

# READ BIG DATA FOR CHIMPS A GUIDE TO MASSIVE SCALE DATA PROCESSING IN PRACTICE

## Big Data for Chimps

Annotation To help you answer big data questions, this unique guide shows you how to use simple, fun, and elegant tools leveraging Apache Hadoop. You'll learn how to break problems into efficient data transformations to meet most of your analysis needs.

## Big Data for Chimps

Large Scale and Big Data: Processing and Management provides readers with a central source of reference on the data management techniques currently available for large-scale data processing. Presenting chapters written by leading researchers, academics, and practitioners, it addresses the fundamental challenges associated with Big Data processing t

## Large Scale and Big Data

DESIGNING BIG DATA PLATFORMS Provides expert guidance and valuable insights on getting the most out of Big Data systems An array of tools are currently available for managing and processing data—some are ready-to-go solutions that can be immediately deployed, while others require complex and time-intensive setups. With such a vast range of options, choosing the right tool to build a solution can be complicated, as can determining which tools work well with each other. Designing Big Data Platforms provides clear and authoritative guidance on the critical decisions necessary for successfully deploying, operating, and maintaining Big Data systems. This highly practical guide helps readers understand how to process large amounts of data with well-known Linux tools and database solutions, use effective techniques to collect and manage data from multiple sources, transform data into meaningful business insights, and much more.

Author Yusuf Aytas, a software engineer with a vast amount of big data experience, discusses the design of the ideal Big Data platform: one that meets the needs of data analysts, data engineers, data scientists, software engineers, and a spectrum of other stakeholders across an organization. Detailed yet accessible chapters cover key topics such as stream data processing, data analytics, data science, data discovery, and data security. This real-world manual for Big Data technologies: Provides up-to-date coverage of the tools currently used in Big Data processing and management Offers step-by-step guidance on building a data pipeline, from basic scripting to distributed systems Highlights and explains how data is processed at scale Includes an introduction to the foundation of a modern data platform Designing Big Data Platforms: How to Use, Deploy, and Maintain Big Data Systems is a must-have for all professionals working with Big Data, as well researchers and students in computer science and related fields.

## Designing Big Data Platforms

Big Data Imperatives, focuses on resolving the key questions on everyone's mind: Which data matters? Do you have enough data volume to justify the usage? How you want to process this amount of data? How long do you really need to keep it active for your analysis, marketing, and BI applications? Big data is emerging from the realm of one-off projects to mainstream business adoption; however, the real value of big data is not

in the overwhelming size of it, but more in its effective use. This book addresses the following big data characteristics: Very large, distributed aggregations of loosely structured data – often incomplete and inaccessible Petabytes/Exabytes of data Millions/billions of people providing/contributing to the context behind the data Flat schema's with few complex interrelationships Involves time-stamped events Made up of incomplete data Includes connections between data elements that must be probabilistically inferred Big Data Imperatives explains 'what big data can do'. It can batch process millions and billions of records both unstructured and structured much faster and cheaper. Big data analytics provide a platform to merge all analysis which enables data analysis to be more accurate, well-rounded, reliable and focused on a specific business capability. Big Data Imperatives describes the complementary nature of traditional data warehouses and big-data analytics platforms and how they feed each other. This book aims to bring the big data and analytics realms together with a greater focus on architectures that leverage the scale and power of big data and the ability to integrate and apply analytics principles to data which earlier was not accessible. This book can also be used as a handbook for practitioners; helping them on methodology, technical architecture, analytics techniques and best practices. At the same time, this book intends to hold the interest of those new to big data and analytics by giving them a deep insight into the realm of big data.

## **Big Data Imperatives**

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

## **Big Data: Concepts, Methodologies, Tools, and Applications**

Find the right big data solution for your business or organization Big data management is one of the major challenges facing business, industry, and not-for-profit organizations. Data sets such as customer transactions for a mega-retailer, weather patterns monitored by meteorologists, or social network activity can quickly outpace the capacity of traditional data management tools. If you need to develop or manage big data solutions, you'll appreciate how these four experts define, explain, and guide you through this new and often confusing concept. You'll learn what it is, why it matters, and how to choose and implement solutions that work. Effectively managing big data is an issue of growing importance to businesses, not-for-profit organizations, government, and IT professionals Authors are experts in information management, big data, and a variety of solutions Explains big data in detail and discusses how to select and implement a solution, security concerns to consider, data storage and presentation issues, analytics, and much more Provides essential information in a no-nonsense, easy-to-understand style that is empowering Big Data For Dummies cuts through the confusion and helps you take charge of big data solutions for your organization.

