

Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

Focusing on quantitative approaches to investigating problems, this title introduces the basic rules and principles of statistics, encouraging the reader to think critically about data analysis and research design, and how these factors can impact upon evidence-based practice.

Did you ever try getting Businesspeople and IT to agree on the project scope for a new application? Or try getting Marketing and Sales to agree on the target audience? Or try bringing new team members up to speed on the hundreds of tables in your data warehouse — without them dozing off? Whether you are a businessperson or an IT professional, you can be the hero in each of these and hundreds of other scenarios by building a High-Level Data Model. The High-Level Data Model is a simplified view of our complex environment. It can be a powerful communication tool of the key concepts within our application development projects, business intelligence and master data management programs, and all enterprise and industry initiatives. Learn about the High-Level Data Model and master the techniques for building one, including a comprehensive ten-step approach and hands-on exercises to help you practice topics on your own. In this book, we review data modeling basics and explain why the core concepts stored in a high-level data model can have significant business impact on an organization. We explain the technical notation used for a data model and walk through some simple examples of building a high-level data model. We also describe how data models relate to other key initiatives you may have heard of or may be implementing in your organization. This book contains best practices for implementing a high-level data model, along with some easy-to-use templates and guidelines for a step-by-step approach. Each step will be illustrated using many examples based on actual projects we have worked on. Names have been changed to protect the innocent, but the pain points and lessons have been preserved. One example spans an entire chapter and will allow you to practice building a high-level data model from beginning to end, and then compare your results to ours. Building a high-level data model following the ten step approach you'll read about is a great way to ensure you will retain the new skills you learn in this book. As is the case in many disciplines, using the right tool for the right job is critical to the overall success of your high-level data model implementation. To help you in your tool selection process, there are several chapters dedicated to discussing what to look for in a high-level data modeling tool and a framework for choosing a data modeling tool, in general. This book concludes with a real-world case study that shows how an international energy company successfully used a high-level data model to streamline their information

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

management practices and increase communication throughout the organization—between both businesspeople and IT. Data modeling is one of the under-exploited, and potentially very valuable, business capabilities that are often hidden away in an organization's Information Technology department. Data Modeling for the Business highlights both the resulting damage to business value, and the opportunities to make things better. As an easy-to follow and comprehensive guide on the 'why' and 'how' of data modeling, it also reminds us that a successful strategy for exploiting IT depends at least as much on the information as the technology. Chris Potts, Corporate IT Strategist and Author of fruITion: Creating the Ultimate Corporate Strategy for Information Technology One of the most critical systems issues is aligning business with IT and fulfilling business needs using data models. The authors of Data Modeling for the Business do a masterful job at simply and clearly describing the art of using data models to communicate with business representatives and meet business needs. The book provides many valuable tools, analogies, and step-by-step methods for effective data modeling and is an important contribution in bridging the much needed connection between data modeling and realizing business requirements. Len Silverston, author of The Data Model Resource Book series Data is at the center of many challenges in system design today. Difficult issues need to be figured out, such as scalability, consistency, reliability, efficiency, and maintainability. In addition, we have an overwhelming variety of tools, including relational databases, NoSQL datastores, stream or batch processors, and message brokers. What are the right choices for your application? How do you make sense of all these buzzwords? In this practical and comprehensive guide, author Martin Kleppmann helps you navigate this diverse landscape by examining the pros and cons of various technologies for processing and storing data. Software keeps changing, but the fundamental principles remain the same. With this book, software engineers and architects will learn how to apply those ideas in practice, and how to make full use of data in modern applications. Peer under the hood of the systems you already use, and learn how to use and operate them more effectively Make informed decisions by identifying the strengths and weaknesses of different tools Navigate the trade-offs around consistency, scalability, fault tolerance, and complexity Understand the distributed systems research upon which modern databases are built Peek behind the scenes of major online services, and learn from their architectures

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

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

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...

Harness the power and simplicity of Informatica PowerCenter 10.x to build and manage efficient data management solutions About This Book Master PowerCenter 10.x components to create, execute, monitor, and schedule ETL processes with a practical approach. An ideal guide to building the necessary skills and competencies to become an expert Informatica PowerCenter developer. A comprehensive guide to fetching/transforming and loading huge volumes of data in a very effective way, with reduced resource consumption Who This Book Is For If you wish to deploy Informatica in enterprise environments and build a career in data warehousing, then this book is for you. Whether you are a software developer/analytic professional and are new to Informatica or an experienced user, you will learn all the features of Informatica 10.x. A basic knowledge of programming and data warehouse concepts is essential. What You Will Learn Install or upgrade the components of the Informatica PowerCenter tool Work on various aspects of administrative skills and on the various developer Informatica PowerCenter screens such as Designer, Workflow Manager, Workflow Monitor, and Repository Manager. Get practical hands-on experience of various sections of Informatica PowerCenter, such as navigator, toolbar, workspace, control panel, and so on Leverage basic and advanced utilities, such as the debugger, target load plan, and incremental aggregation to process data Implement data warehousing concepts such as schemas and SCDs using Informatica Migrate various components, such as sources and targets, to another region using the

