

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

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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.

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.

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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 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. A key task that any aspiring data-driven organization needs to learn is data wrangling, the process of converting raw data into something truly useful. This practical guide provides business analysts with an overview of various data wrangling techniques and tools, and puts the practice of data wrangling into context by asking, "What are you trying to do and why?" Wrangling data consumes roughly 50-80% of an analyst's time before any kind of analysis is possible. Written by key executives at Trifacta, this book walks you through the wrangling process by exploring several factors—time, granularity, scope, and structure—that you need to consider as you begin to work with data. You'll learn a shared language and a comprehensive understanding of data wrangling, with an emphasis on recent agile analytic

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

processes used by many of today's data-driven organizations. Appreciate the importance—and the satisfaction—of wrangling data the right way. Understand what kind of data is available Choose which data to use and at what level of detail Meaningfully combine multiple sources of data Decide how to distill the results to a size and shape that can drive downstream analysis

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 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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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.

A thorough reference on database administration outlines a variety of DBA roles and responsibilities and discusses such topics as data modeling and normalization, database/application design, change management, database security and data integrity, performance issues, disaster planning, and other essentials. Original. (Advanced)

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

and guidelines that will be invaluable to anyone interested in the operation of health care databases. 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 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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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 "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.

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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-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

“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 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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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.

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

for understanding the impacts of poor data quality, and guides practitioners and managers alike in socializing, gaining sponsorship for, planning, and establishing a 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. Data stewards in business and IT are the backbone of a successful data governance implementation because they do the work to make a company's data trusted, dependable, and high quality. Data Stewardship explains everything you need to know to successfully implement the stewardship portion of data governance, including how to organize, train,

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

and work with data stewards, get high-quality business definitions and other metadata, and perform the day-to-day tasks using a minimum of the steward's time and effort. David Plotkin has loaded this book with practical advice on stewardship so you can get right to work, have early successes, and measure and communicate those successes, gaining more support for this critical effort. Provides clear and concise practical advice on implementing and running data stewardship, including guidelines on how to organize based on company structure, business functions, and data ownership Shows how to gain support for your stewardship effort, maintain that support over the long-term, and measure the success of the data stewardship effort and report back to management Includes detailed lists of responsibilities for each type of data steward and strategies to help the Data Governance Program Office work effectively with the data stewards

"A groundbreaking and illuminating look at the state of abortion access in America and the first long-term study of the consequences-emotional, physical, financial, professional, personal, and psychological-of receiving versus being denied an abortion on women's lives"--

Entity Resolution and Information Quality presents topics and definitions, and clarifies confusing terminologies regarding entity resolution and information quality. It takes a very wide view of IQ,

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

including its six-domain framework and the skills formed by the International Association for Information and Data Quality (IAIDQ). The book includes chapters that cover the principles of entity resolution and the principles of Information Quality, in addition to their concepts and terminology. It also discusses the Fellegi-Sunter theory of record linkage, the Stanford Entity Resolution Framework, and the Algebraic Model for Entity Resolution, which are the major theoretical models that support Entity Resolution. In relation to this, the book briefly discusses entity-based data integration (EBDI) and its model, which serve as an extension of the Algebraic Model for Entity Resolution. There is also an explanation of how the three commercial ER systems operate and a description of the non-commercial open-source system known as OYSTER. The book concludes by discussing trends in entity resolution research and practice. Students taking IT courses and IT professionals will find this book invaluable. First authoritative reference explaining entity resolution and how to use it effectively Provides practical system design advice to help you get a competitive advantage Includes a companion site with synthetic customer data for applicatory exercises, and access to a Java-based Entity Resolution program.

Can any subject inspire less excitement than "data quality"? Yet a moment's thought reveals the ever-

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

figures and tables into documents within their organizations Gives step-by-step instructions for applying most techniques and summarizes what "works"

Data-driven insights are a key competitive advantage for any industry today, but deriving insights from raw data can still take days or weeks. Most organizations can't scale data science teams fast enough to keep up with the growing amounts of data to transform. What's the answer? Self-service data. With this practical book, data engineers, data scientists, and team managers will learn how to build a self-service data science platform that helps anyone in your organization extract insights from data. Sandeep Uttamchandani provides a scorecard to track and address bottlenecks that slow down time to insight across data discovery, transformation, processing, and production. This book bridges the gap between data scientists bottlenecked by engineering realities and data engineers unclear about ways to make self-service work. Build a self-service portal to support data discovery, quality, lineage, and governance Select the best approach for each self-service capability using open source cloud technologies Tailor self-service for the people, processes, and technology maturity of your data platform Implement capabilities to democratize data and reduce time to insight Scale your self-service portal to support a large number of users within your