## **Big Data For Dummies**

Making Big Data Work: Real-World Use Cases and Examples, Practical Code, Detailed Solutions Large-scale data analysis is now vitally important to virtually every business. Mobile and social technologies are generating massive datasets; distributed cloud computing offers the resources to store and analyze them; and professionals have radically new technologies at their command, including NoSQL databases. Until now, however, most books on "Big Data" have been little more than business polemics or product catalogs. Data

Just Right is different: It's a completely practical and indispensable guide for every Big Data decision-maker, implementer, and strategist. Michael Manoochchri, a former Google engineer and data hacker, writes for professionals who need practical solutions that can be implemented with limited resources and time. Drawing on his extensive experience, he helps you focus on building applications, rather than infrastructure, because that's where you can derive the most value. Manoochchri shows how to address each of today's key Big Data use cases in a cost-effective way by combining technologies in hybrid solutions. You'll find expert approaches to managing massive datasets, visualizing data, building data pipelines and dashboards, choosing tools for statistical analysis, and more. Throughout, the author demonstrates techniques using many of today's leading data analysis tools, including Hadoop, Hive, Shark, R, Apache Pig, Mahout, and Google BigQuery. Coverage includes Mastering the four guiding principles of Big Data success—and avoiding common pitfalls Emphasizing collaboration and avoiding problems with siloed data Hosting and sharing multi-terabyte datasets efficiently and economically “Building for infinity” to support rapid growth Developing a NoSQL Web app with Redis to collect crowd-sourced data Running distributed queries over massive datasets with Hadoop, Hive, and Shark Building a data dashboard with Google BigQuery Exploring large datasets with advanced visualization Implementing efficient pipelines for transforming immense amounts of data Automating complex processing with Apache Pig and the Cascading Java library Applying machine learning to classify, recommend, and predict incoming information Using R to perform statistical analysis on massive datasets Building highly efficient analytics workflows with Python and Pandas Establishing sensible purchasing strategies: when to build, buy, or outsource Previewing emerging trends and convergences in scalable data technologies and the evolving role of the Data Scientist

## **Data Just Right**

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

## **Frontiers in Massive Data Analysis**

With new technologies, such as computer vision, internet of things, mobile computing, e-governance and e-commerce, and wide applications of social media, organizations generate a huge volume of data and at a much faster rate than several years ago. Big data in large-/small-scale systems, characterized by high volume, diversity, and velocity, increasingly drives decision making and is changing the landscape of business intelligence. From governments to private organizations, from communities to individuals, all areas are being affected by this shift. There is a high demand for big data analytics that offer insights for computing efficiency, knowledge discovery, problem solving, and event prediction. To handle this demand and this increase in big data, there needs to be research on innovative and optimized machine learning algorithms in both large- and small-scale systems. *Applications of Big Data in Large- and Small-Scale Systems* includes state-of-the-art research findings on the latest development, up-to-date issues, and challenges in the field of

big data and presents the latest innovative and intelligent applications related to big data. This book encompasses big data in various multidisciplinary fields from the medical field to agriculture, business research, and smart cities. While highlighting topics including machine learning, cloud computing, data visualization, and more, this book is a valuable reference tool for computer scientists, data scientists and analysts, engineers, practitioners, stakeholders, researchers, academicians, and students interested in the versatile and innovative use of big data in both large-scale and small-scale systems.

## **Applications of Big Data in Large- and Small-Scale Systems**

Principles of Big Data helps readers avoid the common mistakes that endanger all Big Data projects. By stressing simple, fundamental concepts, this book teaches readers how to organize large volumes of complex data, and how to achieve data permanence when the content of the data is constantly changing. General methods for data verification and validation, as specifically applied to Big Data resources, are stressed throughout the book. The book demonstrates how adept analysts can find relationships among data objects held in disparate Big Data resources, when the data objects are endowed with semantic support (i.e., organized in classes of uniquely identified data objects). Readers will learn how their data can be integrated with data from other resources, and how the data extracted from Big Data resources can be used for purposes beyond those imagined by the data creators. Learn general methods for specifying Big Data in a way that is understandable to humans and to computers Avoid the pitfalls in Big Data design and analysis Understand how to create and use Big Data safely and responsibly with a set of laws, regulations and ethical standards that apply to the acquisition, distribution and integration of Big Data resources

