

## Mastering Your Data

Multi-Domain Master Data Management delivers practical guidance and specific instruction to help guide planners and practitioners through the challenges of a multi-domain master data management (MDM) implementation. Authors Mark Allen and Dalton Cervo bring their expertise to you in the only reference you need to help your organization take master data management to the next level by incorporating it across multiple domains. Written in a business friendly style with sufficient program planning guidance, this book covers a comprehensive set of topics and advanced strategies centered on the key MDM disciplines of Data Governance, Data Stewardship, Data Quality Management, Metadata Management, and Data Integration. Provides a logical order toward planning, implementation, and ongoing management of multi-domain MDM from a program manager and data steward perspective. Provides detailed guidance, examples and illustrations for MDM practitioners to apply these insights to their strategies, plans, and processes. Covers advanced MDM strategy and instruction aimed at improving data quality management, lowering data maintenance costs, and reducing corporate risks by applying consistent enterprise-wide practices for the management and control of master data.

Data modeling is one of the most critical phases in the database application development process, but also the phase most likely to fail. A master data modeler must come into any organization, understand its data requirements, and skillfully model the data for applications that most effectively serve organizational needs. Mastering Data Modeling is a complete guide to becoming a successful data modeler. Featuring a requirements-driven approach, this book clearly explains fundamental concepts, introduces a user-oriented data modeling notation, and describes a rigorous, step-by-step process for collecting, modeling, and documenting the kinds of data that users need. Assuming no prior knowledge, Mastering Data Modeling sets forth several fundamental problems of data modeling, such as reconciling the software developer's demand for rigor with the users' equally valid need to speak their own (sometimes vague) natural language. In addition, it describes the good habits that help you respond to these fundamental problems. With these good habits in mind, the book describes the Logical Data Structure (LDS) notation and the process of controlled evolution by which you can create low-cost, user-approved data models that resist premature obsolescence. Also included is an encyclopedic analysis of all data shapes that you will encounter. Most notably, the book describes The Flow, a loosely scripted process by which you and the users gradually but continuously improve an LDS until it faithfully represents the information needs. Essential implementation and technology issues are also covered. You will learn about such vital topics as: The fundamental problems of data modeling The good habits that help a data modeler be effective and economical LDS notation, which encourages these good habits How to read an LDS aloud--in declarative English sentences How to write a well-formed (syntactically correct) LDS How to get users to name the parts of an LDS with words from their own business vocabulary How to visualize data for an LDS A catalog of LDS shapes that recur throughout all data models The Flow--the template for your conversations with users How to document an LDS for users, data modelers, and technologists How to map an LDS to a relational schema How LDS differs from other notations and why "Story interludes" appear throughout the book, illustrating real-world successes of the LDS notation and controlled evolution process.

Numerous exercises help you master critical skills. In addition, two detailed, annotated sample conversations with users show you the process of controlled evolution in action.

Gain sharp insights into your data and solve real-world data science problems with R—from data munging to modeling and visualization About This Book Handle your data with precision and care for optimal business intelligence Restructure and transform your data to inform decision-making Packed with practical advice and tips to help you get to grips with data mining

Who This Book Is For If you are a data scientist or R developer who wants to explore and optimize your use of R's advanced features and tools, this is the book for you. A basic knowledge of R is required, along with an understanding of database logic. What You Will Learn Connect to and load data from R's range of powerful databases Successfully fetch and parse structured and unstructured data Transform and restructure your data with efficient R packages Define and build complex statistical models with glm Develop and train machine learning algorithms Visualize social networks and graph data Deploy supervised and unsupervised classification algorithms Discover how to visualize spatial data with R In Detail R is an essential language for sharp and successful data analysis. Its numerous features and ease of use make it a powerful way of mining, managing, and interpreting large sets of data. In a world where understanding big data has become key, by mastering R you will be able to deal with your data effectively and efficiently. This book will give you the guidance you need to build and develop your knowledge and expertise. Bridging the gap between theory and practice, this book will help you to understand and use data for a competitive advantage. Beginning with taking you through essential data mining and management tasks such as munging, fetching, cleaning, and restructuring, the book then explores different model designs and the core components of effective analysis. You will then discover how to optimize your use of machine learning algorithms for classification and recommendation systems beside the traditional and more recent statistical methods. Style and approach Covering the essential tasks and skills within data science, Mastering Data Analysis provides you with solutions to the challenges of data science. Each section gives you a theoretical overview before demonstrating how to put the theory to work with real-world use cases and hands-on examples.

If you are convinced that the world today is producing more data than the previous decades, then you understand that processing yesterday's data for today's use at times is not enough. The level of data analysis that is needed in highly competitive business environment needs to be processed, analyzed and used immediately for businesses to be ahead of their competition. Having this in mind, you need to understand from the ground up, what data is, the different types of data and how you should identify the right data for your business. To help you understand the simple basics of data and how it needs to be analyzed, then Data Analytics for Beginners is the book that you have been waiting for. The size and type of business you are running doesn't matter because after all, it will depend on your ability to understand the data that your business is exposed to so as to make better business decisions for the current working environment and the future. Are there patterns in your business that you cannot see? Do you want to make sense of the shopping trends of your clients to better enrich their experience? Do you want to know your target market even more? Do you want to better derive insights from the feedback your clients give you? These questions can only be answered when you perform a data analysis for your business. Collecting the data is one thing, analyzing them is another matter entirely as it is not something that can be done haphazardly by just looking at the data. If you hope to understand your data well, you need to understand the data you are collecting, the methods to use and the right tools to use when analyzing the data. Inside you will find valuable steps and tools that will help make your information work for you. Do not let yourself get complacent, stop looking at the data that you collect each day and start analyzing your data to move your business up. Get started by buying this book today! Inside you will find How data should be understood? Terms and concepts used in data analysis. Data mining and the different kinds of databases used to store data. How information can be retrieved and manipulated in the database to create a visual representation of what you want to know? The life cycle of data analysis. And more...

This book will show you how to use Power BI effectively to create a variety of visualizations and BI dashboards. Right from gathering data through various data sources, you will learn to perform effective visual analytics. By the end of this book, you will be able to gain unique,

hidden insights into your data using Microsoft Power BI.