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

Designer and Repository Manager screens Enhance code performance using tips such as pushdown optimization and partitioning In Detail Informatica PowerCenter is an industry-leading ETL tool, known for its accelerated data extraction, transformation, and data management strategies. This book will be your quick guide to exploring Informatica PowerCenter's powerful features such as working on sources, targets, transformations, performance optimization, scheduling, deploying for processing, and managing your data at speed. First, you'll learn how to install and configure tools. You will learn to implement various data warehouse and ETL concepts, and use PowerCenter 10.x components to build mappings, tasks, workflows, and so on. You will come across features such as transformations, SCD, XML processing, partitioning, constraint-based loading, Incremental aggregation, and many more. Moreover, you'll also learn to deliver powerful visualizations for data profiling using the advanced monitoring dashboard functionality offered by the new version. Using data transformation technique, performance tuning, and the many new advanced features, this book will help you understand and process data for training or production purposes. The step-by-step approach and adoption of real-time scenarios will guide you through effectively accessing all core functionalities offered by Informatica PowerCenter version 10.x. Style and approach You'll get hand-on with sources, targets, transformations, performance optimization, scheduling, deploying for processing, and managing your data, and learn everything you need to become a proficient Informatica PowerCenter developer.

Information is currency. Recent studies show that data quality problems are costing businesses billions of dollars each year, with poor data linked to waste and inefficiency, damaged credibility among customers and suppliers, and an organizational inability to make sound decisions. In this important and timely new book, Danette McGilvray presents her "Ten Steps approach to information quality, a proven method for both understanding and creating information quality in the enterprise. Her trademarked approach—in which she has trained Fortune 500 clients and hundreds of workshop attendees—applies to all types of data and to all types of organizations. * Includes numerous templates, detailed examples, and practical advice for executing every step of the "Ten Steps approach. * Allows for quick reference with an easy-to-use format highlighting key concepts and definitions, important checkpoints, communication activities, and best practices. * A companion Web site includes links to numerous data quality resources, including many of the planning and information-gathering templates featured in the text, quick summaries of key ideas from the Ten Step methodology, and other tools and information available online.

A guide for assessing an organization's data quality practice and a roadmap for implementing a viable data and information quality management program, based on rigorous research and drawing on real-world examples. All organizations today confront data quality problems, both systemic and structural. Neither ad hoc approaches nor fixes at

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

the systems level--installing the latest software or developing an expensive data warehouse--solve the basic problem of bad data quality practices. *Journey to Data Quality* offers a roadmap that can be used by practitioners, executives, and students for planning and implementing a viable data and information quality management program. This practical guide, based on rigorous research and informed by real-world examples, describes the challenges of data management and provides the principles, strategies, tools, and techniques necessary to meet them. The authors, all leaders in the data quality field for many years, discuss how to make the economic case for data quality and the importance of getting an organization's leaders on board. They outline different approaches for assessing data, both subjectively (by users) and objectively (using sampling and other techniques). They describe real problems and solutions, including efforts to find the root causes of data quality problems at a healthcare organization and data quality initiatives taken by a large teaching hospital. They address setting company policy on data quality and, finally, they consider future challenges on the journey to data quality.

Human-in-the-Loop Machine Learning lays out methods for humans and machines to work together effectively. Summary Most machine learning systems that are deployed in the world today learn from human feedback. However, most machine learning courses focus almost exclusively on the algorithms, not the human-computer interaction part of the systems. This can leave a big knowledge gap for data scientists working in real-world machine learning, where data scientists spend more time on data management than on building algorithms. *Human-in-the-Loop Machine Learning* is a practical guide to optimizing the entire machine learning process, including techniques for annotation, active learning, transfer learning, and using machine learning to optimize every step of the process. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Machine learning applications perform better with human feedback. Keeping the right people in the loop improves the accuracy of models, reduces errors in data, lowers costs, and helps you ship models faster. About the book *Human-in-the-Loop Machine Learning* lays out methods for humans and machines to work together effectively. You'll find best practices on selecting sample data for human feedback, quality control for human annotations, and designing annotation interfaces. You'll learn to create training data for labeling, object detection, and semantic segmentation, sequence labeling, and more. The book starts with the basics and progresses to advanced techniques like transfer learning and self-supervision within annotation workflows. What's inside Identifying the right training and evaluation data Finding and managing people to annotate data Selecting annotation quality control strategies Designing interfaces to improve accuracy and efficiency About the author Robert (Munro) Monarch is a data scientist and engineer who has built machine learning data for companies such as Apple, Amazon, Google, and IBM. He holds a PhD from Stanford. Robert holds a PhD from Stanford focused on Human-

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

in-the-Loop machine learning for healthcare and disaster response, and is a disaster response professional in addition to being a machine learning professional. A worked example throughout this text is classifying disaster-related messages from real disasters that Robert has helped respond to in the past. Table of Contents PART 1 - FIRST STEPS 1

Introduction to human-in-the-loop machine learning 2 Getting started with human-in-the-loop machine learning PART 2 - ACTIVE LEARNING 3 Uncertainty sampling 4 Diversity sampling 5 Advanced active learning 6 Applying active learning to different machine learning tasks PART 3 - ANNOTATION 7 Working with the people annotating your data 8 Quality control for data annotation 9 Advanced data annotation and augmentation 10 Annotation quality for different machine learning tasks PART 4 - HUMAN-COMPUTER INTERACTION FOR MACHINE LEARNING 11 Interfaces for data annotation 12 Human-in-the-loop machine learning products