## **Principles of Big Data**

Many corporations are finding that the size of their data sets are outgrowing the capability of their systems to store and process them. The data is becoming too big to manage and use with traditional tools. The solution: implementing a big data system. As *Big Data Made Easy: A Working Guide to the Complete Hadoop Toolset* shows, Apache Hadoop offers a scalable, fault-tolerant system for storing and processing data in parallel. It has a very rich toolset that allows for storage (Hadoop), configuration (YARN and ZooKeeper), collection (Nutch and Solr), processing (Storm, Pig, and Map Reduce), scheduling (Oozie), moving (Sqoop and Avro), monitoring (Chukwa, Ambari, and Hue), testing (Big Top), and analysis (Hive). The problem is that the Internet offers IT pros wading into big data many versions of the truth and some outright falsehoods born of ignorance. What is needed is a book just like this one: a wide-ranging but easily understood set of instructions to explain where to get Hadoop tools, what they can do, how to install them, how to configure them, how to integrate them, and how to use them successfully. And you need an expert who has worked in this area for a decade—someone just like author and big data expert Mike Frampton. *Big Data Made Easy* approaches the problem of managing massive data sets from a systems perspective, and it explains the roles for each project (like architect and tester, for example) and shows how the Hadoop toolset can be used at each system stage. It explains, in an easily understood manner and through numerous examples, how to use each tool. The book also explains the sliding scale of tools available depending upon data size and when and how to use them. *Big Data Made Easy* shows developers and architects, as well as testers and project managers, how to: Store big data Configure big data Process big data Schedule processes Move data among SQL and NoSQL systems Monitor data Perform big data analytics Report on big data processes and projects Test big data systems *Big Data Made Easy* also explains the best part, which is that this toolset is free. Anyone can download it and—with the help of this book—start to use it within a day. With the skills this book will teach you under your belt, you will add value to your company or client immediately, not to mention your career.

## **Big Data Made Easy**

PEEK \("UNDER THE HOOD\) OF BIG DATA ANALYTICS The world of big data analytics grows ever more complex. And while many people can work superficially with specific frameworks, far fewer

understand the fundamental principles of large-scale, distributed data processing systems and how they operate. In *Foundations of Data Intensive Applications: Large Scale Data Analytics under the Hood*, renowned big-data experts and computer scientists Drs. Supun Kamburugamuve and Saliya Ekanayake deliver a practical guide to applying the principles of big data to software development for optimal performance. The authors discuss foundational components of large-scale data systems and walk readers through the major software design decisions that define performance, application type, and usability. You will learn how to recognize problems in your applications resulting in performance and distributed operation issues, diagnose them, and effectively eliminate them by relying on the bedrock big data principles explained within. Moving beyond individual frameworks and APIs for data processing, this book unlocks the theoretical ideas that operate under the hood of every big data processing system. Ideal for data scientists, data architects, dev-ops engineers, and developers, *Foundations of Data Intensive Applications: Large Scale Data Analytics under the Hood* shows readers how to:

- Identify the foundations of large-scale, distributed data processing systems
- Make major software design decisions that optimize performance
- Diagnose performance problems and distributed operation issues
- Understand state-of-the-art research in big data
- Explain and use the major big data frameworks and understand what underpins them
- Use big data analytics in the real world to solve practical problems

## **Foundations of Data Intensive Applications**

Data has become a valuable asset like never before. Today the challenge is not a shortage of data but the need for techniques and methods capable enough to be able to glean valuable insights from the fast-flowing mass of Big Data. This new volume, *Handbook of Research for Big Data: Concepts and Techniques*, helps to meet the challenge of managing and using Big Data by presenting new research on various technological advances in the field. The chapters in the book present information on important applications, concepts, and technologies for Big Data in the present industry and market scenario. It looks at research domain issues and their solutions as well as various research case studies, research plans, methodologies, and related data sets for the four Vs: volume, velocity, variety, and veracity. Chapters discuss Big Data in governance, transportation, disaster management, epidemiology, and more. The book covers design and analysis of reconfigurable computing of SoC for IoT, data mining techniques and applications, the use of natural language processing in big data, and more. The volume is a valuable resource for researchers from both academia and industry to learn about and enhance their knowledge and skills in the broad area of big data computing and applications.