Master the intricacies of Tableau to create effective data visualizations About This Book Arm yourself with an arsenal of advanced chart types and geocoding to efficiently and engagingly present information Map a grid over a network node diagram and use that grid to demonstrate loads, processing time, and more in Tableau Integrate R with Tableau by utilizing R functions, libraries, and saved models Who This Book Is For If you are a business analyst without developer-level programming skills, then this book is for you. You are expected to have at least a fundamental understanding of Tableau and basic knowledge of joins, however SQL knowledge is not assumed. You should have basic computer skills, including at least moderate Excel proficiency. What You Will Learn Create a worksheet that can display the current balance for any given period in time Recreate a star schema from in a data warehouse in Tableau Combine level of detail calculations with table calculations, sets, and parameters Create custom polygons to build filled maps for area codes in the USA Visualize data using a set of analytical and advanced charting techniques Know when to use Tableau instead of PowerPoint Build a dashboard and export it to PowerPoint In Detail Tableau has emerged as one of the most popular Business Intelligence solutions in recent times, thanks to its powerful and interactive data visualization capabilities. This book will empower you to become a master in Tableau by exploiting the many new features introduced in Tableau 10.0. You will embark on this exciting journey by getting to know the valuable methods of utilizing advanced calculations to solve complex problems. These techniques include creative use of different types of calculations such as row-level, aggregate-level, and more. You will discover how almost any data visualization challenge can be met in Tableau by getting a proper understanding of the tool's inner workings and creatively exploring possibilities. You'll be armed with an arsenal of advanced chart types and techniques to enable you to efficiently and engagingly present information to a variety of audiences through the use of clear, efficient, and engaging dashboards. Explanations and examples of efficient and inefficient visualization techniques, well-designed and poorly designed dashboards, and compromise options when Tableau consumers will not embrace data visualization will build on your understanding of Tableau and how to use it efficiently. By the end of the book, you will be equipped with all the information you need to create effective dashboards and data visualization solutions using Tableau. Style and approach This book takes a direct approach, to systematically evolve to more involved functionalities such as advanced calculation, parameters & sets, data blending and R integration. This book will help you gain skill in building visualizations previously beyond your capacity.

Master the programming language of choice among statisticians and data analysts worldwide Coming to grips with R can be tough, even for seasoned statisticians and data analysts. Enter R For Dummies, the quick, easy way to master all the R you'll ever need. Requiring no prior programming experience and packed with practical examples, easy, step-by-step exercises, and sample code, this extremely accessible guide is the ideal introduction to R for complete beginners. It also covers many concepts that intermediate-level programmers will find extremely useful. Master your R ABCs ? get up to speed in no time with the basics, from installing and configuring R to writing simple scripts and performing simultaneous calculations on many variables Put data in its place ? get to know your way around lists, data frames, and other R data structures while learning to interact with other programs, such as Microsoft Excel Make data dance to your tune ? learn how to reshape and manipulate data, merge data sets, split and combine data, perform calculations on vectors and arrays, and much more Visualize it ? learn to use R's powerful data visualization features to create beautiful and informative graphical presentations of your data Get statistical ? find out how to do simple statistical analysis, summarize your variables, and conduct classic statistical tests, such as t-tests Expand and customize R ? get the lowdown on how to find, install, and make the most of add-

on packages created by the global R community for a wide variety of purposes Open the book and find: Help downloading, installing, and configuring R Tips for getting data in and out of R Ways to use data frames and lists to organize data How to manipulate and process data Advice on fitting regression models and ANOVA Helpful hints for working with graphics How to code in R What R mailing lists and forums can do for you

This IBM® Redbooks® publication presents a development approach for master data management projects, and in particular, those projects based on IBM InfoSphere® MDM Server. The target audience for this book includes Enterprise Architects, Information, Integration and Solution Architects and Designers, Developers, and Product Managers. Master data management combines a set of processes and tools that defines and manages the non-transactional data entities of an organization. Master data management can provide processes for collecting, consolidating, persisting, and distributing this data throughout an organization. IBM InfoSphere Master Data Management Server creates trusted views of master data that can improve applications and business processes. You can use it to gain control over business information by managing and maintaining a complete and accurate view of master data. You also can use InfoSphere MDM Server to extract maximum value from master data by centralizing multiple data domains. InfoSphere MDM Server provides a comprehensive set of prebuilt business services that support a full range of master data management functionality. As data management and integration continue to evolve rapidly, storing all your data in one place, such as a data warehouse, is no longer scalable. In the very near future, data will need to be distributed and available for several technological solutions. With this practical book, you'll learn how to migrate your enterprise from a complex and tightly coupled data landscape to a more flexible architecture ready for the modern world of data consumption. Executives, data architects, analytics teams, and compliance and governance staff will learn how to build a modern scalable data landscape using the Scaled Architecture, which you can introduce incrementally without a large upfront investment. Author Piethein Strengtholt provides blueprints, principles, observations, best practices, and patterns to get you up to speed. Examine data management trends, including technological developments, regulatory requirements, and privacy concerns Go deep into the Scaled Architecture and learn how the pieces fit together Explore data governance and data security, master data management, self-service data marketplaces, and the importance of metadata

Power Query is the amazing new data cleansing tool in both Excel and Power BI Desktop. Do you find yourself performing the same data cleansing steps day after day? Power Query will make it faster to clean your data the first time. While Power Query is powerful, the interface is subtle--there are tools hiding in plain sight that are easy to miss. Go beyond the obvious and take Power Query to new levels with this book.

In this book, authors Dalton Cervo and Mark Allen show you how to implement Master Data Management (MDM) within your business model to create a more quality controlled approach. Focusing on techniques that can improve data quality management, lower data maintenance costs, reduce corporate and compliance risks, and drive increased efficiency in customer data management practices, the book will guide you in successfully managing and maintaining your customer master data. You'll find the expert guidance you need, complete with tables, graphs, and charts, in planning, implementing, and managing MDM.