The data lake is a daring new approach for harnessing the power of big data technology and providing convenient self-service capabilities. But is it right for your company? This book is based on discussions with practitioners and executives from more than a hundred organizations, ranging from data-driven companies such as Google, LinkedIn, and Facebook, to governments and traditional corporate enterprises. You'll learn what a data lake is, why enterprises need one, and how to build one successfully with the best practices in this book. Alex Gorelik, CTO and founder of Waterline Data, explains why old systems and processes can no longer support data needs in the enterprise. Then, in a collection of essays about data lake implementation, you'll examine data lake initiatives, analytic projects, experiences, and best practices from data experts working in various industries. Get a succinct introduction to data warehousing, big data, and data science Learn various paths enterprises take to build a data lake Explore how to build a self-service model and best practices for providing analysts access to the data Use different methods for architecting your data lake Discover ways to implement a data lake from experts in different industries

Create a competitive advantage with data quality Data is rapidly becoming the powerhouse of industry, but low-quality data can actually put a company at a disadvantage. To be used effectively, data must accurately reflect the real-world scenario it represents, and it must be in a form that is usable and accessible. Quality data involves asking the right questions, targeting the correct parameters, and having an effective internal management, organization, and access system. It must be relevant, complete, and correct, while falling in line with pervasive regulatory oversight programs. Competing with High Quality Data: Concepts, Tools and Techniques for Building a Successful Approach to Data Quality takes a holistic approach to improving data quality, from collection to usage. Author Rajesh Jugulum is globally-recognized as a major voice in the data quality arena, with high-level backgrounds in international corporate finance. In the book, Jugulum provides a roadmap to data quality innovation, covering topics such as: The four-phase approach to data

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

quality control Methodology that produces data sets for different aspects of a business Streamlined data quality assessment and issue resolution A structured, systematic, disciplined approach to effective data gathering The book also contains real-world case studies to illustrate how companies across a broad range of sectors have employed data quality systems, whether or not they succeeded, and what lessons were learned. High-quality data increases value throughout the information supply chain, and the benefits extend to the client, employee, and shareholder. Competing with High Quality Data: Concepts, Tools and Techniques for Building a Successful Approach to Data Quality provides the information and guidance necessary to formulate and activate an effective data quality plan today.

Can any subject inspire less excitement than "data quality"? Yet a moment's thought reveals the ever-growing importance of quality data. From restated corporate earnings, to incorrect prices on the web, to the bombing of the Chinese Embassy, the media reports the impact of poor data quality on a daily basis. Every business operation creates or consumes huge quantities of data. If the data are wrong, time, money, and reputation are lost. In today's environment, every leader, every decision maker, every operational manager, every consumer, indeed everyone has a vested interest in data quality. Data Quality: The Field Guide provides the practical guidance needed to start and advance a data quality program. It motivates interest in data quality, describes the most important data quality problems facing the typical organization, and outlines what an organization must do to improve. It consists of 36 short chapters in an easy-to-use field guide format. Each chapter describes a single issue and how to address it. The book begins with sections that describe why leaders, whether CIOs, CFOs, or CEOs, should be concerned with data quality. It explains the pros and cons of approaches for addressing the issue. It explains what those organizations with the best data do. And it lays bare the social issues that prevent organizations from making headway. "Field tips" at the end of each chapter summarize the most important points. Allows readers to go directly to the topic of interest Provides web-based material so readers can cut and paste figures and tables into documents within their organizations Gives step-by-step instructions for applying most techniques and summarizes what "works"

Providing a practical and step-by-step guide to collecting and managing qualitative data, this book focuses on the three most often used forms of qualitative methods: participant observation, in-depth interviews, and focus groups. Designed to be very applied, this textbook includes many checklists and tips for how to use each technique while doing research. It also includes numerous real-life examples and cases so that the reader will benefit from seeing the broader picture. Collecting Qualitative Data: A Field Manual is intended both for beginning researchers and the more experienced research collector.

Executing Data Quality Projects, Second Edition presents a structured yet flexible approach for creating, improving,

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

sustaining and managing the quality of data and information within any organization. Studies show that data quality problems are costing businesses billions of dollars each year, with poor data linked to waste and inefficiency, damaged credibility among customers and suppliers, and an organizational inability to make sound decisions. Help is here! This book describes a proven Ten Step approach that combines a conceptual framework for understanding information quality with techniques, tools, and instructions for practically putting the approach to work – with the end result of high-quality trusted data and information, so critical to today’s data-dependent organizations. The Ten Steps approach applies to all types of data and all types of organizations – for-profit in any industry, non-profit, government, education, healthcare, science, research, and medicine. This book includes numerous templates, detailed examples, and practical advice for executing every step. At the same time, readers are advised on how to select relevant steps and apply them in different ways to best address the many situations they will face. The layout allows for quick reference with an easy-to-use format highlighting key concepts and definitions, important checkpoints, communication activities, best practices, and warnings. The experience of actual clients and users of the Ten Steps provide real examples of outputs for the steps plus highlighted, sidebar case studies called Ten Steps in Action. This book uses projects as the vehicle for data quality work and the word broadly to include: 1) focused data quality improvement projects, such as improving data used in supply chain management, 2) data quality activities in other projects such as building new applications and migrating data from legacy systems, integrating data because of mergers and acquisitions, or untangling data due to organizational breakups, and 3) ad hoc use of data quality steps, techniques, or activities in the course of daily work. The Ten Steps approach can also be used to enrich an organization’s standard SDLC (whether sequential or Agile) and it complements general improvement methodologies such as six sigma or lean. No two data quality projects are the same but the flexible nature of the Ten Steps means the methodology can be applied to all. The new Second Edition highlights topics such as artificial intelligence and machine learning, Internet of Things, security and privacy, analytics, legal and regulatory requirements, data science, big data, data lakes, and cloud computing, among others, to show their dependence on data and information and why data quality is more relevant and critical now than ever before. Includes concrete instructions, numerous templates, and practical advice for executing every step of The Ten Steps approach Contains real examples from around the world, gleaned from the author’s consulting practice and from those who implemented based on her training courses and the earlier edition of the book Allows for quick reference with an easy-to-use format highlighting key concepts and definitions, important checkpoints, communication activities, and best practices A companion Web site includes links to numerous data quality resources, including many of the templates featured in the text, quick summaries of key ideas from the Ten Steps methodology, and other tools and information that are available online

