabytes of data and now in apache hadoop yarn two hadoop technical leaders show you how to develop new applications and adapt existing code to fully leverage these revolutionary advances yarn project founder arun murthy and project lead vinod kumar vavilapalli demonstrate how yarn increases scalability and cluster utilization enables new programming models and services and opens new options beyond java and batch processing they walk you through the entire yarn project lifecycle from installation through deployment you ll find many examples drawn from the authors cutting edge experience first as hadoop s earliest developers and implementers at yahoo and now as hortonworks developers moving the platform forward and helping customers succeed with it coverage includes yarn s goals design architecture and components how it expands the apache hadoop ecosystem exploring yarn on a single node administering yarn clusters and capacity scheduler running existing mapreduce applications developing a large scale clustered yarn application discovering new open source frameworks that run under yarn

Optimizing Hadoop for MapReduce 2014-02-21 this book is an example based tutorial that deals with optimizing hadoop for mapreduce job performance if you are a hadoop administrator developer mapreduce user or beginner this book is the best choice available if you wish to optimize your clusters and applications having prior knowledge of creating mapreduce applications is not necessary but will help you better understand the concepts and snippets of mapreduce class template code

Professional NoSQL 2011-08-31 a hands on guide to leveraging nosql databases nosql databases are an efficient and powerful tool for storing and manipulating vast quantities of data most nosql databases scale well as data grows in addition they are often malleable and flexible enough to accommodate semi structured and sparse data sets this comprehensive hands on guide presents fundamental concepts and practical solutions for getting you ready to use nosql databases expert author shashank tiwari begins with a helpful introduction on the subject of nosql explains its characteristics and typical uses and looks at where it fits in the application stack unique insights help you choose which nosql solutions are best for solving your specific data storage needs professional nosql demystifies the concepts that relate to nosql databases including column family oriented stores key value databases and document databases delves into installing and configuring a number of nosql products and the hadoop family of products explains ways of storing accessing and querying data in nosql databases through examples that use mongodb hbase cassandra redis couchdb google app engine datastore and more looks at architecture and internals provides guidelines for optimal usage performance tuning and scalable

configurations presents a number of tools and utilities relating to nosql distributed platforms and scalable processing including hive pig rrdtool nagios and more

Practical Ext JS 4 2013-09-20 practical ext js 4 will get you up and running using ext js 4 2 for your projects as quickly as possible after a quick refresher on some javascript basics you will get to grips with ext js 4 s oo concepts such as mixins and familiarize yourself with its ui components and layout you ll learn all the core features of the ext js framework such as its mvc architecture theming and styling your applications and displaying data through components such as grids trees and charts you ll use the ext js components and create an entire application from scratch by following the many practical examples finally you ll learn about unit testing and packaging to build and deploy better applications provides you with a solid knowledge of the building blocks of ext js 4 takes you through developing applications using the mvc architecture demonstrates extending the ui with custom components and plugins shows you how to unit test ext js 4 applications with jasmine and deploy them with sencha cmd completely up to date for the latest ext js 4 2

Modern Big Data Processing with Hadoop 2018-03-30 a comprehensive guide to design build and execute effective big data strategies using hadoop key features get an in depth view of the apache hadoop ecosystem and an overview of the architectural patterns pertaining to the popular big data platform conquer different data processing and analytics challenges using a multitude of tools such as apache spark elasticsearch tableau and more a comprehensive step by step guide that will teach you everything you need to know to be an expert hadoop architect book description the complex structure of data these days requires sophisticated solutions for data transformation to make the information more accessible to the users this book empowers you to build such solutions with relative ease with the help of apache hadoop along with a host of other big data tools this book will give you a complete understanding of the data lifecycle management with hadoop followed by modeling of structured and unstructured data in hadoop it will also show you how to design real time streaming pipelines by leveraging tools such as apache spark and build efficient enterprise search solutions using elasticsearch you will learn to build enterprise grade analytics solutions on hadoop and how to visualize your data using tools such as apache superset this book also covers techniques for deploying your big data solutions on the cloud apache ambari as well as expert techniques for managing and administering your hadoop cluster by the end of this book you will have all the knowledge you need to build expert big data systems what you will learn build an efficient enterprise big data strategy centered around apache hadoop gain a thorough understanding of using hadoop with various big data frameworks such as apache spark elasticsearch and more set up and deploy your big data environment on premises or on the cloud with apache ambari design effective streaming data pipelines and build your own enterprise search solutions utilize the historical data to build your analytics solutions and visualize them using popular tools such as apache superset plan set up and administer your hadoop cluster efficiently who this book is for this book is for big data professionals who want to fast track their career in the hadoop industry and become an expert big data architect project managers and mainframe professionals looking forward to build a career in big data hadoop will also find this book to be useful some understanding of hadoop is required to get the best out of this book