## **Handbook of Research for Big Data**

To help you navigate the large number of new data tools available, this guide describes 60 of the most recent innovations, from NoSQL databases and MapReduce approaches to machine learning and visualization tools. Descriptions are based on first-hand experience with these tools in a production environment. This handy glossary also includes a chapter of key terms that help define many of these tool categories:

- NoSQL Databases—Document-oriented databases using a key/value interface rather than SQL
- MapReduce—Tools that support distributed computing on large datasets
- Storage—Technologies for storing data in a distributed way
- Servers—Ways to rent computing power on remote machines
- Processing—Tools for extracting valuable information from large datasets
- Natural Language Processing—Methods for extracting information from human-created text
- Machine Learning—Tools that automatically perform data analyses, based on results of a one-off analysis
- Visualization—Applications that present meaningful data graphically
- Acquisition—Techniques for cleaning up messy public data sources
- Serialization—Methods to convert data structure or object state into a storable format

## **Big Data Glossary**

*Big Data: Principles and Paradigms* captures the state-of-the-art research on the architectural aspects, technologies, and applications of Big Data. The book identifies potential future directions and technologies

that facilitate insight into numerous scientific, business, and consumer applications. To help realize Big Data's full potential, the book addresses numerous challenges, offering the conceptual and technological solutions for tackling them. These challenges include life-cycle data management, large-scale storage, flexible processing infrastructure, data modeling, scalable machine learning, data analysis algorithms, sampling techniques, and privacy and ethical issues.

## **Big Data**

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.

## **Modern Big Data Processing with Hadoop**

Welcome to this book on "\"Big Data Analytics - Derive Insights\" is a beginners guide to big data, mining, manipulation, integration, processing, visualisation and interpretation based on actual case studies in implementing Data Management frameworks across diverse organisations. Big data is a term used collectively for data sets that are voluminous, large and complicated that cannot be processed by traditional databases and applications. Some of the major challenges are the collection, extraction, mining, integration, analysis, sharing, storage, transfer, loading, querying, updating and visualisation of these data sets from multiple streams and formats. An effective and accurate visualisation of big data will lead to a more informed decision which in turn will result in greater operational efficiency, cost reduction, increased revenues and reduction in risk. Big Data analysis is used to identify business trends, predict accurate supply and demand, prevent diseases, combat crime and growth in revenues. Businessmen, organisations, scientists, hospitals, social media, marketing and governments face difficulties with large data sets in areas including internet search, finance and business trends. This book will give you a basic understanding of how big data works and the process used to get insights from such data.

## **Big Data Analytics**

Handbook of Big Data provides a state-of-the-art overview of the analysis of large-scale datasets. Featuring contributions from well-known experts in statistics and computer science, this handbook presents a carefully curated collection of techniques from both industry and academia. Thus, the text instills a working understanding of key statistical

## **Handbook of Big Data**

Due to the increasing need to solve complex problems, high-performance computing (HPC) is now one of the most fundamental infrastructures for scientific development in all disciplines, and it has progressed massively in recent years as a result. HPC facilitates the processing of big data, but the tremendous research challenges faced in recent years include: the scalability of computing performance for high velocity, high variety and high volume big data; deep learning with massive-scale datasets; big data programming paradigms on multi-core; GPU and hybrid distributed environments; and unstructured data processing with high-performance computing. This book presents 19 selected papers from the TopHPC2017 congress on Advances in High-Performance Computing and Big Data Analytics in the Exascale era, held in Tehran, Iran, in April 2017. The book is divided into 3 sections: State of the Art and Future Scenarios, Big Data Challenges, and HPC Challenges, and will be of interest to all those whose work involves the processing of Big Data and the use of HPC.

## **Big Data and HPC**

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

## **Hadoop: The Definitive Guide**

"Big Data Analytics is a field that dissects, efficiently extricates data from, or in any case manages informational indexes that are excessively huge or complex to be managed by customary information preparing application programming. Information with numerous cases (lines) offers more noteworthy factual force, while information with higher multifaceted nature may prompt a higher bogus disclosure rate. Enormous information challenges incorporate catching information, information stockpiling, information investigation, search, sharing, move, representation, and questioning, refreshing, data security and data source. Large information was initially connected with three key ideas: volume, variety and velocity. Consequently, huge information regularly incorporates information with sizes that surpass the limit of conventional programming to measure inside a satisfactory time and worth. Current utilization of the term enormous information will in general allude to the utilization of predictive analytics, user behavior analytics, or certain other progressed information investigation techniques that concentrate an incentive from information, and sometimes to a specific size of informational index. There is little uncertainty that the amounts of information now accessible are undoubtedly enormous, however that is not the most important quality of this new information biological system. Investigation of informational indexes can discover new relationships to spot business patterns or models. Researchers, business persons, clinical specialists,