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

Regional health care databases are being established around the country with the goal of providing timely and useful information to policymakers, physicians, and patients. But their emergence is raising important and sometimes controversial questions about the collection, quality, and appropriate use of health care data. Based on experience with databases now in operation and in development, *Health Data in the Information Age* provides a clear set of guidelines and principles for exploiting the potential benefits of aggregated health data--without jeopardizing confidentiality. A panel of experts identifies characteristics of emerging health database organizations (HDOs). The committee explores how HDOs can maintain the quality of their data, what policies and practices they should adopt, how they can prepare for linkages with computer-based patient records, and how diverse groups from researchers to health care administrators might use aggregated data. *Health Data in the Information Age* offers frank analysis and guidelines that will be invaluable to anyone interested in the operation of health care databases.

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. This User's Guide is intended to support the design, implementation, analysis, interpretation, and quality evaluation of registries created to increase understanding of patient outcomes. For the purposes of this guide, a patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes. A registry database is a file (or files) derived from the registry. Although registries can serve many purposes, this guide focuses on registries created for one or more of the following purposes: to describe the natural history of disease, to determine clinical effectiveness or cost-effectiveness of health care products and services, to measure or monitor safety and harm, and/or to measure quality of care. Registries are

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

classified according to how their populations are defined. For example, product registries include patients who have been exposed to biopharmaceutical products or medical devices. Health services registries consist of patients who have had a common procedure, clinical encounter, or hospitalization. Disease or condition registries are defined by patients having the same diagnosis, such as cystic fibrosis or heart failure. The User's Guide was created by researchers affiliated with AHRQ's Effective Health Care Program, particularly those who participated in AHRQ's DEcIDE (Developing Evidence to Inform Decisions About Effectiveness) program. Chapters were subject to multiple internal and external independent reviews.

The U.S. Geological Survey (USGS) mission is to provide reliable and impartial scientific information to understand Earth, minimize loss of life and property from natural disasters, and manage water, biological, energy, and mineral resources. Data collection, analysis, interpretation, and dissemination are central to everything the USGS does. Among other activities, the USGS operates some 250 laboratories across the country to analyze physical and biological samples, including water, sediment, rock, plants, invertebrates, fish, and wildlife. The data generated in the laboratories help answer pressing scientific and societal questions or support regulation, resource management, or commercial applications. At the request of the USGS, this study reviews a representative sample of USGS laboratories to examine quality management systems and other approaches for assuring the quality of laboratory results and recommends best practices and procedures for USGS laboratories.

Using case studies and hands-on activities, this book discusses topics in information governance (IG): recognizing hidden development and operational implications of IG--and why it needs to be integrated in the broader organization; integrating IG activities with transactional processing, BI, MDM, and other enterprise information management functions; the information governance organization: defining roles, launching projects, and integrating with ongoing operations; performing IG in transactional projects, including those using agile methods and COTS products; bringing stronger information governance to MDM: strategy, architecture, development, and beyond; governing information throughout the BI or big data project lifecycle; performing ongoing IG and data stewardship operational processes; auditing and enforcing data quality management in the context of enterprise information management; maintaining and evolving metadata management for maximum business value. -- \$c Edited summary from book.

Executing Data Quality Projects Ten Steps to Quality Data and Trusted Information (TM) Academic Press

The Practitioner's Guide to Data Quality Improvement offers a comprehensive look at data quality for business and IT, encompassing people, process, and technology. It shares the fundamentals for understanding the impacts of poor data quality, and guides practitioners and managers alike in socializing, gaining sponsorship for, planning, and establishing a

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

data quality program. It demonstrates how to institute and run a data quality program, from first thoughts and justifications to maintenance and ongoing metrics. It includes an in-depth look at the use of data quality tools, including business case templates, and tools for analysis, reporting, and strategic planning. This book is recommended for data management practitioners, including database analysts, information analysts, data administrators, data architects, enterprise architects, data warehouse engineers, and systems analysts, and their managers. Offers a comprehensive look at data quality for business and IT, encompassing people, process, and technology. Shows how to institute and run a data quality program, from first thoughts and justifications to maintenance and ongoing metrics. Includes an in-depth look at the use of data quality tools, including business case templates, and tools for analysis, reporting, and strategic planning. "What do you need to become a data-driven organization? Far more than having big data or a crack team of unicorn data scientists, it requires establishing an effective, deeply-ingrained data culture. This practical book shows you how true data-drivenness involves processes that require genuine buy-in across your company ... Through interviews and examples from data scientists and analytics leaders in a variety of industries ... Anderson explains the analytics value chain you need to adopt when building predictive business models"--Publisher's description.

