

FAST DATA PROCESSING WITH SPARK SECOND EDITION

SPARK IN ACTION LEARNING SPARK LEARNING SPARK, 2ND EDITION ADVANCED ANALYTICS WITH SPARK DATA LAKE DEVELOPMENT WITH BIG DATA MACHINE LEARNING WITH SPARK CATALOGUE OF PRINTED BOOKS GETTING STARTED WITH HAZELCAST BRITISH MUSEUM CATALOGUE OF PRINTED BOOKS CATALOGUE OF THE PRINTED BOOKS IN THE LIBRARY OF THE FACULTY OF ADVOCATES CATALOGUE OF THE PRINTED BOOKS IN THE LIBRARY OF THE FACULTY OF ADVOCATES SPARK IN ACTION, SECOND EDITION PROCEEDINGS OF THE IEEE. GENERAL CATALOGUE OF PRINTED BOOKS THE ELECTRICIAN THE ELECTRICAL ENGINEER GENERAL CATALOGUE OF PRINTED BOOKS CATALOGUE OF MUSIC CATALOGUE OF THE PRINTED BOOKS IN THE LIBRARY OF THE FACULTY OF ADVOCATES ...: S-ZYPAEUS. 1878 GENERAL CATALOGUE OF PRINTED BOOKS TO 1955 JEAN-GEORGES PERRIN JULES S. DAMJI JULES DAMJI SANDY RYZA PRADEEP PASUPULETI RAJDEEP DUA BRITISH MUSEUM MAT JOHNS FACULTY OF ADVOCATES (SCOTLAND). LIBRARY ADVOCATES' LIBRARY JEAN-GEORGES PERRIN INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS BRITISH MUSEUM. DEPT. OF PRINTED BOOKS BRITISH MUSEUM. DEPARTMENT OF PRINTED BOOKS BRITISH MUSEUM FACULTY OF ADVOCATES (SCOTLAND). LIBRARY BRITISH MUSEUM. DEPT. OF PRINTED BOOKS

SPARK IN ACTION LEARNING SPARK LEARNING SPARK, 2ND EDITION ADVANCED ANALYTICS WITH SPARK DATA LAKE DEVELOPMENT WITH BIG DATA MACHINE LEARNING WITH SPARK CATALOGUE OF PRINTED BOOKS GETTING STARTED WITH HAZELCAST BRITISH MUSEUM CATALOGUE OF PRINTED BOOKS CATALOGUE OF THE PRINTED BOOKS IN THE LIBRARY OF THE FACULTY OF ADVOCATES CATALOGUE OF THE PRINTED BOOKS IN THE LIBRARY OF THE FACULTY OF ADVOCATES SPARK IN ACTION, SECOND EDITION PROCEEDINGS OF THE IEEE. GENERAL CATALOGUE OF PRINTED BOOKS THE ELECTRICIAN THE ELECTRICAL ENGINEER GENERAL CATALOGUE OF PRINTED BOOKS CATALOGUE OF MUSIC CATALOGUE OF THE PRINTED BOOKS IN THE LIBRARY OF THE FACULTY OF ADVOCATES ...: S-ZYPAEUS. 1878 GENERAL CATALOGUE OF PRINTED BOOKS TO 1955 JEAN-GEORGES PERRIN JULES S. DAMJI JULES DAMJI SANDY RYZA PRADEEP PASUPULETI RAJDEEP DUA BRITISH MUSEUM MAT JOHNS FACULTY OF ADVOCATES (SCOTLAND). LIBRARY ADVOCATES' LIBRARY JEAN-GEORGES PERRIN INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS BRITISH MUSEUM. DEPT. OF PRINTED BOOKS BRITISH MUSEUM. DEPARTMENT OF PRINTED BOOKS BRITISH MUSEUM FACULTY OF ADVOCATES (SCOTLAND). LIBRARY BRITISH MUSEUM. DEPT. OF PRINTED BOOKS

SUMMARY THE SPARK DISTRIBUTED DATA PROCESSING PLATFORM PROVIDES AN EASY TO IMPLEMENT TOOL FOR INGESTING STREAMING AND PROCESSING DATA FROM ANY SOURCE IN SPARK IN ACTION SECOND EDITION YOU LL LEARN TO TAKE ADVANTAGE OF SPARK S CORE FEATURES AND INCREDIBLE PROCESSING SPEED WITH APPLICATIONS INCLUDING REAL TIME COMPUTATION DELAYED EVALUATION AND MACHINE LEARNING SPARK SKILLS ARE A HOT COMMODITY IN ENTERPRISES WORLDWIDE AND WITH SPARK S POWERFUL AND FLEXIBLE JAVA APIS YOU CAN REAP ALL THE BENEFITS WITHOUT FIRST LEARNING SCALA OR HADOOP FOREWORD BY ROB THOMAS ABOUT THE TECHNOLOGY ANALYZING ENTERPRISE DATA STARTS BY READING FILTERING AND