Hadoop in Action 2010-11-30 hadoop in action teaches readers how to use hadoop and write mapreduce programs the intended readers are programmers architects and project managers who have to process large amounts of data offline hadoop in action will lead the reader from obtaining a copy of hadoop to setting it up in a cluster and writing data analytic programs the book begins by making the basic idea of hadoop and mapreduce easier to grasp by applying the default hadoop installation to a few easy to follow tasks such as analyzing changes in word frequency across a body of documents the book continues through the basic concepts of mapreduce applications developed using hadoop including a close look at framework components use of hadoop for a variety of data analysis tasks and numerous examples of hadoop in action hadoop in action will explain how to

use hadoop and present design patterns and practices of programming mapreduce mapreduce is a complex idea both conceptually and in its implementation and hadoop users are challenged to learn all the knobs and levers for running hadoop this book takes you beyond the mechanics of running hadoop teaching you to write meaningful programs in a mapreduce framework this book assumes the reader will have a basic familiarity with java as most code examples will be written in java familiarity with basic statistical concepts e g histogram correlation will help the reader appreciate the more advanced data processing examples purchase of the print book comes with an offer of a free pdf epub and kindle ebook from manning also available is all code from the book

Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data 2011-10-22 big data represents a new era in data exploration and utilization and ibm is uniquely positioned to help clients navigate this transformation this book reveals how ibm is leveraging open source big data technology infused with ibm technologies to deliver a robust secure highly available enterprise class big data platform the three defining characteristics of big data volume variety and velocity are discussed you ll get a primer on hadoop and how ibm is hardening it for the enterprise and learn when to leverage ibm infosphere biginsights big data at rest and ibm infosphere streams big data in motion technologies industry use cases are also included in this practical guide learn how ibm hardens hadoop for enterprise class scalability and reliability gain insight into ibm s unique in motion and at rest big data analytics platform learn tips and tricks for big data use cases and solutions get a quick hadoop primer

Official Google Cloud Certified Professional Data Engineer Study Guide 2020-05-11 the proven study guide that prepares you for this new google cloud exam the google cloud certified professional data engineer study guide provides everything you need to prepare for this important exam and master the skills necessary to land that coveted google cloud professional data engineer certification beginning with a pre book assessment quiz to evaluate what you know before you begin each chapter features exam objectives and review questions plus the online learning environment includes additional complete practice tests written by dan sullivan a popular and experienced online course author for machine learning big data and cloud topics google cloud certified professional data engineer study guide is your ace in the hole for deploying and managing analytics and machine learning applications build and operationalize storage systems pipelines and compute infrastructure understand machine learning models and learn how to select pre built models monitor and troubleshoot machine learning models design analytics and machine learning applications that are secure scalable and highly available this exam guide is designed to help you develop an in depth understanding of data engineering and machine learning on google cloud platform

Data Analytics with Hadoop 2016-06-01 ready to use statistical and machine learning techniques across large data sets this practical guide shows you why the hadoop ecosystem is perfect for the job instead of deployment operations or software development usually associated with distributed computing you ll focus on particular analyses you can build the data warehousing techniques that hadoop provides and higher order data workflows this framework can produce data scientists and analysts will learn how to perform a wide range of techniques from writing mapreduce and spark applications with python to using advanced modeling and data management with spark mllib hive and hbase you ll also learn about the analytical processes and data systems available to build and empower data products that can handle and actually require huge amounts of data understand core concepts behind hadoop and cluster computing use design patterns and parallel analytical algorithms to create distributed data analysis jobs learn about data management mining and warehousing in a distributed context using apache hive and hbase use sqoop and apache flume to ingest data from relational databases program complex hadoop and spark applications with apache pig and spark dataframes perform machine learning techniques such as classification clustering and collaborative filtering with spark s mllib

Proceedings of the International Conference on Systems, Science, Control, Communication, Engineering and Technology 2015 2015-08-10 icssccet 2015 will be

the most comprehensive conference focused on the various aspects of advances in systems science management medical sciences communication engineering technology interdisciplinary research theory and technology this conference provides a chance for academic and industry professionals to discuss recent progress in the area of interdisciplinary research theory and technology furthermore we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject the goal of this conference is to bring together the researchers from academia and industry as well as practitioners to share ideas problems and solutions relating to the multifaceted aspects of interdisciplinary research theory and technology

