

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

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 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.

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

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-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

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 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.

The Accuracy Dimension

Measuring Data Quality for Ongoing Improvement

The Discipline of Getting Things Done

The Turnaway Study

Occupational Outlook Handbook

AI 2041

An Emerging Imperative

"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.

This volume explores the scientific frontiers and leading edges of research across the fields of anthropology, economics, political science, psychology, sociology, history, business, education, geography, law, and psychiatry, as well as the newer, more specialized areas of artificial intelligence, child development, cognitive science, communications, demography, linguistics, and management and decision science. It includes recommendations concerning new resources, facilities, and programs that may be needed over the next several years to ensure rapid progress and provide a high level of returns to basic research.

Data sharing can accelerate new discoveries by avoiding duplicative trials, stimulating new ideas for research, and enabling the maximal scientific knowledge and benefits to be gained from the efforts of clinical trial participants and investigators. At the same time, sharing clinical trial data presents risks, burdens, and challenges. These include the need to protect the privacy and honor the consent of clinical trial participants; safeguard the legitimate economic interests of sponsors; and guard against invalid secondary analyses, which could undermine trust in clinical trials or otherwise harm public health. Sharing Clinical Trial Data presents activities and strategies for the responsible sharing of clinical trial data. With the goal of increasing scientific knowledge to lead to better therapies for patients, this book identifies guiding principles and makes recommendations to maximize the benefits and minimize risks. This report offers guidance on the types of clinical trial data available at different points in the process, the points in the process at which each type of data should be shared, methods for sharing data, what groups should have access to data, and future knowledge and infrastructure needs. Responsible sharing of clinical trial data will allow other investigators to replicate published findings and carry out additional analyses, strengthen the evidence base for regulatory and clinical decisions, and increase the scientific knowledge gained from investments by the funders of clinical trials. The recommendations of Sharing Clinical Trial Data will be useful both now and well into the future as improved sharing of data leads to a stronger evidence base for treatment. This book will be of interest to stakeholders across the spectrum of research--from funders, to researchers, to journals, to physicians, and ultimately, to patients.

This is the single best book ever written on data quality. Clear, concise, and actionable. We all want to leverage our data resources to drive growth, but we too often ignore the fundamentals of data quality, which almost always inhibits our success. Tom lays out a clear path for each organization to holistically improve not only its data quality, but more importantly the performance of its business as a whole. "Jeffrey G. McMillan, Chief Analytics and Data Officer, Morgan Stanley This book lays out the roles everyone, up and down the organization chart, can and must play to ensure that data is up to the demands of its use, in day-in, day-out work, decision-making, planning, and analytics. By now, everyone knows that bad data extorts an enormous toll, adding huge (though often hidden) costs, and making it more difficult to make good decisions and leverage advanced analyses. While the problems are pervasive and insidious, they are also solvable! As Tom Redman, "the Data Doc," explains in Getting in Front on Data, the secret lies in getting the right people in the right roles to "get in front" of the management and social issues that lead to bad data in the first place. Everyone should see himself or herself in this book. We are all both data customers and data creators--after all, we use data created by others and create data used by others. And all of us must step up to these roles. As data customers, we must clarify our most important needs and communicate them to data creators. As data creators, we must strive to meet those needs by finding and eliminating the root causes of error. Getting in Front on Data proposes new roles for data professionals as: embedded data managers, in helping data customers and creators complete their work, DQ team leads, in connecting customers and creators, pulling the entire program together, and training people on their new roles, data maestros, in providing deep expertise on the really tough problems, chief data architects, in establishing common data definitions, and technologists, in increasing scale and decreasing unit cost.

Getting in Front on Data introduces a new role, the data provocateur, the motive force in attacking data quality properly! This book urges everyone to unleash their inner provocateur. Finally, it crystallizes what senior leaders must do if their entire organizations are to enjoy the benefits of high-quality data! Data quality has always been important. But now, in the growing digital economy where business transactions and customer experiences are automated and tailored, data quality is critical. This book comes just in time. "Maria C. Villar, Global Vice President, SAP America, Inc. Winning, and more importantly thriving, in the digital age requires more than stating "Data is a strategic corporate asset." Leaders and organizations need a plan of action to make the new vision a reality. Tom's latest book is a how-to for those seeking that reality. "Bob Palermo, Vice President, Performance Excellence, Shell Unconventionals Many, if not most, companies still struggle with their data. With his latest offering, Tom Redman sets out a path they can follow to Get in Front on Data. Based on his decades of experience working with many companies and individuals, this is the most practical guide around. A must read for data professionals, and especially data