The latest techniques for building a customer-focused enterprise environment "The authors have appreciated that MDM is a complex multidimensional area, and have set



out to cover each of these dimensions in sufficient detail to provide adequate practical guidance to anyone implementing MDM. While this necessarily makes the book rather long, it means that the authors achieve a comprehensive treatment of MDM that is lacking in previous works." -- Malcolm Chisholm, Ph.D., President, AskGet.com Consulting, Inc. Regain control of your master data and maintain a master-entity-centric enterprise data framework using the detailed information in this authoritative guide. Master Data Management and Data Governance, Second Edition provides up-to-date coverage of the most current architecture and technology views and system development and management methods. Discover how to construct an MDM business case and roadmap, build accurate models, deploy data hubs, and implement layered security policies. Legacy system integration, cross-industry challenges, and regulatory compliance are also covered in this comprehensive volume. Plan and implement enterprise-scale MDM and Data Governance solutions

- Develop master data model
- Identify, match, and link master records for various domains through entity resolution
- Improve efficiency and maximize integration using SOA and Web services
- Ensure compliance with local, state, federal, and international regulations
- Handle security using authentication, authorization, roles, entitlements, and encryption
- Defend against identity theft, data compromise, spyware attack, and worm infection
- Synchronize components and test data quality and system performance

Power Query is the amazing new data cleansing tool in both Excel and Power BI Desktop. Do you find yourself performing the same data cleansing steps day after day? Power Query will make it faster to clean your data the first time. While Power Query is powerful, the interface is subtle—there are tools hiding in plain sight that are easy to miss. Go beyond the obvious and take Power Query to new levels with this book. Generate effective results in a variety of visually appealing charts using the plotting packages in Python

About This Book Explore various tools and their strengths while building meaningful representations that can make it easier to understand data

Packed with computational methods and algorithms in diverse fields of science

Written in an easy-to-follow categorical style, this book discusses some niche techniques that will make your code easier to work with and reuse

Who This Book Is For If you are a Python developer who performs data visualization and wants to develop existing knowledge about Python to build analytical results and produce some amazing visual display, then this book is for you. A basic knowledge level and understanding of Python libraries is assumed.

What You Will Learn

- Gather, cleanse, access, and map data to a visual framework
- Recognize which visualization method is applicable and learn best practices for data visualization
- Get acquainted with reader-driven narratives and author-driven narratives and the principles of perception
- Understand why Python is an effective tool to be used for numerical computation much like MATLAB, and explore some interesting data structures that come with it
- Explore with various visualization choices how Python can be very useful in computation in the field of finance and statistics
- Get to know why Python is the second choice after Java, and is used frequently in the field of machine learning
- Compare Python with other visualization approaches using Julia and a JavaScript-based framework such as D3.js
- Discover how Python can be used in conjunction with NoSQL such as Hive to produce results efficiently in a distributed environment

In Detail Python has a handful of open source libraries for numerical computations involving optimization, linear algebra, integration,

interpolation, and other special functions using array objects, machine learning, data mining, and plotting. Pandas have a productive environment for data analysis. These libraries have a specific purpose and play an important role in the research into diverse domains including economics, finance, biological sciences, social science, health care, and many more. The variety of tools and approaches available within Python community is stunning, and can bolster and enhance visual story experiences. This book offers practical guidance to help you on the journey to effective data visualization. Commencing with a chapter on the data framework, which explains the transformation of data into information and eventually knowledge, this book subsequently covers the complete visualization process using the most popular Python libraries with working examples. You will learn the usage of Numpy, Scipy, IPython, Matplotlib, Pandas, Patsy, and Scikit-Learn with a focus on generating results that can be visualized in many different ways. Further chapters are aimed at not only showing advanced techniques such as interactive plotting; numerical, graphical linear, and non-linear regression; clustering and classification, but also in helping you understand the aesthetics and best practices of data visualization. The book concludes with interesting examples such as social networks, directed graph examples in real-life, data structures appropriate for these problems, and network analysis. By the end of this book, you will be able to effectively solve a broad set of data analysis problems. Style and approach The approach of this book is not step by step, but rather categorical. The categories are based on fields such as bioinformatics, statistical and machine learning, financial computation, and linear algebra. This approach is beneficial for the community in many different fields of work and also helps you learn how one approach can make sense across many fields

A step-by-step introduction to using SAS® statistical software as a foundational approach to data analysis and interpretation Presenting a straightforward introduction from the ground up, SAS® Essentials: Mastering SAS for Data Analytics, Second Edition illustrates SAS using hands-on learning techniques and numerous real-world examples. Keeping different experience levels in mind, the highly-qualified author team has developed the book over 20 years of teaching introductory SAS courses. Divided into two sections, the first part of the book provides an introduction to data manipulation, statistical techniques, and the SAS programming language. The second section is designed to introduce users to statistical analysis using SAS Procedures. Featuring self-contained chapters to enhance the learning process, the Second Edition also includes: Programming approaches for the most up-to-date version of the SAS platform including information on how to use the SAS University Edition Discussions to illustrate the concepts and highlight key fundamental computational skills that are utilized by business, government, and organizations alike New chapters on reporting results in tables and factor analysis Additional information on the DATA step for data management with an emphasis on importing data from other sources, combining data sets, and data cleaning Updated ANOVA and regression examples as well as other data analysis techniques A companion website with the discussed data sets, additional code, and related PowerPoint® slides SAS Essentials: Mastering SAS for Data Analytics, Second Edition is an ideal textbook for upper-undergraduate and graduate-level courses in statistics, data analytics, applied SAS programming, and statistical computer applications as well as an excellent supplement for statistical methodology

courses. The book is an appropriate reference for researchers and academicians who require a basic introduction to SAS for statistical analysis and for preparation for the Basic SAS Certification Exam.