MERGING FILES AND STREAMS FROM MANY SOURCES THE SPARK DATA PROCESSING ENGINE HANDLES THIS VARIED VOLUME LIKE A CHAMP DELIVERING SPEEDS 100 TIMES FASTER THAN HADOOP SYSTEMS THANKS TO SQL SUPPORT AN INTUITIVE INTERFACE AND A STRAIGHTFORWARD MULTILANGUAGE API YOU CAN USE SPARK WITHOUT LEARNING A COMPLEX NEW ECOSYSTEM ABOUT THE BOOK SPARK IN ACTION SECOND EDITION TEACHES YOU TO CREATE END TO END ANALYTICS APPLICATIONS IN THIS ENTIRELY NEW BOOK YOU LL LEARN FROM INTERESTING JAVA BASED EXAMPLES INCLUDING A COMPLETE DATA PIPELINE FOR PROCESSING NASA SATELLITE DATA AND YOU LL DISCOVER JAVA PYTHON AND SCALA CODE SAMPLES HOSTED ON GITHUB THAT YOU CAN EXPLORE AND ADAPT PLUS APPENDIXES THAT GIVE YOU A CHEAT SHEET FOR INSTALLING TOOLS AND UNDERSTANDING SPARK SPECIFIC TERMS WHAT S INSIDE WRITING SPARK APPLICATIONS IN JAVA SPARK APPLICATION ARCHITECTURE INGESTION THROUGH FILES DATABASES STREAMING AND ELASTICSEARCH QUERYING DISTRIBUTED DATASETS WITH SPARK SQL ABOUT THE READER THIS BOOK DOES NOT ASSUME PREVIOUS EXPERIENCE WITH SPARK SCALA OR HADOOP ABOUT THE AUTHOR JEAN GEORGES PERRIN IS AN EXPERIENCED DATA AND SOFTWARE ARCHITECT HE IS FRANCE S FIRST IBM CHAMPION AND HAS BEEN HONORED FOR 12 CONSECUTIVE YEARS

TABLE OF CONTENTS

PART 1 THE THEORY CRIPPLED BY AWESOME EXAMPLES

1 SO WHAT IS SPARK ANYWAY

2 ARCHITECTURE AND FLOW

3 THE MAJESTIC ROLE OF THE DATAFRAME

4 FUNDAMENTALLY LAZY

5 BUILDING A SIMPLE APP FOR DEPLOYMENT

6 DEPLOYING YOUR SIMPLE APP

PART 2 INGESTION

7 INGESTION FROM FILES

8 INGESTION FROM DATABASES

9 ADVANCED INGESTION FINDING DATA SOURCES AND BUILDING YOUR OWN

10 INGESTION THROUGH STRUCTURED STREAMING

PART 3 TRANSFORMING YOUR DATA

11 WORKING WITH SQL

12 TRANSFORMING YOUR DATA

13 TRANSFORMING ENTIRE DOCUMENTS

14 EXTENDING TRANSFORMATIONS WITH USER DEFINED FUNCTIONS

15 AGGREGATING YOUR DATA

PART 4 GOING FURTHER

16 CACHE AND CHECKPOINT ENHANCING SPARK S PERFORMANCES

17 EXPORTING DATA AND BUILDING FULL DATA PIPELINES

18 EXPLORING DEPLOYMENT

DATA IS BIGGER ARRIVES FASTER AND COMES IN A VARIETY OF FORMATS¹ AND IT ALL NEEDS TO BE PROCESSED AT SCALE FOR ANALYTICS OR MACHINE LEARNING BUT HOW CAN YOU PROCESS SUCH VARIED WORKLOADS EFFICIENTLY ENTER APACHE SPARK UPDATED TO INCLUDE SPARK 3 0 THIS SECOND EDITION SHOWS DATA ENGINEERS AND DATA SCIENTISTS WHY STRUCTURE AND UNIFICATION IN SPARK MATTERS SPECIFICALLY THIS BOOK EXPLAINS HOW TO PERFORM SIMPLE AND COMPLEX DATA ANALYTICS AND EMPLOY MACHINE LEARNING ALGORITHMS THROUGH STEP BY STEP WALK THROUGHS CODE SNIPPETS AND NOTEBOOKS YOU² LL BE ABLE TO LEARN PYTHON SQL SCALA OR JAVA HIGH LEVEL STRUCTURED APIS UNDERSTAND SPARK OPERATIONS AND SQL ENGINE INSPECT TUNE AND DEBUG SPARK OPERATIONS WITH SPARK CONFIGURATIONS AND SPARK UI CONNECT TO DATA SOURCES JSON PARQUET CSV AVRO ORC HIVE S3 OR KAFKA PERFORM ANALYTICS ON BATCH AND STREAMING DATA USING STRUCTURED STREAMING BUILD RELIABLE DATA PIPELINES WITH OPEN SOURCE DELTA LAKE AND SPARK DEVELOP MACHINE LEARNING PIPELINES WITH MLLIB AND PRODUCTIONIZE MODELS USING MLFLOW

DATA IS GETTING BIGGER ARRIVING FASTER AND COMING IN VARIED FORMATS AND IT ALL NEEDS TO BE PROCESSED AT SCALE FOR ANALYTICS OR MACHINE LEARNING HOW CAN YOU PROCESS SUCH VARIED DATA WORKLOADS EFFICIENTLY ENTER APACHE SPARK UPDATED TO EMPHASIZE NEW FEATURES IN SPARK 2 X THIS SECOND EDITION SHOWS DATA ENGINEERS AND SCIENTISTS WHY STRUCTURE AND UNIFICATION IN SPARK MATTERS SPECIFICALLY THIS BOOK EXPLAINS HOW TO PERFORM SIMPLE AND COMPLEX DATA ANALYTICS AND EMPLOY MACHINE LEARNING ALGORITHMS THROUGH DISCOURSE CODE SNIPPETS AND NOTEBOOKS YOU LL BE ABLE TO LEARN PYTHON SQL SCALA OR JAVA HIGH LEVEL APIS DATAFRAMES AND DATASETS PEEK UNDER THE HOOD OF THE SPARK SQL ENGINE TO UNDERSTAND SPARK TRANSFORMATIONS AND PERFORMANCE INSPECT TUNE AND DEBUG YOUR

SPARK OPERATIONS WITH SPARK CONFIGURATIONS AND SPARK UI CONNECT TO DATA SOURCES JSON PARQUET CSV AVRO ORC HIVE S3 OR KAFKA PERFORM ANALYTICS ON BATCH AND STREAMING DATA USING STRUCTURED STREAMING BUILD RELIABLE DATA PIPELINES WITH OPEN SOURCE DELTA LAKE AND SPARK DEVELOP MACHINE LEARNING PIPELINES WITH MLLIB AND PRODUCTIONIZE MODELS USING MLFLOW USE OPEN SOURCE PANDAS FRAMEWORK KOALAS AND SPARK FOR DATA TRANSFORMATION AND FEATURE ENGINEERING