Practical Hadoop Ecosystem 2016-09-30 learn how to use the apache hadoop projects including mapreduce hdfs apache hive apache hbase apache kafka apache mahout and apache solr from setting up the environment to running sample applications each chapter in this book is a practical tutorial on using an apache hadoop ecosystem project while several books on apache hadoop are available most are based on the main projects mapreduce and hdfs and none discusses the other apache hadoop ecosystem projects and how they all work together as a cohesive big data development platform what you will learn set up the environment in linux for hadoop projects using cloudera hadoop distribution cdh 5 run a mapreduce job store data with apache hive and apache hbase index data in hdfs with apache solr develop a kafka messaging system stream logs to hdfs with apache flume transfer data from mysql database to hive hdfs and hbase with sqoop create a hive table over apache solr develop a mahout user recommender system who this book is for apache hadoop developers pre requisite knowledge of linux and some knowledge of hadoop is required

Hadoop 2 Quick-Start Guide 2015-10-28 get started fast with apache hadoop 2 yarn and today s hadoop ecosystem with hadoop 2 x and yarn hadoop moves beyond mapreduce to become practical for virtually any type of data processing hadoop 2 x and the data lake concept represent a radical shift away from conventional approaches to data usage and storage hadoop 2 x installations offer unmatched scalability and breakthrough extensibility that supports new and existing big data analytics processing methods and models hadoop 2 quick start guide is the first easy accessible guide to apache hadoop 2 x yarn and the modern hadoop ecosystem building on his unsurpassed experience teaching hadoop and big data author douglas eadline covers all the basics you need to know to install and use hadoop 2 on personal computers or servers and to navigate the powerful technologies that complement it eadline concisely introduces and explains every key hadoop 2 concept tool and service illustrating each with a simple beginning to end example and identifying trustworthy up to date resources for learning more this guide is ideal if you want to learn about hadoop 2 without getting mired in technical details douglas eadline will bring you up to speed quickly whether you re a user admin devops specialist programmer architect analyst or data scientist coverage includes understanding what hadoop 2 and yarn do and how they improve on hadoop 1 with mapreduce understanding hadoop based data lakes versus rdbms data warehouses installing hadoop 2 and core services on linux machines virtualized sandboxes or clusters exploring the hadoop distributed file system hdfs understanding the essentials of mapreduce and yarn application programming simplifying programming and data movement with apache pig hive sqoop flume oozie and hbase observing application progress controlling jobs and managing workflows managing hadoop efficiently with apache ambari including recipes for hdfs to nfsv3 gateway hdfs snapshots and yarn configuration learning basic hadoop 2 troubleshooting and installing apache hue and apache spark

Implementing Big Data Analytics Using Hadoop 2019-06-12 the ultimate objective of this book is to help you become a professional in the field of big data and hadoop and ensuring you have enough skills to work in an industrial environment and solve real world problems to come up with solutions that make a difference to this world i tried at my best to explain the understanding on how a component in the hadoop ecosystem works why it works that way and how it fits into the design of the overall hadoop framework this book explains the hadoop framework followed by data analysis using mapreduce hive and pig on sample use

cases big data analysis using amazon elastic mapreduce hadoop on amazon cloud is also explained in detail it also focuses on the hadoop architecture as well as explains the hadoop setup using cloudera quickstart vm further mapreduce is also explained using a data analytics use case in addition of the above it also explains apache pig and apache hive respectively and show how these technologies can be used for solving data analysis problems as well as big data analytics using amazon services aws other valuable titles edge computing fog computing python simply in depth formal language and automata theory virtual reality iot programming internet of things 5g technologies

Architecting Modern Data Platforms 2018-12-05 there s a lot of information about big data technologies but splicing these technologies into an end to end enterprise data platform is a daunting task not widely covered with this practical book you ll learn how to build big data infrastructure both on premises and in the cloud and successfully architect a modern data platform ideal for enterprise architects it managers application architects and data engineers this book shows you how to overcome the many challenges that emerge during hadoop projects you ll explore the vast landscape of tools available in the hadoop and big data realm in a thorough technical primer before diving into infrastructure look at all component layers in a modern data platform from the server to the data center to establish a solid foundation for data in your enterprise platform understand aspects of deployment operation security high availability and disaster recovery along with everything you need to know to integrate your platform with the rest of your enterprise it taking hadoop to the cloud learn the important architectural aspects of running a big data platform in the cloud while maintaining enterprise security and high availability