Enterprises today understand the value of employing a master data management (MDM) solution for managing and governing mission critical information assets. Chief data officers and chief information officers drive MDM initiatives with IBM® InfoSphere® Master Data Management to improve business results and operational efficiencies, which can help to lower costs and to reduce the risk of using untrusted master information in business process. Cloud computing introduces new considerations where enterprise IT architectures are extended beyond the corporate networks into the cloud. Many enterprises are now adopting turnkey business applications offered as software as a service (SaaS) solutions, such as customer relationship management (CRM), payroll processing, human resource management, and many more. However, in the context of MDM solutions, many organizations perceive risks in having these solutions deployed on the cloud. In some cases, organizations are concerned with the legal restrictions of deploying solutions on the cloud, whereas in other cases organizations have policies and strategies in force that limit solution deployment on the cloud. Irrespective of what all the cases might be, industry trends point to a prediction that many "extended enterprises" will keep MDM solutions on premises and will want its integrations with SaaS applications, specifically customer and asset domains. This trend puts a key focus on an important component in the solution construct, that is, the cloud integration middleware and how it fits with hybrid cloud architectures that span on premises and cloud services. As this trend pans out, the on-premises MDM solution integration with SaaS applications will be the key pain point for the "extended enterprise." This IBM Redbooks® publication provides guidance to chief data officers, chief information officers, MDM practitioners, integration architects, and others who are interested in the integration of IBM InfoSphere Master Data Management with SaaS applications. This book lays the background on how mastering and governance needs for SaaS applications is quite similar to what on-premises business applications would need. It draws the perspective for serving the on-premises application and the SaaS application with the same MDM hub. This book describes how IBM WebSphere® Cast Iron® Cloud Integration can serve as the "de-facto" cloud integration middleware to integrate the on-premises InfoSphere Master Data Management systems with any SaaS application by using Salesforce.com integration as an example. This book also covers aspects of handling bulk operations with IBM InfoSphere Information Server. After reading this book, you will have a good understanding about the considerations for on-premises InfoSphere Master Data Management integration with SaaS applications in general and Salesforce.com in particular. The MDM practitioners and integration architects will understand the deployable integrations patterns and, in general, will be able to effectively contribute to delivering strategies that involve building solutions in this area. Additionally, SaaS vendors and customers looking to build or implement SaaS solutions that might require trusted master information will be able to use this compilation to ensure that the right architecture is put together and adhered to as a set of standard integrations patterns with all the core building blocks is essential for the longevity of a solution in this space.

A major limitation of conventional web sites is their unorganized and isolated contents,

which is created mainly for human consumption. This limitation can be addressed by organizing and publishing data, using powerful formats that add structure and meaning to the content of web pages and link related data to one another. Computers can "understand" such data better, which can be useful for task automation. The web sites that provide semantics (meaning) to software agents form the Semantic Web, the Artificial Intelligence extension of the World Wide Web. In contrast to the conventional Web (the "Web of Documents"), the Semantic Web includes the "Web of Data", which connects "things" (representing real-world humans and objects) rather than documents meaningless to computers. Mastering Structured Data on the Semantic Web explains the practical aspects and the theory behind the Semantic Web and how structured data, such as HTML5 Microdata and JSON-LD, can be used to improve your site's performance on next-generation Search Engine Result Pages and be displayed on Google Knowledge Panels. You will learn how to represent arbitrary fields of human knowledge in a machine-interpretable form using the Resource Description Framework (RDF), the cornerstone of the Semantic Web. You will see how to store and manipulate RDF data in purpose-built graph databases such as triplestores and quadstores, that are exploited in Internet marketing, social media, and data mining, in the form of Big Data applications such as the Google Knowledge Graph, Wikidata, or Facebook's Social Graph. With the constantly increasing user expectations in web services and applications, Semantic Web standards gain more popularity. This book will familiarize you with the leading controlled vocabularies and ontologies and explain how to represent your own concepts. After learning the principles of Linked Data, the five-star deployment scheme, and the Open Data concept, you will be able to create and interlink five-star Linked Open Data, and merge your RDF graphs to the LOD Cloud. The book also covers the most important tools for generating, storing, extracting, and visualizing RDF data, including, but not limited to, Protégé, TopBraid Composer, Sindice, Apache Marmotta, Callimachus, and Tabulator. You will learn to implement Apache Jena and Sesame in popular IDEs such as Eclipse and NetBeans, and use these APIs for rapid Semantic Web application development. Mastering Structured Data on the Semantic Web demonstrates how to represent and connect structured data to reach a wider audience, encourage data reuse, and provide content that can be automatically processed with full certainty. As a result, your web contents will be integral parts of the next revolution of the Web.

A thought-provoking and wide-ranging exploration of machine learning and the race to build computer intelligences as flexible as our own In the world's top research labs and universities, the race is on to invent the ultimate learning algorithm: one capable of discovering any knowledge from data, and doing anything we want, before we even ask. In *The Master Algorithm*, Pedro Domingos lifts the veil to give us a peek inside the learning machines that power Google, Amazon, and your smartphone. He assembles a blueprint for the future universal learner--the Master Algorithm--and discusses what it will mean for business, science, and society. If data-ism is today's philosophy, this book is its bible.

Explore the world of data science through Python and learn how to make sense of data About This Book Master data science methods using Python and its libraries Create data visualizations and mine for patterns Advanced techniques for the four fundamentals of Data Science with Python - data mining, data analysis, data



visualization, and machine learning

**Who This Book Is For** If you are a Python developer who wants to master the world of data science then this book is for you. Some knowledge of data science is assumed.

**What You Will Learn** Manage data and perform linear algebra in Python Derive inferences from the analysis by performing inferential statistics Solve data science problems in Python Create high-end visualizations using Python Evaluate and apply the linear regression technique to estimate the relationships among variables. Build recommendation engines with the various collaborative filtering algorithms Apply the ensemble methods to improve your predictions Work with big data technologies to handle data at scale

**In Detail** Data science is a relatively new knowledge domain which is used by various organizations to make data driven decisions. Data scientists have to wear various hats to work with data and to derive value from it. The Python programming language, beyond having conquered the scientific community in the last decade, is now an indispensable tool for the data science practitioner and a must-know tool for every aspiring data scientist. Using Python will offer you a fast, reliable, cross-platform, and mature environment for data analysis, machine learning, and algorithmic problem solving. This comprehensive guide helps you move beyond the hype and transcend the theory by providing you with a hands-on, advanced study of data science. Beginning with the essentials of Python in data science, you will learn to manage data and perform linear algebra in Python. You will move on to deriving inferences from the analysis by performing inferential statistics, and mining data to reveal hidden patterns and trends. You will use the matplotlib library to create high-end visualizations in Python and uncover the fundamentals of machine learning. Next, you will apply the linear regression technique and also learn to apply the logistic regression technique to your applications, before creating recommendation engines with various collaborative filtering algorithms and improving your predictions by applying the ensemble methods. Finally, you will perform K-means clustering, along with an analysis of unstructured data with different text mining techniques and leveraging the power of Python in big data analytics.