"provocateurs". "Ken Self, President IAIDQ This book offers a unique perspective on how to think about data and address Data Quality " offering practical guidance and useful instruction from the perspective of each stakeholder. The process " and processes " to go from business need to having the right quality data to address that need is no small task. "John Nicodemo, Global Leader, Data Quality, Dun & Bradstreet Getting in Front on Data is a clearly written survival handbook for the new data-driven economy. It is a "must read" for the employees of any organization expecting to remain relevant and competitive. The "Data Doc" has an extraordinary talent for explaining key concepts with simple examples and understandable analogies making it accessible to everyone in their organization regardless of their role. "John R. Talburt, Director of the Information Quality Graduate Program University of Arkansas at Little Rock

Data Stewardship
Ten Steps to a Results-Based Monitoring and Evaluation System
Registries for Evaluating Patient Outcomes
Data Science on AWS
Infonomics
Learning Informatica PowerCenter 10.x
An Actionable Guide to Effective Data Management and Data Governance

Using case studies and hands-on activities, this book discusses topics in information governance (IG): recognizing hidden development and operational implications of IG to be integrated in the broader organization; integrating IG activities with transactional processing, BI, MDM, and other enterprise information management functions;

governance organization: defining roles, launching projects, and integrating with ongoing operations; performing IG in transactional projects, including those using agile COTS products; bringing stronger information governance to MDM: strategy, architecture, development, and beyond; governing information throughout the BI or big data performing ongoing IG and data stewardship operational processes; auditing and enforcing data quality management in the context of enterprise information management evolving metadata management for maximum business value. -- \$c Edited summary from book.

Data may be the most valuable resource that your organization owns. None of the promise of AI is possible without the ability to access, integrate, and transform data fundamentally changing the way our customers perform data management because changes in consumer expectations, and technology that drive them, continue to evolve rate. SAS offers many different data management solutions to handle and protect your data. The papers included in this special collection demonstrates the latest tools can benefit your data analysis. Also available free as a PDF from sas.com/books.

As you move data to the cloud, you need to consider a comprehensive approach to data governance, along with well-defined and agreed-upon policies to ensure your compliance requirements. Data governance incorporates the ways people, processes, and technology work together to ensure data is trustworthy and can be used effectively. This guide shows you how to effectively implement and scale data governance throughout your organization. Chief information, data, and security officers and their teams are using tooling to support democratizing data and unlocking its value while enforcing security, privacy, and other governance standards. Through good data governance, you can build trust, enable your organization to identify business efficiencies, generate more competitive offerings, and improve customer experience. This book shows you how. You will learn governance strategies addressing people, processes, and tools Benefits and challenges of a cloud-based data governance approach How data governance is conducted Data governance preparation and use How to handle the ongoing improvement of data quality Challenges and techniques in governing streaming data Data protection for authentication and monitoring How to build a data culture in your organization

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 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 accessible 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 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, 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 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 applicability of practical approaches. It is ideally suited for everyone – researchers, students, or professionals – interested in a comprehensive overview of data quality. This book will serve as the basis for an introductory course or for self-study on this topic.

Assuring Data Quality at U.S. Geological Survey Laboratories

Use, Disclosure, and Privacy

The Behavioral and Social Sciences

Getting in Front on Data

Achievements and Opportunities

How to Design, Deploy, and Sustain an Effective Data Governance Program

Maximizing Benefits, Minimizing Risk

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

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.

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.

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.

Execution

Special Collection

Data Cleaning

Managing Data Quality

A Step-by-Step Guide to Making Information Governance Work

A Guide to the Project Management Body of Knowledge (PMBOK® Guide) - Seventh Edition and The Standard for Project Management (RUSSIAN)

A Data Quality Assessment Framework

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.

☐ This is not the kind of book that you ☐ will 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 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.

This volume, developed by the Observatory together with OECD, provides an overall conceptual framework for understanding and applying strategies aimed at improving quality of care. Crucially, it summarizes available evidence on different quality strategies and provides recommendations for their implementation. This book is intended to help policy-makers to understand concepts of quality and to support them to evaluate single strategies and combinations of strategies.