Beginning Hadoop 2015-12-24 there are many challenges in setting up and scaling distributed frameworks like hadoop despite hadoop being an open source product and with so many good documentations and books it is difficult for an individual or an enterprise to define various use cases or working models that too with a clear understanding of its workings and tuning it for optimal performance pro hadoop administration by gurmukh singh a hadoop specialist and an infrastructure architect takes a deep dive into configuring hadoop services and its integration with various tools or frameworks the book covers the processes right from scratch to building a hadoop cluster at the production level with best practices and optimal performance you will learn use cases and set of recipes for the hadoop production environment from compiling hadoop to setting up cluster with highly available services it s integration with various tools like sqoop flume hbase hive and many more performance tuning and cluster planning hadoop security like kerberos encryption and other aspects of security like os and network level what you ll learn how to define a problem at hand and describe a solution for it what is hadoop cluster configuration for both hadoop 1 x and 2 x how to setup high availability for hadoop services how to integrate with pig hive sqoop hbase flume with examples how to setup kerberos and security for hadoop daemons how to optimize hadoop for best best performance based on hadware software parameters who this book is for this book is targeted towards beginners and intermediate users who want to learn or enhance their knowledge about hadoop framework this book will even be useful for advanced users as it covers aspects of cluster from a production perspective

Hadoop in Practice 2014-10-12 summary hadoop in practice second edition provides over 100 tested instantly useful techniques that will help you conquer big data using hadoop this revised new edition covers changes and new features in the hadoop core architecture including mapreduce 2 brand new chapters cover yarn and integrating kafka impala and spark sql with hadoop you ll also get new and updated techniques for flume sqoop and mahout all of which have seen major new versions recently in short this is the most practical up to date coverage of hadoop available anywhere purchase of the print book includes a free ebook in pdf kindle and epub formats from manning publications about the book it s always a good time to upgrade your hadoop skills hadoop in practice second edition provides a collection of 104 tested instantly useful techniques for analyzing real time streams moving data securely machine learning managing large scale clusters and taming big data using hadoop this completely revised

edition covers changes and new features in hadoop core including mapreduce 2 and yarn you ll pick up hands on best practices for integrating spark kafka and impala with hadoop and get new and updated techniques for the latest versions of flume sqoop and mahout in short this is the most practical up to date coverage of hadoop available readers need to know a programming language like java and have basic familiarity with hadoop what s inside thoroughly updated for hadoop 2 how to write yarn applications integrate real time technologies like storm impala and spark predictive analytics using mahout and rr readers need to know a programming language like java and have basic familiarity with hadoop about the author alex holmes works on tough big data problems he is a software engineer author speaker and blogger specializing in large scale hadoop projects table of contents part 1 background and fundamentals hadoop in a heartbeat introduction to yarn part 2 data logistics data serialization working with text and beyond organizing and optimizing data in hdfs moving data into and out of hadoop part 3 big data patterns applying mapreduce patterns to big data utilizing data structures and algorithms at scale tuning debugging and testing part 4 beyond mapreduce sql on hadoop writing a yarn application

Practical Data Science with Hadoop and Spark 2016-12-08 the complete guide to data science with hadoop for technical professionals businesspeople and students demand is soaring for professionals who can solve real data science problems with hadoop and spark practical data science with hadoop and spark is your complete guide to doing just that drawing on immense experience with hadoop and big data three leading experts bring together everything you need high level concepts deep dive techniques real world use cases practical applications and hands on tutorials the authors introduce the essentials of data science and the modern hadoop ecosystem explaining how hadoop and spark have evolved into an effective platform for solving data science problems at scale in addition to comprehensive application coverage the authors also provide useful guidance on the important steps of data ingestion data munging and visualization once the groundwork is in place the authors focus on specific applications including machine learning predictive modeling for sentiment analysis clustering for document analysis anomaly detection and natural language processing nlp this guide provides a strong technical foundation for those who want to do practical data science and also presents business driven guidance on how to apply hadoop and spark to optimize roi of data science initiatives learn what data science is how it has evolved and how to plan a data science career how data volume variety and velocity shape data science use cases hadoop and its ecosystem including hdfs mapreduce yarn and spark data importation with hive and spark data quality preprocessing preparation and modeling visualization surfacing insights from huge data sets machine learning classification regression clustering and anomaly detection algorithms and hadoop tools for predictive modeling cluster analysis and similarity functions large scale anomaly detection nlp applying data science to human language