**Style and approach** This book is an easy-to-follow, comprehensive guide on data science using Python. The topics covered in the book can all be used in real world scenarios.

Data science is the most exciting skill you can master. Data has dramatically changed how our world works. From entertainment to politics, from technology to advertising and from science to the business world, data is integral and its only limit is our imagination. If you want to have a vibrant and valuable professional life, being skilled with data is the key to a cutting-edge career. Learning how to work with data may seem intimidating or difficult but with *Confident Data Skills* you will be able to master the fundamentals and supercharge your professional abilities. This essential book covers data mining, preparing data, analysing data, communicating data, financial modelling, visualizing insights and presenting data through film making and dynamic simulations. In-depth international case studies from a wide range of organizations, including Netflix, LinkedIn, Goodreads, Deep Blue, Alpha Go and Mike's Hard Lemonade Co. show successful data techniques in practice and inspire you to turn knowledge into innovation. *Confident Data Skills* also provides insightful guidance on how you can use data skills to enhance your employability and improve how your industry or company works through your data skills. Expert author and instructor, Kirill Eremenko, is committed to making the complex simple and inspiring you to have the confidence to

develop an understanding, adeptness and love of data.

Leverage Azure security services to architect robust cloud solutions in Microsoft Azure  
Key Features Secure your Azure cloud workloads across applications and networks  
Protect your Azure infrastructure from cyber attacks Discover tips and techniques for implementing, deploying, and maintaining secure cloud services using best practices  
Book Description Security is always integrated into cloud platforms, causing users to let their guard down as they take cloud security for granted. Cloud computing brings new security challenges, but you can overcome these with Microsoft Azure's shared responsibility model. Mastering Azure Security covers the latest security features provided by Microsoft to identify different threats and protect your Azure cloud using innovative techniques. The book takes you through the built-in security controls and the multi-layered security features offered by Azure to protect cloud workloads across apps and networks. You'll get to grips with using Azure Security Center for unified security management, building secure application gateways on Azure, protecting the cloud from DDoS attacks, safeguarding with Azure Key Vault, and much more. Additionally, the book covers Azure Sentinel, monitoring and auditing, Azure security and governance best practices, and securing PaaS deployments. By the end of this book, you'll have developed a solid understanding of cybersecurity in the cloud and be able to design secure solutions in Microsoft Azure. What you will learn Understand cloud security concepts Get to grips with managing cloud identities Adopt the Azure security cloud infrastructure Grasp Azure network security concepts Discover how to keep cloud resources secure Implement cloud governance with security policies and rules Who this book is for This book is for Azure cloud professionals, Azure architects, and security professionals looking to implement secure cloud services using Azure Security Centre and other Azure security features. A fundamental understanding of security concepts and prior exposure to the Azure cloud will help you understand the key concepts covered in the book more effectively.

This is my latest book on Data Architecture focusing on the subject of MDM (Master Data Management). It is intended to provide a overview of the subject with chapters covering key topics such as: the business case, data privacy, the challenges of global MDM, golden source and authoritative source explanations, the different MDM styles and the record matching process. The back cover text says the following: " Master Data Management (MDM for short) has become a whole industry, within an industry. There are many companies now claiming to be MDM software (or services) providers. Everyone wants a master data project on their CV, and in general it has become hip and trendy to talk about and do. The reality is that MDM is in fact the reincarnation of the problem of how to manage the consistency and integrity of the myriads of data assets that exist across the enterprise. This book provides an understanding of MDM, the business drivers behind it, the various techniques that are critical to its success and gives a good architectural grounding in the subject. It is perfect for anyone embarking on an 'adventure' in this problem space." I hope you find this book enjoyable and useful. Andy

The concept of a data lake is less than 10 years old, but they are already hugely implemented within large companies. Their goal is to efficiently deal with ever-growing volumes of heterogeneous data, while also facing various sophisticated user needs. However, defining and building a data lake is still a challenge, as no consensus has

been reached so far. Data Lakes presents recent outcomes and trends in the field of data repositories. The main topics discussed are the data-driven architecture of a data lake; the management of metadata – supplying key information about the stored data, master data and reference data; the roles of linked data and fog computing in a data lake ecosystem; and how gravity principles apply in the context of data lakes. A variety of case studies are also presented, thus providing the reader with practical examples of data lake management.

"Analyzing data is not easy, due to the fact that you have to figure out which type of data analytics you are going to use, as well as defeat the challenges that you will come up against when it comes to analyzing data. With this book, it is our goal to show you the easiest way to work with data analytics and how you are going to avoid some of the challenges and risks that you will be putting yourself up against when you are working with data. You will realize that analyzing data is not the easiest thing in the world.

However, it is going to get easier the more that you practice. Just guarantee that you are taking the time to practice and do not put too much pressure on yourself. In this book, you are going to learn: The risks of data analytics The types of data analytics that are out there in the world What the decision tree is The benefits of using data analytics Real world examples that will show you how you are going to be able to take this knowledge and apply it to your everyday life. Data analysis happens no matter what line of work you are in, and it is my hope that with this book, you are able to learn everything that pushes you further in your knowledge of data analysis!"--