PMBOK® Guide is the go-to resource for project management practitioners. The project management profession has significantly evolved due to emerging technology, new approaches and rapid market changes. Reflecting this evolution, The Standard for Project Management enumerates 12 principles of project management and the PMBOK® Guide &- Seventh Edition is structured around eight project

performance domains. This edition is designed to address practitioners' current and future needs and to help them be more proactive, innovative and nimble in enabling desired project outcomes. This edition of the PMBOK® Guide:

- Reflects the full range of development approaches (predictive, adaptive, hybrid, etc.);
- Provides an entire section devoted to tailoring the development approach and processes;
- Includes an expanded list of models, methods, and artifacts;
- Focuses on not just delivering project outputs but also enabling outcomes; and
- Integrates with PMI Standards+™ for information and standards application content based on project type, development approach, and industry sector.

A Practical Guide

Improving Healthcare Quality in Europe Characteristics, Effectiveness and Implementation of Different Strategies

Data Quality Assessment

Creating a Data-Driven Organization

Competing with High Quality Data

Who Does What

Data Driven

Data quality is one of the most important problems in data management, since dirty data often leads to inaccurate data analytics results and incorrect business decisions. Poor data across businesses and the U.S. government are reported to cost trillions of dollars a year. Multiple surveys show that dirty data is the most common barrier faced by data scientists. Not surprisingly, developing effective and efficient data cleaning solutions is challenging and is rife with deep theoretical and engineering problems. This book is about data cleaning, which is used to refer to all kinds of tasks and activities to detect and repair errors in the data. Rather than focus on a particular data cleaning task, we give an overview of the end-to-end data cleaning process, describing various error detection and repair methods, and attempt to anchor these proposals with multiple taxonomies and views. Specifically, we cover four of the most common and important data cleaning tasks, namely, outlier detection, data transformation, error repair (including imputing missing values), and data deduplication. Furthermore, due to the increasing popularity and applicability of machine learning techniques, we include a chapter that specifically explores how machine learning techniques are used for data cleaning, and how data cleaning is used to improve machine learning models. This book is intended to serve as a useful reference for researchers and practitioners who are interested in the area of data quality and data cleaning. It can also be used as a textbook for a graduate course. Although we aim at covering state-of-the-art algorithms and techniques, we recognize that data cleaning is still an active field of research and therefore provide future directions of research whenever appropriate.

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 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.

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.

For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn,

and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use: IPython and Jupyter: provide computational environments for data scientists using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib: includes capabilities for a flexible range of data visualizations in Python Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms

The Practitioner's Guide to Data Quality Improvement

Data Governance: The Definitive Guide

Health Data in the Information Age

Ten Steps to Quality Data and Trusted Information (TM)

Ten Years, a Thousand Women, and the Consequences of Having—or Being Denied—an Abortion

Visualizing Your Data Using Real-World Business Scenarios

Performing Information Governance

Your company's data has the potential to add enormous value to every facet of the organization -- from marketing and new product development to strategy to financial management. Yet if your company is like most, it's not using its data to create strategic advantage. Data sits around unused -- or incorrect data fouls up operations and decision making. In Data Driven, Thomas Redman, the "Data Doc," shows how to leverage and deploy data to sharpen your company's competitive edge and enhance its profitability. The author reveals: · The special properties that make data such a powerful asset · The hidden costs of flawed, outdated, or otherwise poor-quality data · How to improve data quality for competitive advantage · Strategies for exploiting your data to make better business decisions · The many ways to bring data to market · Ideas for dealing with political struggles over data and concerns about privacy rights Your company's data is a key business asset, and you need to manage it aggressively and professionally. Whether you're a top executive, an aspiring leader, or a product-line manager, this eye-opening book provides the tools and thinking you need to do that.

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. 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 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 Offers advice on how to lead an organization into change, including establishing a sense of urgency, developing a vision and strategy, and generating short-term wins.

Characteristics, Effectiveness and Implementation of Different Strategies

A Handbook for Development Practitioners

The Self-Service Data Roadmap

Achieving Your Wildly Important Goals

The Big Book of Dashboards

Python Data Science Handbook

Data Governance

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 organization

"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"--

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, 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

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 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.