Pro Docker 2015-12-30 in this fast paced book on the docker open standards platform for developing packaging and running portable distributed applications deepak vorhadiscusses how to build ship and run applications on any platform such as a pc the cloud data center or a virtual machine he describes how to install and create docker images and the advantages off docker containers the remainder of the book is devoted to discussing using docker with important software solutions he begins by discussing using docker with a traditional rdbms using oracle and mysql next he moves on to nosql with chapter on mongodb cassandra and couchbase then he addresses the use of docker in the hadoop ecosystem with complete chapters on utilizing not only hadoop but hive hbase sqoop kafka solr and spark what you will learn how to install a docker image how to create a docker container how to run an application in a docker container use docker with apache hadoop ecosystem use docker with nosql databases use docker with rdbms who this book is forapache hadoop developers database developers nosql developers

Hadoop Practice Guide 2019-08-19 this book is a complete practical approach for hadoop lovers it is mainly aimed at beginners who want to have a hands on experience with hadoop and its ecosystem its simplicity and step by step explanation will help students and other readers in the computer science

industry to use this book as a reference manual the book has been divided into various chapters that cover hadoop installation summary on hadoop core components general commands in hadoop with examples sqoop import export commands with verification steps pig latin commands analysis using pig latin pig script examples hiveql queries and expected outputs and hbase with crud operations in short this book is a guide for programmers and non programmers to begin their projects in hadoop it is also suitable as a reference manual for students and professionals who are new to the hadoop ecosystems

Big Data Analytics Beyond Hadoop 2014 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

Information Technology - New Generations 2017-07-15 this volume presents a collection of peer reviewed scientific articles from the 14th international conference on information technology new generations held at the university of nevada at las vegas on april 10 12 at tuscany suites hotel in las vegas the book of chapters addresses critical areas of information technology including web technology communications computing architectures software engineering security and data mining

Field Guide to Hadoop 2015-03-02 annotation it managers developers data analysts system architects and similar technical workers are now encountering the largest and most disruptive change in their profession since the ascendancy of the relational database in early 1980s you hear that nosql and big data analytics are about to replace the systems and skills you now own and possess but there s often no easy way to make that transition to exacerbate the issue the transition may not be gradual but forced on you by a new project in your enterprisenamely hadoopthat will immediately require new ways of thinking new tools and new techniques this book helps you understand the components of the hadoop ecosystem and how they relate to each other you ll discover how to get started on that project in an efficient manner that lays out the possibilities the authors suggest a path and resources that will guide you on their journey from the status quo to the brave new world you face

Securing Hadoop 2013-11-22 this book is a step by step tutorial filled with practical examples which will focus mainly on the key security tools and implementation techniques of hadoop security this book is great for hadoop practitioners solution architects hadoop administrators developers and hadoop project managers who are looking to get a good grounding in what kerberos is all about and who wish to learn how to implement end to end hadoop security within an enterprise setup it s assumed that you will have some basic understanding of hadoop as well as be familiar with some basic security concepts

Hadoop For Dummies 2014-04-14 let hadoop for dummies help harness the power of your data and rein in the information overload big data has become big business and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed enter hadoop and this easy to understand for dummies guide hadoop for dummies helps readers understand the value of big data make a business case for using hadoop navigate the hadoop ecosystem and build and manage hadoop applications and clusters explains the origins of hadoop its economic benefits and its functionality and practical applications helps you find your way around the hadoop ecosystem program mapreduce utilize design patterns and get your hadoop cluster up and running quickly and easily details how to use hadoop applications for data mining web analytics and personalization large scale text processing data science and problem solving shows you how to improve the value of your hadoop cluster maximize your investment in hadoop and avoid common pitfalls when building your hadoop cluster from programmers challenged with building and maintaining affordable scaleable data systems to administrators who must deal with huge volumes of information effectively and efficiently this how to has something to help you with hadoop

Pro Tableau 2016-12-23 leverage the power of visualization in business intelligence and data science to make quicker and better decisions use statistics and data mining to make compelling and interactive dashboards this book will help those familiar with tableau software chart their journey to being a visualization expert pro tableau demonstrates the power of visual analytics and teaches you how to connect to various data sources such as spreadsheets text files relational databases microsoft sql server mysql etc non relational databases nosql such as mongodb cassandra r data files etc write your own custom sql etc perform statistical analysis in tableau using r use a multitude of charts pie bar stacked bar line scatter plots dual axis histograms heat maps tree maps highlight tables box and whisker etc what you ll learn connect to various data sources such as relational databases microsoft sql server mysql non relational databases nosql such as mongodb cassandra write your own custom sql join and blend data sources etc leverage table calculations moving average year over year growth lod level of detail etc integrate tableau with r tell a compelling story with data by creating highly interactive dashboards who this book is for all levels of it professionals from executives responsible for determining it strategies to systems administrators to data analysts to decision makers responsible for driving strategic initiatives etc the book will help those familiar with tableau software chart their journey to a visualization expert