Regular expressions are an extremely powerful tool for manipulating text and data. They are now standard features in a wide range of languages and popular tools, including Perl, Python, Ruby, Java, VB.NET and C# (and any language using the .NET Framework), PHP, and MySQL. If you don't use regular expressions yet, you will discover in this book a whole new world of mastery over your data. If you already use them, you'll appreciate this book's unprecedented detail and breadth of coverage. If you think you know all you need to know about regular expressions, this book is a stunning eye-opener. As this book shows, a command of regular expressions is an invaluable skill. Regular expressions allow you to code complex and subtle text processing that you never imagined could be automated. Regular expressions can save you time and aggravation. They can be used to craft elegant solutions to a wide range of problems. Once you've mastered regular expressions, they'll become an invaluable part of your toolkit. You will wonder how you ever got by without them. Yet despite their wide availability, flexibility, and unparalleled power, regular expressions are frequently underutilized. Yet what is power in the hands of an expert can be fraught with peril for the unwary. Mastering Regular Expressions will help you navigate the minefield to becoming an expert and help you optimize your use of regular expressions. Mastering Regular Expressions, Third Edition, now includes a full chapter devoted to PHP and its powerful and expressive suite of regular expression functions, in addition to enhanced PHP coverage in the central "core" chapters. Furthermore, this edition has been updated throughout to reflect advances in other languages, including expanded in-depth coverage of Sun's `java.util.regex` package, which has emerged as the standard Java regex implementation. Topics include: A comparison of features among different versions of many languages and tools How the regular expression engine works Optimization (major savings available here!) Matching just what you want, but not what you don't want Sections and chapters on individual languages Written in the lucid, entertaining tone that makes a complex, dry topic become crystal-clear to programmers, and sprinkled with solutions to complex real-world problems, Mastering Regular Expressions, Third Edition offers a wealth of information that you can put to immediate use. Reviews of this new edition and the

second edition: "There isn't a better (or more useful) book available on regular expressions."  
--Zak Greant, Managing Director, eZ Systems "A real tour-de-force of a book which not only covers the mechanics of regexes in extraordinary detail but also talks about efficiency and the use of regexes in Perl, Java, and .NET...If you use regular expressions as part of your professional work (even if you already have a good book on whatever language you're programming in) I would strongly recommend this book to you." --Dr. Chris Brown, Linux Format "The author does an outstanding job leading the reader from regexnovice to master. The book is extremely easy to read and chock full of useful and relevant examples...Regular expressions are valuable tools that every developer should have in their toolbox. Mastering Regular Expressions is the definitive guide to the subject, and an outstanding resource that belongs on every programmer's bookshelf. Ten out of Ten Horseshoes." --Jason Menard, Java Ranch

If you want to learn more about Data Science or how to master it with the Python Programming Language, then keep reading. Data Science is one of the biggest buzzwords in the business world nowadays. Many businesses know the importance of collecting information, but as they can collect so much data in a short period, the real question is: "what is the next step?" Data Science includes all the different steps that you take with the data: collecting and cleaning them if they come from more than one source, analyzing them, applying Machine Learning algorithms and models, and then presenting your findings from the analysis with some good Data Visualizations. And this is what you will learn in Python Data Science. You will learn about the main steps that are needed to correctly implement Data Science techniques and the algorithms to help you sort through the data and see some amazing results. Some of the topics that we will discuss inside include: What data science is all about and why so many companies are using it to give them a competitive edge. Why Python and how to use it to implement Data Science What is the intersection between Machine Learning and Data Science and how to combine them The main Data Structures & Object-Oriented Python, with practical codes and exercises to use Python Functions and Modules in Python The 7 most important algorithms and models in Data Science Data Aggregation and Group Operations 9 important Data Mining techniques in Data Science Interaction with databases and data in the cloud And Much More! Where most books only focus on how collecting and cleaning the data, this book goes further, providing guidance on how to perform a proper analysis in order to extract precious information that may be vital for a business. Don't miss the opportunity to learn more about these topics. Even if you have never implemented Data Science techniques, learning them is easier than it looks. You just need the right guidance. And Python Data Science provides all the knowledge you need in a simple and practical way. Regardless of your previous experience, you will learn, the techniques to manipulate and process datasets, the principles of Python programming, and its most important real-world applications. Would You Like To Know More? Scroll up and click on the BUY NOW button to get your copy now!

Power Query is one component of the Power BI (Business Intelligence) product from Microsoft, and "M" is the name of the programming language created by it. As more business intelligence pros begin using Power Pivot, they find that they do not have the Excel skills to clean the data in Excel; Power Query solves this problem. This book shows how to use the Power Query tool to get difficult data sets into both Excel and Power Pivot, and is solely devoted to Power Query dashboarding and reporting.

Best Practices for Deploying and Managing Master Data Services (MDS) Effectively manage master data and drive better decision making across your enterprise with detailed instruction from two MDS experts. Microsoft SQL Server 2008 R2 Master Data Services Implementation & Administration shows you how to use MDS to centralize the management of key data within your organization. Find out how to build an MDS model, establish hierarchies, govern data access, and enforce business rules. Legacy system integration and security are also covered.



Real-world programming examples illustrate the material presented in this comprehensive guide. Create a process-agnostic solution for managing your business domains Learn how to take advantage of the data modeling capabilities of MDS Manage hierarchies and consolidations across your organization Import data by using SQL Server Integration Services and T-SQL statements Ensure data accuracy and completeness by using business rules and versioning Employ role-based security at functional, object, and attribute levels Design export views and publish data to subscribing systems Use Web services to programmatically interact with your implementation

Data warehousing is one of the hottest business topics, and there's more to understanding data warehousing technologies than you might think. Find out the basics of data warehousing and how it facilitates data mining and business intelligence with *Data Warehousing For Dummies, 2nd Edition*. Data is probably your company's most important asset, so your data warehouse should serve your needs. The fully updated Second Edition of *Data Warehousing For Dummies* helps you understand, develop, implement, and use data warehouses, and offers a sneak peek into their future. You'll learn to: Analyze top-down and bottom-up data warehouse designs Understand the structure and technologies of data warehouses, operational data stores, and data marts Choose your project team and apply best development practices to your data warehousing projects Implement a data warehouse, step by step, and involve end-users in the process Review and upgrade existing data storage to make it serve your needs Comprehend OLAP, column-wise databases, hardware assisted databases, and middleware Use data mining intelligently and find what you need Make informed choices about consultants and data warehousing products *Data Warehousing For Dummies, 2nd Edition* also shows you how to involve users in the testing process and gain valuable feedback, what it takes to successfully manage a data warehouse project, and how to tell if your project is on track. You'll find it's the most useful source of data on the topic!