IN THE SECOND EDITION OF THIS PRACTICAL BOOK FOUR CLOUDERA DATA SCIENTISTS PRESENT A SET OF SELF CONTAINED PATTERNS FOR PERFORMING LARGE SCALE DATA ANALYSIS WITH SPARK THE AUTHORS BRING SPARK STATISTICAL METHODS AND REAL WORLD DATA SETS TOGETHER TO TEACH YOU HOW TO APPROACH ANALYTICS PROBLEMS BY EXAMPLE UPDATED FOR SPARK 2.1 THIS EDITION ACTS AS AN INTRODUCTION TO THESE TECHNIQUES AND OTHER BEST PRACTICES IN SPARK PROGRAMMING YOU WILL START WITH AN INTRODUCTION TO SPARK AND ITS ECOSYSTEM AND THEN DIVE INTO PATTERNS THAT APPLY COMMON TECHNIQUES INCLUDING CLASSIFICATION CLUSTERING COLLABORATIVE FILTERING AND ANOMALY DETECTION TO FIELDS SUCH AS GENOMICS SECURITY AND FINANCE IF YOU HAVE AN ENTRY LEVEL UNDERSTANDING OF MACHINE LEARNING AND STATISTICS AND YOU PROGRAM IN JAVA PYTHON OR SCALA YOU WILL FIND THE BOOK'S PATTERNS USEFUL FOR WORKING ON YOUR OWN DATA APPLICATIONS WITH THIS BOOK YOU WILL FAMILIARIZE YOURSELF WITH THE SPARK PROGRAMMING MODEL BECOME COMFORTABLE WITHIN THE SPARK ECOSYSTEM LEARN GENERAL APPROACHES IN DATA SCIENCE EXAMINE COMPLETE IMPLEMENTATIONS THAT ANALYZE LARGE PUBLIC DATA SETS DISCOVER WHICH MACHINE LEARNING TOOLS MAKE SENSE FOR PARTICULAR PROBLEMS ACQUIRE CODE THAT CAN BE ADAPTED TO MANY USES

EXPLORE ARCHITECTURAL APPROACHES TO BUILDING DATA LAKES THAT INGEST INDEX MANAGE AND ANALYZE MASSIVE AMOUNTS OF DATA USING BIG DATA TECHNOLOGIES ABOUT THIS BOOK COMPREHEND THE INTRICACIES OF ARCHITECTING A DATA LAKE AND BUILD A DATA STRATEGY AROUND YOUR CURRENT DATA ARCHITECTURE EFFICIENTLY MANAGE VAST AMOUNTS OF DATA AND DELIVER IT TO MULTIPLE APPLICATIONS AND SYSTEMS WITH A HIGH DEGREE OF PERFORMANCE AND SCALABILITY PACKED WITH INDUSTRY BEST PRACTICES AND USE CASE SCENARIOS TO GET YOU UP AND RUNNING WHO THIS BOOK IS FOR THIS BOOK IS FOR ARCHITECTS AND SENIOR MANAGERS WHO ARE RESPONSIBLE FOR BUILDING A STRATEGY AROUND THEIR CURRENT DATA ARCHITECTURE HELPING THEM IDENTIFY THE NEED FOR A DATA LAKE IMPLEMENTATION IN AN ENTERPRISE CONTEXT THE READER WILL NEED A GOOD KNOWLEDGE OF MASTER DATA MANAGEMENT AND INFORMATION LIFECYCLE MANAGEMENT AND EXPERIENCE OF BIG DATA TECHNOLOGIES WHAT YOU WILL LEARN IDENTIFY THE NEED FOR A DATA LAKE IN YOUR ENTERPRISE CONTEXT AND LEARN TO ARCHITECT A DATA LAKE LEARN TO BUILD VARIOUS TIERS OF A DATA LAKE SUCH AS DATA INTAKE MANAGEMENT CONSUMPTION AND GOVERNANCE WITH A FOCUS ON PRACTICAL IMPLEMENTATION SCENARIOS FIND OUT THE KEY CONSIDERATIONS TO BE TAKEN INTO ACCOUNT WHILE BUILDING EACH TIER OF THE DATA LAKE UNDERSTAND HADOOP ORIENTED DATA TRANSFER MECHANISM TO INGEST DATA IN BATCH MICRO BATCH AND REAL TIME MODES EXPLORE VARIOUS DATA INTEGRATION NEEDS AND LEARN HOW TO PERFORM DATA ENRICHMENT AND DATA TRANSFORMATIONS USING BIG DATA TECHNOLOGIES ENABLE DATA DISCOVERY ON THE DATA LAKE TO ALLOW USERS TO DISCOVER THE DATA DISCOVER HOW DATA IS PACKAGED AND PROVISIONED FOR CONSUMPTION COMPREHEND THE IMPORTANCE OF INCLUDING DATA GOVERNANCE DISCIPLINES WHILE BUILDING A DATA LAKE IN DETAIL A DATA LAKE IS A HIGHLY SCALABLE PLATFORM FOR STORING HUGE VOLUMES OF MULTISTRUCTURED DATA FROM DISPARATE SOURCES WITH CENTRALIZED DATA MANAGEMENT SERVICES THIS BOOK EXPLORES THE POTENTIAL OF DATA LAKES AND EXPLORES ARCHITECTURAL APPROACHES TO BUILDING DATA LAKES THAT INGEST INDEX MANAGE AND ANALYZE MASSIVE AMOUNTS OF DATA USING BATCH AND REAL TIME PROCESSING FRAMEWORKS IT GUIDES YOU ON HOW TO GO