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm organization

Managing information within the enterprise has always been a vital and important task to support the day-to-day business operations and to enable analysis of that data for decision making to better manage and grow the business for improved profitability. To do all that, clearly the data must be accurate and organized so it is accessible and understandable to all who need it. That task has grown in importance as the volume of enterprise data has been growing significantly (analyst estimates of 40 - 50% growth per year are not uncommon) over the years. However, most of that data has been what we call "structured" data, which is the type that can fit neatly into rows and columns and be more easily analyzed. Now we are in the era of "big data." This significantly increases the volume of data available, but it is in a form called "unstructured" data. That is, data from sources that are not as easily organized, such as data from emails, spreadsheets, sensors, video, audio, and social media sites. There is valuable information in all that data but it calls for new processes to enable it to be analyzed. All this has brought with it a renewed and critical need to manage and organize that data with clarity of meaning, understandability, and interoperability. That is, you must be able to integrate this data when it is from within an enterprise but also importantly when it is from many

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

different external sources. What is described here has been and is being done to varying extents. It is called "information governance." Governing this information however has proven to be challenging. But without governance, much of the data can be less useful and perhaps even used incorrectly, significantly impacting enterprise decision making. So we must also respect the needs for information security, consistency, and validity or else suffer the potential economic and legal consequences. Implementing sound governance practices needs to be an integral part of the information control in our organizations. This IBM® Redbooks® publication focuses on the building blocks of a solid governance program. It examines some familiar governance initiative scenarios, identifying how they underpin key governance initiatives, such as Master Data Management, Quality Management, Security and Privacy, and Information Lifecycle Management. IBM Information Management and Governance solutions provide a comprehensive suite to help organizations better understand and build their governance solutions. The book also identifies new and innovative approaches that are developed by IBM practice leaders that can help as you implement the foundation capabilities in your organizations. Managing data continues to grow as a necessity for modern organizations. There are seemingly infinite opportunities for organic growth, reduction of costs,

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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 self-study on this topic. With clear instructions for developing a research design and complementary research tools, this book is not about describing or theorizing qualitative methods, but how researchers actually create and execute these methods. Helping students conquer the practical issues many novice researchers face, the book provides them with the tools they need to answer critical questions such as: what are some ways to sample potential participants? how do I construct an interview schedule? should I be thinking of a single case study or a comparative study? what and how should I record in the field? what other sources of data should I consider?

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

To support the broadening spectrum of project delivery approaches, PMI is offering A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Sixth Edition as a bundle with its latest, the Agile Practice Guide. The PMBOK® Guide – Sixth Edition now contains detailed information about agile; while the Agile Practice Guide, created in partnership with Agile Alliance®, serves as a bridge to connect waterfall and agile. Together they are a powerful tool for project managers. The PMBOK® Guide – Sixth Edition – PMI's flagship publication has been updated to reflect the latest good practices in project management. New to the Sixth Edition, each knowledge area will contain a section entitled Approaches for Agile, Iterative and Adaptive Environments, describing how these practices integrate in project settings. It will also contain more emphasis on strategic and business knowledge—including discussion of project management business documents—and information on the PMI Talent Triangle™ and the essential skills for success in today's market. Agile Practice Guide has been developed as a resource to understand, evaluate, and use agile and hybrid agile approaches. This practice guide provides guidance on when, where, and how to apply agile approaches and provides practical tools for practitioners and organizations wanting to increase agility. This practice guide is aligned with other PMI standards,

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

including A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Sixth Edition, and was developed as the result of collaboration between the Project Management Institute and the Agile Alliance.

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

reducing corporate risks by applying consistent enterprise-wide practices for the management and control of master data.

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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.

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.