Data in its raw state is rarely ready for productive analysis. This book not only teaches you data preparation, but also what questions you should ask of your data. It focuses on the thought processes necessary for successful data cleaning as much as on concise and precise code examples that express these thoughts.

Poor data quality can seriously hinder or damage the efficiency and effectiveness of organizations and businesses. The growing awareness of such repercussions has led to major public initiatives like the "Data Quality Act" in the USA and the "European 2003/98" directive of the European Parliament. Batini and Scannapieco present a comprehensive and systematic introduction to the wide set of issues related to data quality. They start with a detailed description of different data quality dimensions, like accuracy, completeness, and consistency, and their importance in different types of data, like federated data, web data, or time-dependent data, and in different data categories classified according to frequency of change, like stable, long-term, and frequently changing data. The book's extensive description of techniques and methodologies from core data quality research as well as from related fields like data mining, probability theory, statistical data analysis, and machine learning gives an excellent overview of the current state of the art. The presentation is completed by a short description and critical comparison of tools and practical methodologies, which will help readers to resolve their own quality problems. This book is an ideal combination of the soundness of theoretical foundations and the applicability of practical approaches. It is ideally suited for everyone – researchers, students, or professionals – interested in a comprehensive overview of data quality issues. In addition, it will serve as the basis for an introductory course or for

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

self-study on this topic.

Many senior executives talk about information as one of their most important assets, but few behave as if it is. They report to the board on the health of their workforce, their financials, their customers, and their partnerships, but rarely the health of their information assets. Corporations typically exhibit greater discipline in tracking and accounting for their office furniture than their data. Infonomics is the theory, study, and discipline of asserting economic significance to information. It strives to apply both economic and asset management principles and practices to the valuation, handling, and deployment of information assets. This book specifically shows: CEOs and business leaders how to more fully wield information as a corporate asset CIOs how to improve the flow and accessibility of information CFOs how to help their organizations measure the actual and latent value in their information assets. More directly, this book is for the burgeoning force of chief data officers (CDOs) and other information and analytics leaders in their valiant struggle to help their organizations become more infosavvy. Author Douglas Laney has spent years researching and developing Infonomics and advising organizations on the infinite opportunities to monetize, manage, and measure information. This book delivers a set of new ideas, frameworks, evidence, and even approaches adapted from other disciplines on how to administer, wield, and understand the value of information. Infonomics can help organizations not only to better develop, sell, and market their offerings, but to transform their organizations altogether. "Doug Laney masterfully weaves together a collection of great examples with a solid framework to guide readers on how to gain competitive advantage through what he labels "the unruly asset" – data. The framework is comprehensive, the advice practical and the success stories global and across industries and applications." Liz Rowe, Chief Data Officer, State of New Jersey "A must read for anybody who wants to survive in a data centric world." Shaun Adams, Head of Data Science, Betterbathrooms.com "Phenomenal! An absolute must read for data practitioners, business leaders and technology strategists. Doug's lucid style has set a new standard in providing intelligible material in the field of information economics. His passion and knowledge on the subject exudes thru his literature and inspires individuals like me." Ruchi Rajasekhar, Principal Data Architect, MISO Energy "I highly recommend Infonomics to all aspiring analytics leaders. Doug Laney's work gives readers a deeper understanding of how and why information should be monetized and managed as an enterprise asset. Laney's assertion that accounting should recognize information as a capital asset is quite convincing and one I agree with. Infonomics enjoyably echoes that sentiment!" Matt Green, independent business analytics consultant, Atlanta area "If you care about the digital economy, and you should, read this book." Tanya Shuckhart, Analyst Relations Lead, IRI Worldwide

In a world without political freedom, personal freedom and precious little faith in anything comes a mysterious man in a

white porcelain mask who fights political oppressors through terrorism and seemingly absurd acts. It's a gripping tale of the blurred lines between ideological good and evil. The inspiration for the hit 2005 movie starring Natalie Portman and Hugo Weaving, this amazing graphic novel is packaged with a collectable reproduction of the iconic "V" mask. Foreword by Oliver Schabenberger, PhD Executive Vice President, Chief Operating Officer and Chief Technology Officer SAS Dive into deep learning! Machine learning and deep learning are ubiquitous in our homes and workplaces—from machine translation to image recognition and predictive analytics to autonomous driving. Deep learning holds the promise of improving many everyday tasks in a variety of disciplines. Much deep learning literature explains the mechanics of deep learning with the goal of implementing cognitive applications fueled by Big Data. This book is different. Written by an expert in high-performance analytics, *Deep Learning for Numerical Applications with SAS* introduces a new field: Deep Learning for Numerical Applications (DL4NA). Contrary to deep learning, the primary goal of DL4NA is not to learn from data but to dramatically improve the performance of numerical applications by training deep neural networks. *Deep Learning for Numerical Applications with SAS* presents deep learning concepts in SAS along with step-by-step techniques that allow you to easily reproduce the examples on your high-performance analytics systems. It also discusses the latest hardware innovations that can power your SAS programs: from many-core CPUs to GPUs to FPGAs to ASICs. This book assumes the reader has no prior knowledge of high-performance computing, machine learning, or deep learning. It is intended for SAS developers who want to develop and run the fastest analytics. In addition to discovering the latest trends in hybrid architectures with GPUs and FPGAS, readers will learn how to Use deep learning in SAS Speed up their analytics using deep learning Easily write highly parallel programs using the many task computing paradigms This book is part of the SAS Press program.