ABOUT BUILDING A DATA LAKE THAT IS MANAGED BY HADOOP AND ACCESSED AS REQUIRED BY OTHER BIG DATA APPLICATIONS THIS BOOK WILL GUIDE READERS USING BEST PRACTICES IN DEVELOPING DATA LAKE S CAPABILITIES IT WILL FOCUS ON ARCHITECT DATA GOVERNANCE SECURITY DATA QUALITY DATA LINEAGE TRACKING METADATA MANAGEMENT AND SEMANTIC DATA TAGGING BY THE END OF THIS BOOK YOU WILL HAVE A GOOD UNDERSTANDING OF BUILDING A DATA LAKE FOR BIG DATA STYLE AND APPROACH DATA LAKE DEVELOPMENT WITH BIG DATA PROVIDES ARCHITECTURAL APPROACHES TO BUILDING A DATA LAKE IT FOLLOWS A USE CASE BASED APPROACH WHERE PRACTICAL IMPLEMENTATION SCENARIOS OF EACH KEY COMPONENT ARE EXPLAINED IT ALSO HELPS YOU UNDERSTAND HOW THESE USE CASES ARE IMPLEMENTED IN A DATA LAKE THE CHAPTERS ARE ORGANIZED IN A WAY THAT MIMICS THE SEQUENTIAL DATA FLOW EVIDENCED IN A DATA LAKE

CREATE SCALABLE MACHINE LEARNING APPLICATIONS TO POWER A MODERN DATA DRIVEN BUSINESS USING SPARK 2 X ABOUT THIS BOOK GET TO THE GRIPS WITH THE LATEST VERSION OF APACHE SPARK UTILIZE SPARK S MACHINE LEARNING LIBRARY TO IMPLEMENT PREDICTIVE ANALYTICS LEVERAGE SPARK S POWERFUL TOOLS TO LOAD ANALYZE CLEAN AND TRANSFORM YOUR DATA WHO THIS BOOK IS FOR IF YOU HAVE A BASIC KNOWLEDGE OF MACHINE LEARNING AND WANT TO IMPLEMENT VARIOUS MACHINE LEARNING CONCEPTS IN THE CONTEXT OF SPARK ML THIS BOOK IS FOR YOU YOU SHOULD BE WELL VERSED WITH THE SCALA AND PYTHON LANGUAGES WHAT YOU WILL LEARN GET HANDS ON WITH THE LATEST VERSION OF SPARK ML CREATE YOUR FIRST SPARK PROGRAM WITH SCALA AND PYTHON SET UP AND CONFIGURE A DEVELOPMENT ENVIRONMENT FOR SPARK ON YOUR OWN COMPUTER AS WELL AS ON AMAZON EC2 ACCESS PUBLIC MACHINE LEARNING DATASETS AND USE SPARK TO LOAD PROCESS CLEAN AND TRANSFORM DATA USE SPARK S MACHINE LEARNING LIBRARY TO IMPLEMENT PROGRAMS BY UTILIZING WELL KNOWN MACHINE LEARNING MODELS DEAL WITH LARGE SCALE TEXT DATA INCLUDING FEATURE EXTRACTION AND USING TEXT DATA AS INPUT TO YOUR MACHINE LEARNING MODELS WRITE SPARK FUNCTIONS TO EVALUATE THE PERFORMANCE OF YOUR MACHINE LEARNING MODELS IN DETAIL THIS BOOK WILL TEACH YOU ABOUT POPULAR MACHINE LEARNING ALGORITHMS AND THEIR IMPLEMENTATION YOU WILL LEARN HOW VARIOUS MACHINE LEARNING CONCEPTS ARE IMPLEMENTED IN THE CONTEXT OF SPARK ML YOU WILL START BY INSTALLING SPARK IN A SINGLE AND MULTINODE CLUSTER NEXT YOU LL SEE HOW TO EXECUTE SCALA AND PYTHON BASED PROGRAMS FOR SPARK ML THEN WE WILL TAKE A FEW DATASETS AND GO DEEPER INTO CLUSTERING CLASSIFICATION AND REGRESSION TOWARD THE END WE WILL ALSO COVER TEXT PROCESSING USING SPARK ML ONCE YOU HAVE LEARNED THE CONCEPTS THEY CAN BE APPLIED TO IMPLEMENT ALGORITHMS IN EITHER GREEN FIELD IMPLEMENTATIONS OR TO MIGRATE EXISTING SYSTEMS TO THIS NEW PLATFORM YOU CAN MIGRATE FROM MAHOUT OR SCIKIT TO USE SPARK ML BY THE END OF THIS BOOK YOU WILL ACQUIRE THE SKILLS TO LEVERAGE SPARK S FEATURES TO CREATE YOUR OWN SCALABLE MACHINE LEARNING APPLICATIONS AND POWER A MODERN DATA DRIVEN BUSINESS STYLE AND APPROACH THIS PRACTICAL TUTORIAL WITH REAL WORLD USE CASES ENABLES YOU TO DEVELOP YOUR OWN MACHINE LEARNING SYSTEMS WITH SPARK THE EXAMPLES WILL HELP YOU COMBINE VARIOUS TECHNIQUES AND MODELS INTO AN INTELLIGENT MACHINE LEARNING SYSTEM