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

read for data practitioners, business leaders and technology strategists. Doug's lucid style has a 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

The Data and Analytics Playbook: Proven Methods for Governed Data and Analytic Quality explores the way in which data continues to dominate budgets, along with the varying efforts made across a variety of business enablement projects, including applications, web and mobile computing, big data analytics, and traditional data integration. The book teaches readers how to use proven methods and accelerators to break through data obstacles to provide faster, higher quality delivery of mission critical programs. Drawing upon years of practical experience, and using numerous examples and an easy to understand playbook, Lowell Fryman, Gregory Lampshire, and Dan Meers discuss a simple,

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

proven approach to the execution of multiple data oriented activities. In addition, they present a clear set of methods to provide reliable governance, controls, risk, and exposure management for enterprise data and the programs that rely upon it. In addition, they discuss a cost-effective approach to providing sustainable governance and quality outcomes that enhance project delivery, while also ensuring ongoing controls. Example activities, templates, outputs, resources, and roles are explored, along with different organizational models in common use today and the ways they can be mapped to leverage playbook data governance throughout the organization. Provides a mature and proven playbook approach (methodology) to enabling data governance that supports agile implementation Features specific examples of current industry challenges in enterprise risk management, including anti-money laundering and fraud prevention Describes business benefit measures and funding approaches using exposure based cost models that augment risk models for cost avoidance analysis and accelerated delivery approaches using data integration sprints for application, integration, and information delivery success

“We finally have the definitive treatise on PyTorch! It covers the basics and abstractions in great detail. I hope this book becomes your extended reference document.”

—Soumith Chintala, co-creator of PyTorch Key Features

Written by PyTorch’s creator and key contributors

Develop deep learning models in a familiar Pythonic way

Use PyTorch to build an image classifier for cancer

detection Diagnose problems with your neural network

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

and improve training with data augmentation Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About The Book Every other day we hear about new ways to put deep learning to good use: improved medical imaging, accurate credit card fraud detection, long range weather forecasting, and more. PyTorch puts these superpowers in your hands. Instantly familiar to anyone who knows Python data tools like NumPy and Scikit-learn, PyTorch simplifies deep learning without sacrificing advanced features. It's great for building quick models, and it scales smoothly from laptop to enterprise. Deep Learning with PyTorch teaches you to create deep learning and neural network systems with PyTorch. This practical book gets you to work right away building a tumor image classifier from scratch. After covering the basics, you'll learn best practices for the entire deep learning pipeline, tackling advanced projects as your PyTorch skills become more sophisticated. All code samples are easy to explore in downloadable Jupyter notebooks. What You Will Learn Understanding deep learning data structures such as tensors and neural networks Best practices for the PyTorch Tensor API, loading data in Python, and visualizing results Implementing modules and loss functions Utilizing pretrained models from PyTorch Hub Methods for training networks with limited inputs Sifting through unreliable results to diagnose and fix problems in your neural network Improve your results with augmented data, better model architecture, and fine tuning This Book Is Written For For Python programmers with an

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

interest in machine learning. No experience with PyTorch or other deep learning frameworks is required. About The Authors Eli Stevens has worked in Silicon Valley for the past 15 years as a software engineer, and the past 7 years as Chief Technical Officer of a startup making medical device software. Luca Antiga is co-founder and CEO of an AI engineering company located in Bergamo, Italy, and a regular contributor to PyTorch. Thomas Viehmann is a Machine Learning and PyTorch speciality trainer and consultant based in Munich, Germany and a PyTorch core developer. Table of Contents PART 1 - CORE PYTORCH 1 Introducing deep learning and the PyTorch Library 2 Pretrained networks 3 It starts with a tensor 4 Real-world data representation using tensors 5 The mechanics of learning 6 Using a neural network to fit the data 7 Telling birds from airplanes: Learning from images 8 Using convolutions to generalize PART 2 - LEARNING FROM IMAGES IN THE REAL WORLD: EARLY DETECTION OF LUNG CANCER 9 Using PyTorch to fight cancer 10 Combining data sources into a unified dataset 11 Training a classification model to detect suspected tumors 12 Improving training with metrics and augmentation 13 Using segmentation to find suspected nodules 14 End-to-end nodule analysis, and where to go next PART 3 - DEPLOYMENT 15 Deploying to production

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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.