promoting and governments consistently meet challenges with huge informational collections in territories including Internet look, fintech, metropolitan informatics, and business informatics. Researchers experience constraints in e-Science work, including meteorology, genomics, connectomics, complex material science reproductions, science and ecological exploration. The main objective of this book is to write about issues, challenges, opportunities, and solutions in novel research projects about big data in various domains. The topics of interest include, but are not limited to: efficient storage, management and sharing large scale of data; novel approaches for analyzing data using big data technologies; implementation of high performance and/or scalable and/or real-time computation algorithms for analyzing big data; usage of various data sources like historical data, social networking media, machine data and crowd-sourcing data; using machine learning, visual analytics, data mining, spatio-temporal data analysis and statistical inference in different domains (with large scale datasets); Legal and ethical issues and solutions for using, sharing and publishing large datasets; and the results of data analytics, security and privacy issues"--

## **A Closer Look at Big Data Analytics**

Annotation Foundations for Architecting Data Solutions provides everyone from CIOs and COOs to lead architects and lead developers with the fundamental concepts of big data development. Authors Ted Malaska and Jonathan Seidman guide you through all the major components necessary to start, architect, and develop successful big data projects. This practical book covers a variety of different big data architectures and applications, from massive data pipelines to web scale applications. Each chapter addresses a different part of the software development life cycle and identifies patterns that build on one another to maximize success throughout the life of your project. You'll learn how to: Build a Big Data center of excellence in your company for the first time Identify and manage risk in your data project Retain and motivate teams to increase engagement and innovation Maximize Big Data ROI and align cost structure to help your company attain success.

## **Foundations for Architecting Data Solutions**

The book describes the emergence of big data technologies and the role of Spark in the entire big data stack. It compares Spark and Hadoop and identifies the shortcomings of Hadoop that have been overcome by Spark. The book mainly focuses on the in-depth architecture of Spark and our understanding of Spark RDDs and how RDD complements big data's immutable nature, and solves it with lazy evaluation, cacheable and type inference. It also addresses advanced topics in Spark, starting with the basics of Scala and the core Spark framework, and exploring Spark data frames, machine learning using Mllib, graph analytics using Graph X and real-time processing with Apache Kafka, AWS Kinesis, and Azure Event Hub. It then goes on to investigate Spark using PySpark and R. Focusing on the current big data stack, the book examines the interaction with current big data tools, with Spark being the core processing layer for all types of data. The book is intended for data engineers and scientists working on massive datasets and big data technologies in the cloud. In addition to industry professionals, it is helpful for aspiring data processing professionals and students working in big data processing and cloud computing environments.

## **Big Data Processing Using Spark in Cloud**

Your Python code may run correctly, but you need it to run faster. Updated for Python 3, this expanded edition shows you how to locate performance bottlenecks and significantly speed up your code in high-data-volume programs. By exploring the fundamental theory behind design choices, High Performance Python helps you gain a deeper understanding of Python's implementation. How do you take advantage of multicore architectures or clusters? Or build a system that scales up and down without losing reliability? Experienced Python programmers will learn concrete solutions to many issues, along with war stories from companies that use high-performance Python for social media analytics, productionized machine learning, and more. Get a better grasp of NumPy, Cython, and profilers Learn how Python abstracts the underlying computer architecture Use profiling to find bottlenecks in CPU time and memory usage Write efficient programs by



choosing appropriate data structures Speed up matrix and vector computations Use tools to compile Python down to machine code Manage multiple I/O and computational operations concurrently Convert multiprocessing code to run on local or remote clusters Deploy code faster using tools like Docker