The Structure of Digital Computing takes a fifty year perspective on computing and discusses what is significant, what is novel, what endures, and why it is all so confusing. The book tries to balance two point of views: digital computing as viewed from a business perspective, where the focus is on marketing and selling, and digital computing from a research perspective, where the focus is on developing fundamentally new technology.

Imagine a group of prehistoric hunters armed with stone-tipped spears. Their primitive weapons made hunting large animals, such as mammoths, dangerous work. Over time, however, a new breed of hunters developed. They would stretch the skin of a previously killed mammoth on the wall and throw their spears, while observing which spear, thrown from which angle and distance, penetrated the skin the best. The data gathered helped them make better spears and develop better hunting strategies. Quality data is the key to any advancement, whether it's from the Stone Age to the Bronze Age. Or from the Information Age to whatever Age comes next. The success of corporations and government

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

institutions largely depends on the efficiency with which they can collect, organize, and utilize data about products, customers, competitors, and employees. Fortunately, improving your data quality doesn't have to be such a mammoth task. DATA QUALITY ASSESSMENT is a must read for anyone who needs to understand, correct, or prevent data quality issues in their organization. Skipping theory and focusing purely on what is practical and what works, this text contains a proven approach to identifying, warehousing, and analyzing data errors – the first step in any data quality program. Master techniques in:

- Data profiling and gathering metadata
- Identifying, designing, and implementing data quality rules
- Organizing rule and error catalogues
- Ensuring accuracy and completeness of the data quality assessment
- Constructing the dimensional data quality scorecard
- Executing a recurrent data quality assessment

This is one of those books that marks a milestone in the evolution of a discipline. Arkady's insights and techniques fuel the transition of data quality management from art to science -- from crafting to engineering. From deep experience, with thoughtful structure, and with engaging style Arkady brings the discipline of data quality to practitioners. David Wells, Director of Education, Data Warehousing Institute

The Data Quality Assessment Framework shows you how to measure and monitor data quality, ensuring quality over time. You'll start with general concepts of measurement and work your way through a detailed framework of more than three dozen measurement types related to five objective dimensions of quality: completeness, timeliness, consistency, validity, and integrity. Ongoing measurement, rather than one time activities will help your organization reach a new level of data quality. This plain-language approach to measuring data can be understood by both business and IT and provides practical guidance on how to apply the DQAF within any organization enabling you to prioritize measurements and effectively report on results. Strategies for using data measurement to govern and improve the quality of data and guidelines for applying the framework within a data asset are included. You'll come away able to prioritize which measurement types to implement, knowing where to place them in a data flow and how frequently to measure. Common conceptual models for defining and storing of data quality results for purposes of trend analysis are also included as well as generic business requirements for ongoing measuring and monitoring including calculations and comparisons that make the measurements meaningful and help understand trends and detect anomalies. Demonstrates how to leverage a technology independent data quality measurement framework for your specific business priorities and data quality challenges Enables discussions between business and IT with a non-technical vocabulary for data quality measurement Describes how to measure data quality on an ongoing basis with generic measurement types that can be applied to any situation

Managing data continues to grow as a necessity for modern organizations. There are seemingly infinite opportunities for

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

organic growth, reduction of costs, and creation of new products and services. It has become apparent that none of these opportunities can happen smoothly without data governance. The cost of exponential data growth and privacy / security concerns are becoming burdensome. Organizations will encounter unexpected consequences in new sources of risk. The solution to these challenges is also data governance; ensuring balance between risk and opportunity. Data Governance, Second Edition, is for any executive, manager or data professional who needs to understand or implement a data governance program. It is required to ensure consistent, accurate and reliable data across their organization. This book offers an overview of why data governance is needed, how to design, initiate, and execute a program and how to keep the program sustainable. This valuable resource provides comprehensive guidance to beginning professionals, managers or analysts looking to improve their processes, and advanced students in Data Management and related courses. With the provided framework and case studies all professionals in the data governance field will gain key insights into launching successful and money-saving data governance program. Incorporates industry changes, lessons learned and new approaches Explores various ways in which data analysts and managers can ensure consistent, accurate and reliable data across their organizations Includes new case studies which detail real-world situations Explores all of the capabilities an organization must adopt to become data driven Provides guidance on various approaches to data governance, to determine whether an organization should be low profile, central controlled, agile, or traditional Provides guidance on using technology and separating vendor hype from sincere delivery of necessary capabilities Offers readers insights into how their organizations can improve the value of their data, through data quality, data strategy and data literacy Provides up to 75% brand-new content compared to the first edition