SAS software has been in existence for a long time and has been implemented in large, data-intensive environments, including data warehouses. This SAS book covers practical programming considerations to make when involving SAS in a data warehouse environment. You'll be able to develop the skills you need to apply SAS in your working environment.

Learn how to create more powerful data mining applications with this comprehensive Python guide to advance data analytics techniques About This Book Dive deeper into data mining with Python – don't be complacent, sharpen your skills! From the most common elements of data mining to cutting-edge techniques, we've got you covered for any data-related challenge Become a more fluent and confident Python data-analyst, in full control of its extensive range of libraries Who This Book Is For This book is for data scientists who are already familiar with some basic data mining techniques such as SQL and machine learning, and who are comfortable with Python. If you are ready to learn some more advanced techniques in data mining in order to become a data mining expert, this is the book for you! What You Will Learn Explore techniques for finding frequent itemsets and association rules in large data sets Learn identification methods for entity matches across many different types of data Identify the basics of network mining and how to apply it to real-world data sets Discover methods for detecting the sentiment of text and for locating named entities in text Observe multiple techniques for automatically extracting summaries and generating topic models for text See how to use data mining to fix data anomalies and how to use machine learning to identify outliers in a data set In Detail Data mining is an integral part of the data science pipeline. It is the foundation of any successful data-driven strategy – without it, you'll never be able to uncover truly transformative insights. Since data is vital to just about every modern organization, it is worth taking the next step to unlock even greater value and more meaningful understanding. If you already know the fundamentals of data mining with Python, you are now ready to experiment with more interesting, advanced data analytics techniques using Python's easy-to-use interface and

extensive range of libraries. In this book, you'll go deeper into many often overlooked areas of data mining, including association rule mining, entity matching, network mining, sentiment analysis, named entity recognition, text summarization, topic modeling, and anomaly detection. For each data mining technique, we'll review the state-of-the-art and current best practices before comparing a wide variety of strategies for solving each problem. We will then implement example solutions using real-world data from the domain of software engineering, and we will spend time learning how to understand and interpret the results we get. By the end of this book, you will have solid experience implementing some of the most interesting and relevant data mining techniques available today, and you will have achieved a greater fluency in the important field of Python data analytics. **Style and approach** This book will teach you the intricacies in applying data mining using real-world scenarios and will act as a very practical solution to your data mining needs.

The key to a successful MDM initiative isn't technology or methods, it's people: the stakeholders in the organization and their complex ownership of the data that the initiative will affect. Master Data Management equips you with a deeply practical, business-focused way of thinking about MDM—an understanding that will greatly enhance your ability to communicate with stakeholders and win their support. Moreover, it will help you deserve their support: you'll master all the details involved in planning and executing an MDM project that leads to measurable improvements in business productivity and effectiveness. \* Presents a comprehensive roadmap that you can adapt to any MDM project. \* Emphasizes the critical goal of maintaining and improving data quality. \* Provides guidelines for determining which data to “master. \* Examines special issues relating to master data metadata. \* Considers a range of MDM architectural styles. \* Covers the synchronization of master data across the application infrastructure.

Helps users understand the breadth of Azure services by organizing them into a reference framework they can use when crafting their own big-data analytics solution.

Mastering Your DataKoios Associates Ltd

This book is a refreshingly practical yet theoretically sound roadmap to leveraging data analytics and data science. The vast amount of data generated about us and our world is useless without plans and strategies that are designed to cope with its size and complexity, and which enable organizations to leverage the information to create value in marketing. **Creating Value with Data Analytics in Marketing** provides a nuanced view of big data developments and data science, arguing that big data is not a revolution but an evolution of the increasing availability of data that has been observed in recent times. Building on the authors' extensive academic and practical knowledge, this book aims to provide managers and analysts with strategic directions and practical analytical solutions on how to create value from existing and new big data. The second edition of this bestselling text has been fully updated in line with developments in the field and includes a selection of new, international cases and examples, exercises, techniques and methodologies. Tying data and analytics to specific goals and processes for implementation makes this essential reading for advanced undergraduate and postgraduate students and specialists of data analytics, marketing research, marketing management and customer relationship management. Online resources include chapter-by-chapter lecture slides and data sets and corresponding R code for selected chapters.

Use Java to create a diverse range of Data Science applications and bring Data Science into production **About This Book** An overview of modern Data Science and Machine Learning libraries available in Java **Coverage** of a broad set of topics, going from the basics of Machine Learning to Deep Learning and Big Data frameworks. Easy-to-follow illustrations and the running example of building a search engine. **Who This Book Is For** This book is intended for software engineers who are comfortable with developing Java applications and are familiar with the basic concepts of data science. Additionally, it will also be useful for data scientists

who do not yet know Java but want or need to learn it. If you are willing to build efficient data science applications and bring them in the enterprise environment without changing the existing stack, this book is for you! What You Will Learn Get a solid understanding of the data processing toolbox available in Java Explore the data science ecosystem available in Java Find out how to approach different machine learning problems with Java Process unstructured information such as natural language text or images Create your own search engine Get state-of-the-art performance with XGBoost Learn how to build deep neural networks with DeepLearning4j Build applications that scale and process large amounts of data Deploy data science models to production and evaluate their performance In Detail Java is the most popular programming language, according to the TIOBE index, and it is a typical choice for running production systems in many companies, both in the startup world and among large enterprises. Not surprisingly, it is also a common choice for creating data science applications: it is fast and has a great set of data processing tools, both built-in and external. What is more, choosing Java for data science allows you to easily integrate solutions with existing software, and bring data science into production with less effort. This book will teach you how to create data science applications with Java. First, we will revise the most important things when starting a data science application, and then brush up the basics of Java and machine learning before diving into more advanced topics. We start by going over the existing libraries for data processing and libraries with machine learning algorithms. After that, we cover topics such as classification and regression, dimensionality reduction and clustering, information retrieval and natural language processing, and deep learning and big data. Finally, we finish the book by talking about the ways to deploy the model and evaluate it in production settings. Style and approach This is a practical guide where all the important concepts such as classification, regression, and dimensionality reduction are explained with the help of examples.