THIS BOOK IS AN EASY TO FOLLOW HANDS ON INTRODUCTION THAT GUIDES YOU THROUGH THIS INNOVATIVE NEW TECHNOLOGY IT COVERS EVERYTHING FROM DATA GRIDS TO THE SIMPLE TO USE DISTRIBUTED DATA STORAGE COLLECTIONS QUEUING AND TOPIC MESSAGING CAPABILITIES AS WELL AS LOCKING AND TRANSACTION SUPPORT TO GUARD AGAINST CONCURRENCY RACE CONDITIONS ARE SOME OF THE TOPICS THAT WE WILL COVER WE WILL THEN MOVE ON TO DISTRIBUTED TASK EXECUTION IN PLACE DATA MANIPULATIONS AND BIG DATA ANALYTICAL PROCESSING USING MAPREDUCE AT THE END OF ALL THIS YOU WILL BE ARMED WITH EVERYTHING YOU NEED TO BRING AMAZING POWER AND DATA SCALABILITY TO YOUR

APPLICATIONS AS WELL AS MAKING THEM TRULY GLOBAL AND READY FOR A WORLDWIDE AUDIENCE

SUMMARY THE SPARK DISTRIBUTED DATA PROCESSING PLATFORM PROVIDES AN EASY TO IMPLEMENT TOOL FOR INGESTING STREAMING AND PROCESSING DATA FROM ANY SOURCE IN SPARK IN ACTION SECOND EDITION YOU LL LEARN TO TAKE ADVANTAGE OF SPARK S CORE FEATURES AND INCREDIBLE PROCESSING SPEED WITH APPLICATIONS INCLUDING REAL TIME COMPUTATION DELAYED EVALUATION AND MACHINE LEARNING SPARK SKILLS ARE A HOT COMMODITY IN ENTERPRISES WORLDWIDE AND WITH SPARK S POWERFUL AND FLEXIBLE JAVA APIS YOU CAN REAP ALL THE BENEFITS WITHOUT FIRST LEARNING SCALA OR HADOOP FOREWORD BY ROB THOMAS PURCHASE OF THE PRINT BOOK INCLUDES A FREE EBOOK IN PDF KINDLE AND EPUB FORMATS FROM MANNING PUBLICATIONS ABOUT THE TECHNOLOGY ANALYZING ENTERPRISE DATA STARTS BY READING FILTERING AND MERGING FILES AND STREAMS FROM MANY SOURCES THE SPARK DATA PROCESSING ENGINE HANDLES THIS VARIED VOLUME LIKE A CHAMP DELIVERING SPEEDS 100 TIMES FASTER THAN HADOOP SYSTEMS THANKS TO SQL SUPPORT AN INTUITIVE INTERFACE AND A STRAIGHTFORWARD MULTILANGUAGE API YOU CAN USE SPARK WITHOUT LEARNING A COMPLEX NEW ECOSYSTEM ABOUT THE BOOK SPARK IN ACTION SECOND EDITION TEACHES YOU TO CREATE END TO END ANALYTICS APPLICATIONS IN THIS ENTIRELY NEW BOOK YOU LL LEARN FROM INTERESTING JAVA BASED EXAMPLES INCLUDING A COMPLETE DATA PIPELINE FOR PROCESSING NASA SATELLITE DATA AND YOU LL DISCOVER JAVA PYTHON AND SCALA CODE SAMPLES HOSTED ON GITHUB THAT YOU CAN EXPLORE AND ADAPT PLUS APPENDIXES THAT GIVE YOU A CHEAT SHEET FOR INSTALLING TOOLS AND UNDERSTANDING SPARK SPECIFIC TERMS WHAT S INSIDE WRITING SPARK APPLICATIONS IN JAVA SPARK APPLICATION ARCHITECTURE INGESTION THROUGH FILES DATABASES STREAMING AND ELASTICSEARCH QUERYING DISTRIBUTED DATASETS WITH SPARK SQL ABOUT THE READER THIS BOOK DOES NOT ASSUME PREVIOUS EXPERIENCE WITH SPARK SCALA OR HADOOP ABOUT THE AUTHOR JEAN GEORGES PERRIN IS AN EXPERIENCED DATA AND SOFTWARE ARCHITECT HE IS FRANCE S FIRST IBM CHAMPION AND HAS BEEN HONORED FOR 12 CONSECUTIVE YEARS TABLE OF CONTENTS PART 1 THE THEORY CRIPPLED BY AWESOME EXAMPLES 1 SO WHAT IS SPARK ANYWAY 2 ARCHITECTURE AND FLOW 3 THE MAJESTIC ROLE OF THE DATAFRAME 4 FUNDAMENTALLY LAZY 5 BUILDING A SIMPLE APP FOR DEPLOYMENT 6 DEPLOYING YOUR SIMPLE APP PART 2 INGESTION 7 INGESTION FROM FILES 8 INGESTION FROM DATABASES 9 ADVANCED INGESTION FINDING DATA SOURCES AND BUILDING YOUR OWN 10 INGESTION THROUGH STRUCTURED STREAMING PART 3 TRANSFORMING YOUR DATA 11 WORKING WITH SQL 12 TRANSFORMING YOUR DATA 13 TRANSFORMING ENTIRE DOCUMENTS 14 EXTENDING TRANSFORMATIONS WITH USER DEFINED FUNCTIONS 15 AGGREGATING YOUR DATA PART 4 GOING FURTHER 16 CACHE AND CHECKPOINT ENHANCING SPARK S PERFORMANCES 17 EXPORTING DATA AND BUILDING FULL DATA PIPELINES 18 EXPLORING DEPLOYMENT

VOLS 34 INCLUDE SECTION WAVES AND ELECTRONS

THE COLLECTIONS OF THE ADVOCATES LIBRARY WITH THE EXCEPTION OF ITS LEGAL BOOKS AND MANUSCRIPTS WERE GIVEN BY THE ADVOCATES TO THE NATIONAL LIBRARY OF SCOTLAND IN 1925