Are there any constraints known that bear on the ability to perform Agile Management for Software Engineering work? How is the team addressing them? In a project to restructure Agile Management for Software Engineering outcomes, which stakeholders would you involve? How much are sponsors, customers, partners, stakeholders involved in Agile Management for Software Engineering? In other words, what are the risks, if Agile Management for Software Engineering does not deliver successfully? How does the organization define, manage, and improve its Agile Management for Software Engineering processes? What are the business goals Agile Management for Software Engineering is aiming to achieve? Defining, designing, creating, and implementing a process to solve a business challenge or meet a business objective is the most valuable role... In EVERY company, organization and department. Unless you are talking a one-time, single-use project within a business, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

we really trying to accomplish here? And is there a different way to look at it?' For more than twenty years, The Art of Service's Self-Assessments empower people who can do just that - whether their title is marketer, entrepreneur, manager, salesperson, consultant, business process manager, executive assistant, IT Manager, CxO etc... - they are the people who rule the future. They are people who watch the process as it happens, and ask the right questions to make the process work better. This book is for managers, advisors, consultants, specialists, professionals and anyone interested in Agile Management for Software Engineering assessment. All the tools you need to an in-depth Agile Management for Software Engineering Self-Assessment. Featuring 616 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Agile Management for Software Engineering improvements can be made. In using the questions you will be better able to: - diagnose Agile Management for Software Engineering projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent advances in Agile Management for Software Engineering and process design strategies into practice according to best practice guidelines Using a Self-Assessment tool known as the Agile Management for Software Engineering Scorecard, you will develop a clear picture of which Agile Management for Software Engineering areas need attention. Included with your purchase of the book is the Agile Management for Software Engineering Self-Assessment downloadable resource, which contains all questions and Self-Assessment areas of this book in a ready to use Excel dashboard, including the self-assessment, graphic insights, and project planning automation - all with examples to get you started with the assessment right away. Access instructions can be found in the book. You are free to use the Self-Assessment contents in your presentations and materials for customers without asking us - we are here to help.

Ten Strategies of a World-Class Cyber Security Operations Center conveys MITRE's accumulated expertise on enterprise-grade computer network defense. It covers ten key qualities of leading Cyber Security Operations Centers (CSOCs), ranging from their structure and organization, to processes that best enable smooth operations, to approaches that extract maximum value from key CSOC technology investments. This book offers perspective and context for key decision points in structuring a CSOC, such as what capabilities to offer, how to architect large-scale data collection and analysis, and how to prepare the CSOC team for agile, threat-based response. If you manage, work in, or are standing up a CSOC, this book is for you. It is also available on MITRE's website, www.mitre.org.

“This is not the kind of book that you’ll read one time and be done with. So scan it quickly the first time through to get an idea of its breadth. Then dig in on one topic of special importance to your work. Finally, use it as a reference to guide

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

your next steps, learn details, and broaden your perspective.” from the foreword by Thomas C. Redman, Ph.D., “the Data Doc” Good data is a source of myriad opportunities, while bad data is a tremendous burden. Companies that manage their data effectively are able to achieve a competitive advantage in the marketplace, while bad data, like cancer, can weaken and kill an organization. In this comprehensive book, Rupa Mahanti provides guidance on the different aspects of data quality with the aim to be able to improve data quality. Specifically, the book addresses: -Causes of bad data quality, bad data quality impacts, and importance of data quality to justify the case for data quality-Butterfly effect of data quality- A detailed description of data quality dimensions and their measurement-Data quality strategy approach-Six Sigma - DMAIC approach to data quality-Data quality management techniques-Data quality in relation to data initiatives like data migration, MDM, data governance, etc.-Data quality myths, challenges, and critical success factors Students, academicians, professionals, and researchers can all use the content in this book to further their knowledge and get guidance on their own specific projects. It balances technical details (for example, SQL statements, relational database components, data quality dimensions measurements) and higher-level qualitative discussions (cost of data quality, data quality strategy, data quality maturity, the case made for data quality, and so on) with case studies, illustrations, and real-world examples throughout.

Data Quality: The Accuracy Dimension is about assessing the quality of corporate data and improving its accuracy using the data profiling method. Corporate data is increasingly important as companies continue to find new ways to use it. Likewise, improving the accuracy of data in information systems is fast becoming a major goal as companies realize how much it affects their bottom line. Data profiling is a new technology that supports and enhances the accuracy of databases throughout major IT shops. Jack Olson explains data profiling and shows how it fits into the larger picture of data quality. * Provides an accessible, enjoyable introduction to the subject of data accuracy, peppered with real-world anecdotes. * Provides a framework for data profiling with a discussion of analytical tools appropriate for assessing data accuracy. * Is written by one of the original developers of data profiling technology. * Is a must-read for any data management staff, IT management staff, and CIOs of companies with data assets.

It's necessary for enterprises nowadays to collect massive amounts of data for their business - data concerning everything from products and services to market conditions and customer demographics. With so much information at your fingertips, it's difficult to find the pieces that are accurate and relevant. This problem is exacerbated by the growth that comes from acquisition. Consolidating data becomes a real issue for newly integrated companies or recently launched branches. After all, standardizing software - considering international legal requirements, different roles, responsibilities, and processes-is challenging when you're running a global enterprise. Helmut Schindlwick proposes a

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

solution for these headaches through IT governance. You must put a system in place for inaccurate or redundant data, as well as data that becomes obsolete, and then continuously monitor data quality to ensure optimization. Instead of focusing on software, you should first take time to determine the structure of your master data. Schindlwick shares all the knowledge and experience he's gleaned from eighteen years of consulting work, approaching a complex topic in a way that's practical and easily understood. His unique Transformation Eco System, or TESy, will guide readers through concepts and communicate information plainly.