Apache Oozie 2015-05-12 get a solid grounding in apache oozie the workflow scheduler system for managing hadoop jobs with this hands on guide two experienced hadoop practitioners walk you through the intricacies of this powerful and flexible platform with numerous examples and real world use cases once you set up your oozie server you ll dive into techniques for writing and coordinating workflows and learn how to write complex data pipelines advanced topics show you how to handle shared libraries in oozie as well as how to implement and manage oozie s security capabilities install and configure an oozie server and get an overview of basic concepts journey through the world of writing and configuring workflows learn how the oozie coordinator schedules and executes workflows based on triggers understand how oozie manages data dependencies use oozie bundles to package several coordinator apps into a data pipeline learn about security features and shared library management implement custom extensions and write your own el functions and actions debug workflows and manage oozie s operational details

BIG DATA AND HADOOP 2018-06-02 the book contains the latest trend in it industry bigdata and hadoop it explains how big is big data and why everybody is trying to implement this into their it project it includes research work on various topics theoretical and practical approach each component of the architecture is described along with current industry trends big data and hadoop have taken together are a new skill as per the industry standards readers will get a compact book along with the industry experience and would be a reference to help readers key features overview of big data basics of hadoop

hadoop distributed file system hbase mapreduce hive the dataware house of hadoop pig the higher level programming environment sqoop importing data from heterogeneous sources flume ozzie zookeeper big data stream mining chapter wise questions previous years questions

Beginning Apache Hadoop Administration 2017-09-07 bigdata is one of the most demanding markets in the it sector if you are an administrator or a have a passion for knowing the internal configurations of hadoop then this book is for you this book enables a professional to learn about hadoop in terms of installation configuration and management this book will help the reader to jumpstart with hadoop frameworks its eco system components and slowly progress towards learning the administration part of hadoop the level of this book goes from beginner to intermediate with 70 hands on exercises some of the techniques that you will learn include installation and configuration of hadoop cluster performing hadoop cluster upgrade understanding and implementing hdfs federation understanding and implementing high availability implementing ha on a federated cluster zookeeper cli apache hive installation and security hbase multi master setup oozie installation configuration and job submission setting up hdfs quotas setting up hdfs nfs gateway understanding and implementing rolling upgrade and much more

- [practical guide to troubleshooting installation and maintenance of variable frequency drives practical guide series Copy](#)
- [crooked river a novel valerie geary (2023)](#)
- [ncert solutions for class 11 chemistry chapter 1 Copy](#)
- [rabbit proof fence oxford bookworms library [PDF]](#)
- [icse selina concise biology guide [PDF]](#)
- [sony tv guide not working Copy](#)
- [darkhenge (Read Only)](#)
- [the mba application roadmap the essential guide to getting into a top business school Full PDF](#)
- [confessions of an economic hitman [PDF]](#)
- [suzuki forenza (2023)](#)
- [rock candy kisses 3am 5 addison moore kulturoffensive Full PDF](#)
- [engineering mathematics 3 by np bali solutions (Download Only)](#)
- [ap american government 11th edition Full PDF](#)
- [examples of statistics papers .pdf](#)
- [e2020 algebra 2 prescriptive test answers (Download Only)](#)
- [take back your life using microsoft office outlook 2007 to get organized and stay organized business skills (Read Only)](#)
- [measuring l2 proficiency perspectives from sla second language acquisition [PDF]](#)
- [service manual of konica minolta bizhub 164 [PDF]](#)
- [scaricare libri gratis google books (Read Only)](#)
- [sensation amp perception third edition sinauer associates (2023)](#)
- [answer key cost accounting 14th edition (PDF)](#)
- [ruzi nazar ciann t rk casusu [PDF]](#)
- [bat boy Full PDF](#)
- [psa test by cbse sample papers 11th Full PDF](#)
- [face2face intermediate workbook answer key second edition (Read Only)](#)
- [four seasons the story of a business philosophy (Read Only)](#)
- [ugly s electrical references (Read Only)](#)