## **HADOOP**

Executive summary: Tourism is often proposed 1) as a strategy to fund conservation efforts to protect great apes and their habitats, 2) as a way for local communities to participate in, and benefit from, conservation activities on behalf of great apes, or 3) as a business. A few very successful sites point to the considerable potential of conservation-based great ape tourism, but it will not be possible to replicate this success everywhere. The number of significant risks to great apes that can arise from tourism require a cautious approach. If great ape tourism is not based on sound conservation principles right from the start, the odds are that economic objectives will take precedence, the consequences of which in all likelihood would be damaging to the well-being and eventual survival of the apes, and detrimental to the continued preservation of their habitat. All great ape species and subspecies are classified as Endangered or Critically Endangered on the IUCN Red List of Threatened Species (IUCN 2010), therefore it is imperative that great ape tourism adhere to the best practice guidelines in this document. The guiding principles of best practice in great ape tourism are: Tourism is not a panacea for great ape conservation or revenue generation; Tourism can enhance long-term support for the conservation of great apes and their habitat; Conservation comes first—it must be the primary goal at any great ape site and tourism can be a tool to help fund it; Great ape tourism should only be developed if the anticipated conservation benefits, as identified in impact studies, significantly outweigh the risks; Enhanced conservation investment and action at great ape tourism sites must be sustained in perpetuity; Great ape tourism management must be based on sound and objective science; Benefits and profit for communities adjacent to great ape habitat should be maximised; Profit to private sector partners and others who earn income associated with tourism is also important, but should not be the driving force for great ape tourism development or expansion; Comprehensive understanding of potential impacts must guide tourism development. positive impacts from tourism must be maximised and negative impacts must be avoided or, if inevitable, better understood and mitigated. The ultimate success or failure of great ape tourism can lie in variables that may not be obvious to policymakers who base their decisions primarily on earning revenue for struggling conservation programmes. However, a number of biological, geographical, economic and global factors can affect a site so as to render ape tourism ill-advised or unsustainable. This can be due, for example, to the failure of the tourism market for a particular site to provide revenue sufficient to cover the development and operating costs, or it can result from failure to protect the target great apes from the large number of significant negative aspects inherent in tourism. Either of these failures will have serious consequences for the great ape population. Once apes are habituated to human observers, they are at increased risk from poaching and other forms of conflict with humans. They must be protected in perpetuity even if tourism fails or ceases for any reason. Great ape tourism should not be developed without conducting critical feasibility analyses to ensure there is sufficient potential for success. Strict attention must be paid to the design of the enterprise, its implementation and continual management capacity in a manner that avoids, or at least minimises, the negative impacts of tourism on local communities and on the apes themselves. Monitoring programmes to track costs and impacts, as well as benefits, [is] essential to inform management on how to optimise tourism for conservation benefits. These guidelines have been developed for both existing and potential great ape tourism sites that wish to improve the degree to which their programme contributes to the conservation rather than the exploitation of great apes.

## **High Performance Python**

National Book Award Finalist: “This man’s ideas may be the most influential, not to say controversial, of the second half of the twentieth century.”—Columbus Dispatch At the heart of this classic, seminal book is Julian Jaynes's still-controversial thesis that human consciousness did not begin far back in animal evolution but instead is a learned process that came about only three thousand years ago and is still developing. The implications of this revolutionary scientific paradigm extend into virtually every aspect of our psychology,

our history and culture, our religion—and indeed our future. “Don’t be put off by the academic title of Julian Jaynes’s *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Its prose is always lucid and often lyrical...he unfolds his case with the utmost intellectual rigor.”—The New York Times “When Julian Jaynes . . . speculates that until late in the twentieth millennium BC men had no consciousness but were automatically obeying the voices of the gods, we are astounded but compelled to follow this remarkable thesis.”—John Updike, *The New Yorker* “He is as startling as Freud was in *The Interpretation of Dreams*, and Jaynes is equally as adept at forcing a new view of known human behavior.”—*American Journal of Psychiatry*

## **Best Practice Guidelines for Great Ape Tourism**

From the website: Although the IUCN has previously established working protocols for plant and animal re-introduction, the great apes present unique challenges and concerns owing to their singular cognitive development. This prompted the Primate Specialist Group to reconsider the existing guidelines in terms of the specific needs of great apes. The resulting synthesis, representing the expert opinion of primatologists and re-introduction practitioners, is presented here as part of the series of best-practices documents. Specifically designed for rehabilitators and specialists in re-introduction, these guidelines start from the fundamental assumption that re-introductions should not endanger wild populations of great apes or the ecosystems they inhabit. Equally important is the health and welfare of the individual great apes being re-introduced, as well as the caretaker staff and the residents of the surrounding areas. The re-introduction guidelines also require that the factors which first threatened great apes in the proposed site of release have been addressed and resolved.