The definitive reference book with real-world solutions you won't find anywhere else The Big Book of Dashboards presents a comprehensive reference for those tasked with building or overseeing the development of business dashboards. Comprising dozens of examples that address different industries and departments (healthcare, transportation, finance, human resources, marketing, customer service, sports, etc.) and different platforms (print, desktop, tablet, smartphone, and conference room display) The Big Book of Dashboards is the only book that matches great dashboards with real-world business scenarios. By organizing the book based on these scenarios and offering practical and effective visualization examples, The Big Book of Dashboards will be the trusted resource that you open when you need to build an effective business dashboard. In addition to the scenarios there's an entire section of the book that is devoted to addressing many practical and psychological factors you will encounter in your work. It's great to have theory and evidenced-based research at your disposal, but what will you do when somebody asks you to make your dashboard 'cooler' by adding packed bubbles and donut charts? The expert authors have a combined 30-plus years of hands-on experience helping people in hundreds of organizations build effective visualizations. They have fought many 'best practices' battles and having endured bring an uncommon empathy to help you, the reader of this book, survive and thrive in the data visualization world. A well-designed dashboard can point out risks, opportunities, and more; but common challenges and misconceptions can make your dashboard useless at best, and misleading at worst. The Big Book of Dashboards gives you the tools, guidance, and models you need to produce great dashboards that inform, enlighten, and engage.

In 1997, Congress, in the conference report, H.R. 105-271, to the FY1998 Energy and Water Development Appropriation Bill, directed the National Research Council (NRC) to carry out a series of assessments of project management at the Department of Energy (DOE). The final report in that series noted that DOE lacked an objective set of measures for assessing project management quality. The department set up a committee to develop performance measures and benchmarking procedures and asked the NRC for assistance in this effort. This report presents information and guidance for use as a first step toward development of a viable methodology to suit DOE's needs. It provides a number of possible performance measures, an

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

analysis of the benchmarking process, and a description ways to implement the measures and benchmarking process.

To order please visit <https://onlineacademiccommunity.uvic.ca/press/books/ordering/>

The issue of data quality is as old as data itself. However, the proliferation of diverse, large-scale and often publically available data on the Web has increased the risk of poor data quality and misleading data interpretations. On the other hand, data is now exposed at a much more strategic level e.g. through business intelligence systems, increasing manifold the stakes involved for individuals, corporations as well as government agencies. There, the lack of knowledge about data accuracy, currency or completeness can have erroneous and even catastrophic results. With these changes, traditional approaches to data management in general, and data quality control specifically, are challenged. There is an evident need to incorporate data quality considerations into the whole data cycle, encompassing managerial/governance as well as technical aspects. Data quality experts from research and industry agree that a unified framework for data quality management should bring together organizational, architectural and computational approaches. Accordingly, Sadiq structured this handbook in four parts: Part I is on organizational solutions, i.e. the development of data quality objectives for the organization, and the development of strategies to establish roles, processes, policies, and standards required to manage and ensure data quality. Part II, on architectural solutions, covers the technology landscape required to deploy developed data quality management processes, standards and policies. Part III, on computational solutions, presents effective and efficient tools and techniques related to record linkage, lineage and provenance, data uncertainty, and advanced integrity constraints. Finally, Part IV is devoted to case studies of successful data quality initiatives that highlight the various aspects of data quality in action. The individual chapters present both an overview of the respective topic in terms of historical research and/or practice and state of the art, as well as specific techniques, methodologies and frameworks developed by the individual contributors. Researchers and students of computer science, information systems, or business management as well as data professionals and practitioners will benefit most from this handbook by not only focusing on the various sections relevant to their research area or particular practical work, but by also studying chapters that they may initially consider not to be directly relevant to them, as there they will learn about new perspectives and approaches.

This Handbook provides a comprehensive ten-step model that will help guide development practitioners through the process of designing and building a results-based monitoring and evaluation system.

With this practical book, AI and machine learning practitioners will learn how to successfully build and deploy data science projects on Amazon Web Services. The Amazon AI and machine learning stack unifies data science, data engineering, and application development to help level up your skills. This guide shows you how to build and run pipelines in the cloud, then integrate the results into applications in minutes instead of days. Throughout the book, authors Chris Fregly and Antje Barth demonstrate how to reduce cost and improve performance. Apply the Amazon AI and ML stack to real-world use cases for natural language processing, computer vision, fraud detection, conversational devices, and more Use automated machine learning to implement a specific subset of use cases with SageMaker Autopilot Dive deep into the complete model development lifecycle for a BERT-

Read Book Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information

based NLP use case including data ingestion, analysis, model training, and deployment Tie everything together into a repeatable machine learning operations pipeline Explore real-time ML, anomaly detection, and streaming analytics on data streams with Amazon Kinesis and Managed Streaming for Apache Kafka Learn security best practices for data science projects and workflows including identity and access management, authentication, authorization, and more

[Copyright: 2fbd9bc2eb159c3bbb71cc6842331ae9](https://www.amazon.com/Executing-Data-Quality-Projects-Ten-Steps-To-Quality-Data-And-Trusted-Information/dp/1617095111)