IF YOU ALLY INFATUATION SUCH A REFERRED **FAST DATA PROCESSING WITH SPARK SECOND EDITION** BOOKS THAT WILL GIVE YOU WORTH, ACQUIRE THE NO QUESTION BEST SELLER FROM US CURRENTLY FROM SEVERAL PREFERRED AUTHORS. IF YOU DESIRE TO HUMOROUS BOOKS, LOTS OF NOVELS, TALE, JOKES, AND MORE FICTIONS COLLECTIONS ARE NEXT LAUNCHED, FROM BEST SELLER TO ONE OF THE MOST CURRENT RELEASED. YOU MAY NOT BE

PERPLEXED TO ENJOY ALL BOOK COLLECTIONS FAST DATA PROCESSING WITH SPARK SECOND EDITION THAT WE WILL CERTAINLY OFFER. IT IS NOT AS REGARDS THE COSTS. ITS PRACTICALLY WHAT YOU DEPENDENCE CURRENTLY. THIS FAST DATA PROCESSING WITH SPARK SECOND EDITION, AS ONE OF THE MOST OPERATING SELLERS HERE WILL CERTAINLY BE AMONG THE BEST OPTIONS TO REVIEW.

1. WHAT IS A FAST DATA PROCESSING WITH SPARK SECOND EDITION PDF? A PDF (PORTABLE DOCUMENT FORMAT) IS A FILE FORMAT DEVELOPED BY ADOBE THAT PRESERVES THE LAYOUT AND FORMATTING OF A DOCUMENT, REGARDLESS OF THE SOFTWARE, HARDWARE, OR OPERATING SYSTEM USED TO VIEW OR PRINT IT.
2. HOW DO I CREATE A FAST DATA PROCESSING WITH SPARK SECOND EDITION PDF? THERE ARE SEVERAL WAYS TO CREATE A PDF:
3. USE SOFTWARE LIKE ADOBE ACROBAT, MICROSOFT WORD, OR GOOGLE DOCS, WHICH OFTEN HAVE BUILT-IN PDF CREATION TOOLS. PRINT TO PDF: MANY APPLICATIONS AND OPERATING SYSTEMS HAVE A "PRINT TO PDF" OPTION THAT ALLOWS YOU TO SAVE A DOCUMENT AS A PDF FILE INSTEAD OF PRINTING IT ON PAPER. ONLINE CONVERTERS: THERE ARE VARIOUS ONLINE TOOLS THAT CAN CONVERT DIFFERENT FILE TYPES TO PDF.
4. HOW DO I EDIT A FAST DATA PROCESSING WITH SPARK SECOND EDITION PDF? EDITING A PDF CAN BE DONE WITH SOFTWARE LIKE ADOBE ACROBAT, WHICH ALLOWS DIRECT EDITING OF TEXT, IMAGES, AND OTHER ELEMENTS WITHIN THE PDF. SOME FREE TOOLS, LIKE PDFESCAPE OR SMALLPDF, ALSO OFFER BASIC EDITING CAPABILITIES.
5. HOW DO I CONVERT A FAST DATA PROCESSING WITH SPARK SECOND EDITION PDF TO ANOTHER FILE FORMAT? THERE ARE MULTIPLE WAYS TO CONVERT A PDF TO ANOTHER FORMAT:
6. USE ONLINE CONVERTERS LIKE SMALLPDF, ZAMZAR, OR ADOBE ACROBATS EXPORT FEATURE TO CONVERT PDFs TO FORMATS LIKE WORD, EXCEL, JPEG, ETC. SOFTWARE LIKE ADOBE ACROBAT, MICROSOFT WORD, OR OTHER PDF EDITORS MAY HAVE OPTIONS TO EXPORT OR SAVE PDFs IN DIFFERENT FORMATS.
7. HOW DO I PASSWORD-PROTECT A FAST DATA PROCESSING WITH SPARK SECOND EDITION PDF? MOST PDF EDITING SOFTWARE ALLOWS YOU TO ADD PASSWORD PROTECTION. IN ADOBE ACROBAT, FOR INSTANCE, YOU CAN GO TO "FILE" -> "PROPERTIES" -> "SECURITY" TO SET A PASSWORD TO RESTRICT ACCESS OR EDITING CAPABILITIES.
8. ARE THERE ANY FREE ALTERNATIVES TO ADOBE ACROBAT FOR WORKING WITH PDFs? YES, THERE ARE MANY FREE ALTERNATIVES FOR WORKING WITH PDFs, SUCH AS:
9. LIBREOFFICE: OFFERS PDF EDITING FEATURES. PDFSAM: ALLOWS SPLITTING, MERGING, AND EDITING PDFs. FOXIT READER: PROVIDES BASIC PDF VIEWING AND EDITING CAPABILITIES.
10. HOW DO I COMPRESS A PDF FILE? YOU CAN USE ONLINE TOOLS LIKE SMALLPDF, ILOVEPDF, OR DESKTOP SOFTWARE LIKE ADOBE ACROBAT TO COMPRESS PDF FILES WITHOUT SIGNIFICANT QUALITY LOSS. COMPRESSION REDUCES THE FILE SIZE, MAKING IT EASIER TO SHARE AND DOWNLOAD.
11. CAN I FILL OUT FORMS IN A PDF FILE? YES, MOST PDF VIEWERS/EDITORS LIKE ADOBE ACROBAT, PREVIEW (ON MAC), OR VARIOUS ONLINE TOOLS ALLOW YOU TO FILL OUT FORMS IN PDF FILES BY SELECTING TEXT FIELDS AND ENTERING INFORMATION.
12. ARE THERE ANY RESTRICTIONS WHEN WORKING WITH PDFs? SOME PDFs MIGHT HAVE RESTRICTIONS SET BY THEIR CREATOR, SUCH AS PASSWORD PROTECTION, EDITING RESTRICTIONS, OR PRINT RESTRICTIONS. BREAKING THESE RESTRICTIONS MIGHT REQUIRE SPECIFIC SOFTWARE OR TOOLS, WHICH MAY OR MAY NOT BE LEGAL DEPENDING ON THE CIRCUMSTANCES AND LOCAL LAWS.