## **The Origin of Consciousness in the Breakdown of the Bicameral Mind**

This title brings together work on embodiment, action, and the predictive mind. At the core is the vision of human minds as prediction machines - devices that constantly try to stay one step ahead of the breaking waves of sensory stimulation, by actively predicting the incoming flow. In every situation we encounter, that complex prediction machinery is already buzzing, proactively trying to anticipate the sensory barrage. The book shows in detail how this strange but potent strategy of self-anticipation ushers perception, understanding, and imagination simultaneously onto the cognitive stage.

## **Best Practice Guidelines for the Re-Introduction of Great Apes**

Wild chimpanzees are only found in tropical Africa, where their populations have declined by more than 66% in the last 30 years. This Action Plan focuses on one of the four chimpanzee subspecies, the western chimpanzee, which is one of the two subspecies most threatened with extinction. This publication presents a plan for action that represents a consensus among all parties concerned with the conservation of chimpanzees.

## **Surfing Uncertainty**

For many organizations, Hadoop is the first step for dealing with massive amounts of data. The next step? Processing and analyzing datasets with the Apache Pig scripting platform. With Pig, you can batch-process data without having to create a full-fledged application, making it easy to experiment with new datasets. Updated with use cases and programming examples, this second edition is the ideal learning tool for new and experienced users alike. You’ll find comprehensive coverage on key features such as the Pig Latin scripting language and the Grunt shell. When you need to analyze terabytes of data, this book shows you how to do it efficiently with Pig. Delve into Pig’s data model, including scalar and complex data types Write Pig Latin scripts to sort, group, join, project, and filter your data Use Grunt to work with the Hadoop Distributed File System (HDFS) Build complex data processing pipelines with Pig’s macros and modularity features Embed Pig Latin in Python for iterative processing and other advanced tasks Use Pig with Apache Tez to build high-

performance batch and interactive data processing applications Create your own load and store functions to handle data formats and storage mechanisms

## **West African Chimpanzees**

In this \"frightening and fascinating masterpiece\" (Walter Isaacson), David Quammen explores the true origins of HIV/AIDS.

## **Programming Pig**

How our collective intelligence has helped us to evolve and prosper Humans are a puzzling species. On the one hand, we struggle to survive on our own in the wild, often failing to overcome even basic challenges, like obtaining food, building shelters, or avoiding predators. On the other hand, human groups have produced ingenious technologies, sophisticated languages, and complex institutions that have permitted us to successfully expand into a vast range of diverse environments. What has enabled us to dominate the globe, more than any other species, while remaining virtually helpless as lone individuals? This book shows that the secret of our success lies not in our innate intelligence, but in our collective brains—on the ability of human groups to socially interconnect and learn from one another over generations. Drawing insights from lost European explorers, clever chimpanzees, mobile hunter-gatherers, neuroscientific findings, ancient bones, and the human genome, Joseph Henrich demonstrates how our collective brains have propelled our species' genetic evolution and shaped our biology. Our early capacities for learning from others produced many cultural innovations, such as fire, cooking, water containers, plant knowledge, and projectile weapons, which in turn drove the expansion of our brains and altered our physiology, anatomy, and psychology in crucial ways. Later on, some collective brains generated and recombined powerful concepts, such as the lever, wheel, screw, and writing, while also creating the institutions that continue to alter our motivations and perceptions. Henrich shows how our genetics and biology are inextricably interwoven with cultural evolution, and how culture-gene interactions launched our species on an extraordinary evolutionary trajectory. Tracking clues from our ancient past to the present, *The Secret of Our Success* explores how the evolution of both our cultural and social natures produce a collective intelligence that explains both our species' immense success and the origins of human uniqueness.

## **The Chimp and the River**

Lecturers - request an e-inspection copy of this text or contact your local SAGE representative to discuss your course needs. Watch Andy Field's introductory video to *Discovering Statistics Using R* Keeping the uniquely humorous and self-deprecating style that has made students across the world fall in love with Andy Field's books, *Discovering Statistics Using R* takes students on a journey of statistical discovery using R, a free, flexible and dynamically changing software tool for data analysis that is becoming increasingly popular across the social and behavioural sciences throughout the world. The journey begins by explaining basic statistical and research concepts before a guided tour of the R software environment. Next you discover the importance of exploring and graphing data, before moving onto statistical tests that are the foundations of the rest of the book (for example correlation and regression). You will then stride confidently into intermediate level analyses such as ANOVA, before ending your journey with advanced techniques such as MANOVA and multilevel models. Although there is enough theory to help you gain the necessary conceptual understanding of what you're doing, the emphasis is on applying what you learn to playful and real-world examples that should make the experience more fun than you might expect. Like its sister textbooks, *Discovering Statistics Using R* is written in an irreverent style and follows the same ground-breaking structure and pedagogical approach. The core material is augmented by a cast of characters to help the reader on their way, together with hundreds of examples, self-assessment tests to consolidate knowledge, and additional website material for those wanting to learn more. Given this book's accessibility, fun spirit, and use of bizarre real-world research it should be essential for anyone wanting to learn about statistics using the freely-available R software.