Design quality SAS software and evaluate SAS software quality SAS Data Analytic Development is the developer's compendium for writing better-performing software and the manager's guide to building comprehensive software performance requirements. The text introduces and parallels the International Organization for Standardization (ISO) software product quality model, demonstrating 15 performance requirements that represent dimensions of software quality, including: reliability, recoverability, robustness, execution efficiency (i.e., speed), efficiency, scalability, portability, security, automation, maintainability, modularity, readability, testability, stability, and reusability. The text is intended to be read cover-to-cover or used as a reference tool to instruct, inspire, deliver, and evaluate software quality. A common fault in many software development environments is a focus on functional requirements—the what and how—to the detriment of performance requirements, which specify instead how well software should function (assessed through software execution) or how easily software should be maintained (assessed through code inspection). Without the definition and communication of performance requirements, developers risk either

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

building software that lacks intended quality or wasting time delivering software that exceeds performance objectives—thus, either underperforming or gold-plating, both of which are undesirable. Managers, customers, and other decision makers should also understand the dimensions of software quality both to define performance requirements at project outset as well as to evaluate whether those objectives were met at software completion. As data analytic software, SAS transforms data into information and ultimately knowledge and data-driven decisions. Not surprisingly, data quality is a central focus and theme of SAS literature; however, code quality is far less commonly described and too often references only the speed or efficiency with which software should execute, omitting other critical dimensions of software quality. SAS® software project definitions and technical requirements often fall victim to this paradox, in which rigorous quality requirements exist for data and data products yet not for the software that undergirds them. By demonstrating the cost and benefits of software quality inclusion and the risk of software quality exclusion, stakeholders learn to value, prioritize, implement, and evaluate dimensions of software quality within risk management and project management frameworks of the software development life cycle (SDLC). Thus, SAS Data Analytic Development recalibrates business value, placing code quality on par with data quality, and performance requirements on par with functional requirements.

Information is currency. Recent studies show that data quality problems are costing businesses billions of

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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.

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.

How well does your organization manage the risks associated with information quality? Managing information risk is becoming a top priority on the organizational agenda. The increasing sophistication of IT capabilities along with the constantly changing dynamics of global competition are forcing businesses to

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

make use of their information more effectively.

Information is becoming a core resource and asset for all organizations; however, it also brings many potential risks to an organization, from strategic, operational, financial, compliance, and environmental to societal. If you continue to struggle to understand and measure how information and its quality affects your business, this book is for you. This reference is in direct response to the new challenges that all managers have to face. Our process helps your organization to understand the "pain points" regarding poor data and information quality so you can concentrate on problems that have a high impact on core business objectives. This book provides you with all the fundamental concepts, guidelines and tools to ensure core business information is identified, protected and used effectively, and written in a language that is clear and easy to understand for non-technical managers. Shows how to manage information risk using a holistic approach by examining information from all sources Offers varied perspectives of an author team that brings together academics, practitioners and researchers (both technical and managerial) to provide a comprehensive guide Provides real-life case studies with practical insight into the management of information risk and offers a basis for broader discussion among managers and practitioners

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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 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.

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,

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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 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.

This book is a revised edition of the best selling title

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

Implementing IT Governance (ISBN 978 90 8753 119 5). For trainers free additional material of this book is available. This can be found under the "Training Material" tab. Log in with your trainer account to access the material. In all enterprises around the world, the issues, opportunities and challenges of aligning IT more closely with the organization and effectively governing an organization's IT investments, resources, major initiatives and superior uninterrupted service is becoming a major concern of the Board and executive management. An integrated and comprehensive approach to the alignment, planning, execution and governance of IT and its resources has become critical to more effectively align, integrate, invest, measure, deploy, service and sustain the strategic and tactical direction and value proposition of IT in support of organizations. Much has been written and documented about the individual components of IT Governance such as strategic planning, demand management, program and project management, IT service management, strategic sourcing and outsourcing, performance management, metrics, compliance and others. Much less has been written about a comprehensive and integrated approach for IT/Business Alignment, Planning, Execution and Governance. This title fills that need in the marketplace and offers readers structured and practical solutions using the best of the best practices available today. The book is divided into two parts, which cover the three critical pillars necessary to develop, execute and sustain a robust and effective IT governance environment:- Leadership, people, organization and

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

strategy,- IT governance, its major component processes and enabling technologies. Each of the chapters also covers one or more of the following action oriented topics: - the why and what of IT: strategic planning, portfolio investment management, decision authority, etc.; - the how of IT: Program/Project Management, IT Service Management (including ITIL); Strategic Sourcing and outsourcing; performance, risk and contingency management (including COBIT, the Balanced Scorecard etc.) and leadership, team management and professional competences.

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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

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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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 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

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

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. Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use

File Type PDF Executing Data Quality Projects Ten Steps To Quality Data And Trusted Information Tm

flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples

[Copyright: 86695ac460c7ce5dc9cf00220398073f](https://www.pdfdrive.com/executing-data-quality-projects-ten-steps-to-quality-data-and-trusted-information-tm.html)