INTRODUCTION

THE DIGITAL AGE HAS REVOLUTIONIZED THE WAY WE READ, MAKING BOOKS MORE ACCESSIBLE THAN EVER. WITH THE RISE OF EBOOKS, READERS CAN NOW CARRY ENTIRE LIBRARIES IN THEIR POCKETS. AMONG THE VARIOUS SOURCES FOR EBOOKS, FREE EBOOK SITES HAVE EMERGED AS A POPULAR CHOICE. THESE SITES OFFER A TREASURE TROVE OF KNOWLEDGE AND ENTERTAINMENT WITHOUT THE COST. BUT WHAT MAKES THESE SITES SO VALUABLE, AND WHERE CAN YOU FIND THE BEST ONES? LET'S DIVE INTO THE WORLD OF FREE EBOOK SITES.

BENEFITS OF FREE EBOOK SITES

WHEN IT COMES TO READING, FREE EBOOK SITES OFFER NUMEROUS ADVANTAGES.

COST SAVINGS

FIRST AND FOREMOST, THEY SAVE YOU MONEY. BUYING BOOKS CAN BE EXPENSIVE, ESPECIALLY IF YOU'RE AN AVID READER. FREE EBOOK SITES ALLOW YOU TO ACCESS A VAST ARRAY OF BOOKS WITHOUT SPENDING A DIME.

ACCESSIBILITY

THESE SITES ALSO ENHANCE ACCESSIBILITY. WHETHER YOU'RE AT HOME, ON THE GO, OR HALFWAY AROUND THE WORLD, YOU CAN ACCESS YOUR FAVORITE TITLES ANYTIME, ANYWHERE, PROVIDED YOU HAVE AN INTERNET CONNECTION.

VARIETY OF CHOICES

MOREOVER, THE VARIETY OF CHOICES AVAILABLE IS ASTOUNDING. FROM CLASSIC LITERATURE TO CONTEMPORARY NOVELS, ACADEMIC TEXTS TO CHILDREN'S BOOKS, FREE EBOOK SITES COVER ALL GENRES AND INTERESTS.

TOP FREE EBOOK SITES

THERE ARE COUNTLESS FREE EBOOK SITES, BUT A FEW STAND OUT FOR THEIR QUALITY AND RANGE OF OFFERINGS.

PROJECT GUTENBERG

PROJECT GUTENBERG IS A PIONEER IN OFFERING FREE EBOOKS. WITH OVER 60,000 TITLES, THIS SITE PROVIDES A WEALTH OF CLASSIC LITERATURE IN THE PUBLIC DOMAIN.

OPEN LIBRARY

OPEN LIBRARY AIMS TO HAVE A WEBPAGE FOR EVERY BOOK EVER PUBLISHED. IT OFFERS MILLIONS OF FREE EBOOKS, MAKING IT A FANTASTIC RESOURCE FOR READERS.

GOOGLE BOOKS

GOOGLE BOOKS ALLOWS USERS TO SEARCH AND PREVIEW MILLIONS OF BOOKS FROM LIBRARIES AND PUBLISHERS WORLDWIDE. WHILE NOT ALL BOOKS ARE AVAILABLE FOR FREE, MANY ARE.

MANYBOOKS

MANYBOOKS OFFERS A LARGE SELECTION OF FREE EBOOKS IN VARIOUS GENRES. THE SITE IS USER-FRIENDLY AND OFFERS BOOKS IN MULTIPLE FORMATS.

BOOKBOON

BOOKBOON SPECIALIZES IN FREE TEXTBOOKS AND BUSINESS BOOKS, MAKING IT AN EXCELLENT RESOURCE FOR STUDENTS AND PROFESSIONALS.

HOW TO DOWNLOAD EBOOKS SAFELY

DOWNLOADING EBOOKS SAFELY IS CRUCIAL TO AVOID PIRATED CONTENT AND PROTECT YOUR DEVICES.

AVOIDING PIRATED CONTENT

STICK TO REPUTABLE SITES TO ENSURE YOU'RE NOT DOWNLOADING PIRATED CONTENT. PIRATED EBOOKS NOT ONLY HARM AUTHORS AND PUBLISHERS BUT CAN ALSO POSE SECURITY RISKS.

ENSURING DEVICE SAFETY

ALWAYS USE ANTIVIRUS SOFTWARE AND KEEP YOUR DEVICES UPDATED TO PROTECT AGAINST MALWARE THAT CAN BE HIDDEN IN DOWNLOADED FILES.

LEGAL CONSIDERATIONS

BE AWARE OF THE LEGAL CONSIDERATIONS WHEN DOWNLOADING EBOOKS. ENSURE THE SITE HAS THE RIGHT TO DISTRIBUTE THE BOOK AND THAT YOU'RE NOT VIOLATING COPYRIGHT LAWS.

USING FREE EBOOK SITES FOR EDUCATION

FREE EBOOK SITES ARE INVALUABLE FOR EDUCATIONAL PURPOSES.

ACADEMIC RESOURCES

SITES LIKE PROJECT GUTENBERG AND OPEN LIBRARY OFFER NUMEROUS ACADEMIC RESOURCES, INCLUDING TEXTBOOKS AND SCHOLARLY ARTICLES.

LEARNING NEW SKILLS

YOU CAN ALSO FIND BOOKS ON VARIOUS SKILLS, FROM COOKING TO PROGRAMMING, MAKING THESE SITES GREAT FOR PERSONAL DEVELOPMENT.

SUPPORTING HOMESCHOOLING

FOR HOMESCHOOLING PARENTS, FREE EBOOK SITES PROVIDE A WEALTH OF EDUCATIONAL MATERIALS FOR DIFFERENT GRADE LEVELS AND SUBJECTS.