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience"--

Become an expert at using Python for advanced statistical analysis of data using real-world examples About This Book Clean, format, and explore data using graphical and numerical summaries Leverage the IPython environment to efficiently analyze data with Python Packed with easy-to-follow examples to develop advanced computational skills for the analysis of complex data Who This Book Is For If you are a competent Python developer who wants to take your data analysis skills to the next level by solving complex problems, then this advanced guide is for you. Familiarity with the basics of applying Python libraries to data sets is assumed. What You Will Learn Read, sort, and map various data into Python and Pandas Recognise patterns so you can understand and explore data Use statistical models to discover patterns in data Review classical statistical inference using Python, Pandas, and SciPy Detect similarities and differences in data with clustering Clean your data to make it useful Work in Jupyter Notebook to produce publication ready figures to be included in reports In Detail Python, a multi-paradigm programming language, has become the language of choice for data scientists for data analysis, visualization, and machine learning. Ever imagined how to become an expert at effectively approaching data analysis problems, solving them, and extracting all of the available information from your data? Well, look no further, this is the book you want! Through this comprehensive guide, you will explore data and present results and conclusions from statistical analysis in a meaningful way. You'll be able to quickly and accurately perform the hands-on sorting, reduction, and subsequent analysis, and fully appreciate how data analysis methods can support business decision-making. You'll start off by learning about the tools available for data analysis in Python and will then explore the statistical models that are used to identify patterns in data. Gradually, you'll move on to review statistical inference using Python, Pandas, and SciPy. After that, we'll focus on performing regression using

computational tools and you'll get to understand the problem of identifying clusters in data in an algorithmic way. Finally, we delve into advanced techniques to quantify cause and effect using Bayesian methods and you'll discover how to use Python's tools for supervised machine learning. Style and approach This book takes a step-by-step approach to reading, processing, and analyzing data in Python using various methods and tools. Rich in examples, each topic connects to real-world examples and retrieves data directly online where possible. With this book, you are given the knowledge and tools to explore any data on your own, encouraging a curiosity befitting all data scientists.

The Only Complete Technical Primer for MDM Planners, Architects, and Implementers Companies moving toward flexible SOA architectures often face difficult information management and integration challenges. The master data they rely on is often stored and managed in ways that are redundant, inconsistent, inaccessible, non-standardized, and poorly governed. Using Master Data Management (MDM), organizations can regain control of their master data, improve corresponding business processes, and maximize its value in SOA environments. Enterprise Master Data Management provides an authoritative, vendor-independent MDM technical reference for practitioners: architects, technical analysts, consultants, solution designers, and senior IT decisionmakers. Written by the IBM® data management innovators who are pioneering MDM, this book systematically introduces MDM's key concepts and technical themes, explains its business case, and illuminates how it interrelates with and enables SOA. Drawing on their experience with cutting-edge projects, the authors introduce MDM patterns, blueprints, solutions, and best practices published nowhere else—everything you need to establish a consistent, manageable set of master data, and use it for competitive advantage. Coverage includes How MDM and SOA complement each other Using the MDM Reference Architecture to position and design MDM solutions within an enterprise Assessing the value and risks to master data and applying the right security controls Using PIM-MDM and CDI-MDM Solution Blueprints to address industry-specific information management challenges Explaining MDM patterns as enablers to accelerate consistent MDM deployments Incorporating MDM solutions into existing IT landscapes via MDM Integration Blueprints Leveraging master data as an enterprise asset—bringing people, processes, and technology together with MDM and data governance Best practices in MDM deployment, including data warehouse and SAP integration

Master the techniques and sophisticated analytics used to construct Spark-based solutions that scale to deliver production-grade data science products About This Book Develop and apply advanced analytical techniques with Spark Learn how to tell a compelling story with data science using Spark's ecosystem Explore data at scale and work with cutting edge data science methods Who This Book Is For This book is for those who have beginner-level familiarity with the Spark architecture and data science applications, especially those who are looking for a challenge and want to learn cutting edge techniques. This book assumes working knowledge of data science, common machine learning methods, and popular data science tools, and assumes you have previously run proof of concept studies and built prototypes. What You Will Learn Learn the design patterns that integrate Spark into industrialized data science pipelines See how commercial data scientists design scalable code and reusable code for data science services Explore cutting edge data science methods so that you can study trends and causality Discover advanced programming techniques using RDD and the DataFrame and Dataset APIs Find out how Spark can be used as a universal ingestion engine tool and as a web scraper Practice the implementation of advanced topics in graph processing, such as community detection and contact chaining Get to know the best practices when performing Extended Exploratory Data Analysis, commonly used in commercial data science teams Study advanced Spark concepts, solution design patterns, and integration architectures Demonstrate powerful data science pipelines In Detail Data science seeks to transform the



world using data, and this is typically achieved through disrupting and changing real processes in real industries. In order to operate at this level you need to build data science solutions of substance –solutions that solve real problems. Spark has emerged as the big data platform of choice for data scientists due to its speed, scalability, and easy-to-use APIs. This book deep dives into using Spark to deliver production-grade data science solutions. This process is demonstrated by exploring the construction of a sophisticated global news analysis service that uses Spark to generate continuous geopolitical and current affairs insights. You will learn all about the core Spark APIs and take a comprehensive tour of advanced libraries, including Spark SQL, Spark Streaming, MLlib, and more. You will be introduced to advanced techniques and methods that will help you to construct commercial-grade data products. Focusing on a sequence of tutorials that deliver a working news intelligence service, you will learn about advanced Spark architectures, how to work with geographic data in Spark, and how to tune Spark algorithms so they scale linearly. Style and approach This is an advanced guide for those with beginner-level familiarity with the Spark architecture and working with Data Science applications. Mastering Spark for Data Science is a practical tutorial that uses core Spark APIs and takes a deep dive into advanced libraries including: Spark SQL, visual streaming, and MLlib. This book expands on titles like: Machine Learning with Spark and Learning Spark. It is the next learning curve for those comfortable with Spark and looking to improve their skills.

[Copyright: 313eb18622148077610186c002323839](#)