## **The Secret of Our Success**

An introduction to the science of neuroplasticity recounts the case stories of patients with mental limitations or brain damage whose seemingly unalterable conditions were improved through treatments that involved the thought re-alteration of brain structure.

## **Discovering Statistics Using R**

How to rewire your brain to improve virtually every aspect of your life-based on the latest research in neuroscience and psychology on neuroplasticity and evidence-based practices Not long ago, it was thought that the brain you were born with was the brain you would die with, and that the brain cells you had at birth were the most you would ever possess. Your brain was thought to be “hardwired” to function in predetermined ways. It turns out that's not true. Your brain is not hardwired, it's “softwired” by experience. This book shows you how you can rewire parts of the brain to feel more positive about your life, remain calm during stressful times, and improve your social relationships. Written by a leader in the field of Brain-Based Therapy, it teaches you how to activate the parts of your brain that have been underactivated and calm down those areas that have been hyperactivated so that you feel positive about your life and remain calm during stressful times. You will also learn to improve your memory, boost your mood, have better relationships, and get a good night sleep. Reveals how cutting-edge developments in neuroscience, and evidence-based practices can be used to improve your everyday life Other titles by Dr. Arden include: Brain-Based Therapy-Adult, Brain-Based Therapy-Child, Improving Your Memory For Dummies and Heal Your Anxiety Workbook Dr. Arden is a leader in integrating the new developments in neuroscience with psychotherapy and Director of Training in Mental Health for Kaiser Permanente for the Northern California Region Explaining exciting new developments in neuroscience and their applications to daily living, Rewire Your Brain will guide you through the process of changing your brain so you can change your life and be free of self-imposed limitations.

## **The Brain That Changes Itself**

“What makes you the way you are--and what makes each of us different from everyone else? In *Innate*, leading neuroscientist and popular science blogger Kevin Mitchell traces human diversity and individual differences to their deepest level: in the wiring of our brains. Deftly guiding us through important new research, including his own groundbreaking work, he explains how variations in the way our brains develop before birth strongly influence our psychology and behavior throughout our lives, shaping our personality, intelligence, sexuality, and even the way we perceive the world. We all share a genetic program for making a human brain, and the program for making a brain like yours is specifically encoded in your DNA. But, as Mitchell explains, the way that program plays out is affected by random processes of development that manifest uniquely in each person, even identical twins. The key insight of *Innate* is that the combination of these developmental and genetic variations creates innate differences in how our brains are wired--differences that impact all aspects of our psychology--and this insight promises to transform the way we see the interplay of nature and nurture. *Innate* also explores the genetic and neural underpinnings of disorders such as autism, schizophrenia, and epilepsy, and how our understanding of these conditions is being revolutionized. In addition, the book examines the social and ethical implications of these ideas and of new technologies that may soon offer the means to predict or manipulate human traits. Compelling and original, *Innate* will change the way you think about why and how we are who we are.”--Provided by the publisher.

## **Rewire Your Brain**

Provides a pan-African synthesis of community-based natural resource management (CBNRM), drawing on multiple authors and a wide range of documented experiences from Southern, Eastern, Western and Central Africa. This title discusses the degree to which CBNRM has met poverty alleviation, economic development

and nature conservation objectives.

## **Language Comprehension in Ape and Child**

Innate

[international business law](#)

[fda regulatory affairs third edition](#)

[james stewart calculus concepts and contexts 4th edition solutions](#)

[hi anxiety life with a bad case of nerves](#)

[hewlett packard hp vectra vl400 manual](#)

[robocut manual](#)

[sk garg environmental engineering vol 2 free download](#)

[sky above clouds finding our way through creativity aging and illness](#)

[honda easy start mower manual](#)

[cardiac surgery certification study guide](#)