GENRES AVAILABLE ON FREE EBOOK SITES

THE DIVERSITY OF GENRES AVAILABLE ON FREE EBOOK SITES ENSURES THERE'S SOMETHING FOR EVERYONE.

FICTION

FROM TIMELESS CLASSICS TO CONTEMPORARY BESTSELLERS, THE FICTION SECTION IS BRIMMING WITH OPTIONS.

Non-Fiction

Non-fiction enthusiasts can find biographies, self-help books, historical texts, and more.

Textbooks

Students can access textbooks on a wide range of subjects, helping reduce the financial burden of education.

Children's Books

Parents and teachers can find a plethora of children's books, from picture books to young adult novels.

Accessibility Features of Ebook Sites

Ebook sites often come with features that enhance accessibility.

Audiobook Options

Many sites offer audiobooks, which are great for those who prefer listening to reading.

Adjustable Font Sizes

You can adjust the font size to suit your reading comfort, making it easier for those with visual impairments.

Text-to-Speech Capabilities

Text-to-speech features can convert written text into audio, providing an alternative way to enjoy books.

Tips for Maximizing Your Ebook Experience

To make the most out of your ebook reading experience, consider these tips.

CHOOSING THE RIGHT DEVICE

WHETHER IT'S A TABLET, AN E-READER, OR A SMARTPHONE, CHOOSE A DEVICE THAT OFFERS A COMFORTABLE READING EXPERIENCE FOR YOU.

ORGANIZING YOUR EBOOK LIBRARY

USE TOOLS AND APPS TO ORGANIZE YOUR EBOOK COLLECTION, MAKING IT EASY TO FIND AND ACCESS YOUR FAVORITE TITLES.

SYNCING ACROSS DEVICES

MANY EBOOK PLATFORMS ALLOW YOU TO SYNC YOUR LIBRARY ACROSS MULTIPLE DEVICES, SO YOU CAN PICK UP RIGHT WHERE YOU LEFT OFF, NO MATTER WHICH DEVICE YOU'RE USING.

CHALLENGES AND LIMITATIONS

DESPITE THE BENEFITS, FREE EBOOK SITES COME WITH CHALLENGES AND LIMITATIONS.

QUALITY AND AVAILABILITY OF TITLES

NOT ALL BOOKS ARE AVAILABLE FOR FREE, AND SOMETIMES THE QUALITY OF THE DIGITAL COPY CAN BE POOR.

DIGITAL RIGHTS MANAGEMENT (DRM)

DRM CAN RESTRICT HOW YOU USE THE EBOOKS YOU DOWNLOAD, LIMITING SHARING AND TRANSFERRING BETWEEN DEVICES.

INTERNET DEPENDENCY

ACCESSING AND DOWNLOADING EBOOKS REQUIRES AN INTERNET CONNECTION, WHICH CAN BE A LIMITATION IN AREAS WITH POOR CONNECTIVITY.

FUTURE OF FREE EBOOK SITES

THE FUTURE LOOKS PROMISING FOR FREE EBOOK SITES AS TECHNOLOGY CONTINUES TO ADVANCE.

TECHNOLOGICAL ADVANCES

IMPROVEMENTS IN TECHNOLOGY WILL LIKELY MAKE ACCESSING AND READING EBOOKS EVEN MORE SEAMLESS AND ENJOYABLE.

EXPANDING ACCESS

EFFORTS TO EXPAND INTERNET ACCESS GLOBALLY WILL HELP MORE PEOPLE BENEFIT FROM FREE EBOOK SITES.

ROLE IN EDUCATION

AS EDUCATIONAL RESOURCES BECOME MORE DIGITIZED, FREE EBOOK SITES WILL PLAY AN INCREASINGLY VITAL ROLE IN LEARNING.

CONCLUSION

IN SUMMARY, FREE EBOOK SITES OFFER AN INCREDIBLE OPPORTUNITY TO ACCESS A WIDE RANGE OF BOOKS WITHOUT THE FINANCIAL BURDEN. THEY ARE INVALUABLE RESOURCES FOR READERS OF ALL AGES AND INTERESTS, PROVIDING EDUCATIONAL MATERIALS, ENTERTAINMENT, AND ACCESSIBILITY FEATURES. SO WHY NOT EXPLORE THESE SITES AND DISCOVER THE WEALTH OF KNOWLEDGE THEY OFFER?

FAQs

ARE FREE EBOOK SITES LEGAL? YES, MOST FREE EBOOK SITES ARE LEGAL. THEY TYPICALLY OFFER BOOKS THAT ARE IN THE PUBLIC DOMAIN OR HAVE THE RIGHTS TO DISTRIBUTE THEM. HOW DO I KNOW IF AN EBOOK SITE IS SAFE? STICK TO WELL-KNOWN AND REPUTABLE SITES LIKE PROJECT GUTENBERG, OPEN LIBRARY, AND GOOGLE BOOKS. CHECK REVIEWS AND ENSURE THE SITE HAS PROPER SECURITY MEASURES. CAN I DOWNLOAD EBOOKS TO ANY DEVICE? MOST FREE EBOOK SITES OFFER DOWNLOADS IN MULTIPLE FORMATS, MAKING THEM COMPATIBLE WITH VARIOUS DEVICES LIKE E-READERS, TABLETS, AND SMARTPHONES. DO FREE EBOOK SITES OFFER AUDIOBOOKS? MANY FREE EBOOK SITES OFFER AUDIOBOOKS, WHICH ARE PERFECT FOR THOSE WHO PREFER LISTENING TO THEIR BOOKS. HOW CAN I SUPPORT AUTHORS IF I USE FREE EBOOK SITES? YOU CAN SUPPORT AUTHORS BY PURCHASING THEIR BOOKS WHEN POSSIBLE, LEAVING REVIEWS, AND SHARING THEIR WORK WITH OTHERS.