Data Quality

Concepts, Tools, and Techniques for Building a Successful Approach to Data Quality

Dimensions, Measurement, Strategy, Management, and Governance

Ten Visions for Our Future

Data Management with SAS

Practical Advice from the Trenches

Executing Data Quality Projects

BUSINESS STRATEGY. "The 4 Disciplines of Execution" offers the what but also how effective execution is achieved. They share numerous examples of companies that have done just that, not once, but over and over again. This is a book that every leader should read! (Clayton Christensen, Professor, Harvard Business School, and author of "The Innovator's Dilemma"). Do you remember the last major initiative you watched die in your organization? Did it go down with a loud crash? Or was it slowly and quietly suffocated by other competing priorities? By the time it finally disappeared, it's likely no one even noticed. What happened? The whirlwind of urgent activity required to keep things running day-to-day devoured all the time and energy you needed to invest in executing your strategy for tomorrow. "The 4 Disciplines of Execution" can change all that forever.

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.

#1 NEW YORK TIMES BESTSELLER • More than two million copies in print! The premier resource for how to deliver results in an uncertain world, whether you're running an entire company or in your first management job. "A must-read for anyone who cares about business."—The New York Times When Execution was first published, it changed the way we did our jobs by focusing on the critical importance of "the discipline of execution": the ability to make the final leap to success by actually getting things done. Larry Bossidy and Ram Charan now reframe their empowering message for a world in which the old rules have been shattered, radical change is becoming routine, and the ability to execute is more important than ever. Now and for the foreseeable future: • Growth will be slower. But the company that executes well will have the confidence, speed, and resources to move fast as new opportunities emerge. • Competition will be fiercer, with companies searching for any possible advantage in every area from products and technologies to location and management. • Governments will take on new roles in their national economies, some as partners to business, others imposing constraints. Companies that execute well will be more attractive to government entities as partners and suppliers and better prepared to adapt to a new wave of regulation. • Risk management will become a top priority for every leader. Execution gives you an edge in detecting new internal and external threats and in weathering crises that can never be fully predicted. Execution shows how to link together people, strategy, and operations, the three core processes of every business. Leading these processes is the real job of running a business, not formulating a "vision" and leaving the work of carrying it out to others. Bossidy and Charan show the importance of being deeply and passionately engaged in an organization and why robust dialogues about people, strategy, and operations result in a business based on intellectual honesty and realism. With paradigmatic case histories from the real world—including examples like the diverging paths taken by Jamie Dimon at JPMorgan Chase and Charles Prince at Citigroup—Execution provides the realistic and hard-nosed approach to business success that could come only from authors as accomplished and insightful as Bossidy and Charan.

Written by a leading expert in the field, this account focuses on the convergence of two major trends in information management—big data and information governance—by taking a strategic approach oriented around business cases and industry imperatives. With the advent of new technologies, enterprises are expanding and handling very large volumes of data; this book, nontechnical in nature and geared toward business audiences, encourages the practice of establishing appropriate governance over big data initiatives and addresses how to manage and govern big data, highlighting the relevant processes, procedures, and policies. It teaches readers to understand how big data fits within an overall information governance program; quantify the business value of big data; apply information governance concepts such as stewardship, metadata, and organization structures to big data; appreciate the wide-ranging business benefits for various industries and job functions; sell the value of big data governance to businesses; and establish step-by-step processes to implement big data governance.

Essential Tools for Working with Data

Leading Change

The 4 Disciplines of Execution

How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage

Profiting from Your Most Important Business Asset

Big Data Governance

Sharing Clinical Trial Data

This book explains data quality management in practical terms, focusing on three key areas - the nature of data in enterprises, the purpose and scope of data quality management, and implementing a data quality management system, in line with ISO 8000-61. Examples of good practice in data quality management are also included.

How will artificial intelligence change our world within twenty years? A WALL STREET JOURNAL, WASHINGTON POST, AND FINANCIAL TIMES BEST BOOK OF THE YEAR • “This inspired collaboration between a pioneering technologist and a visionary writer of science fiction offers bold and urgent insights.”—Yann LeCun, winner of the Turing Award; chief AI scientist, Facebook “Amazingly entertaining . . . Lee and Chen take us on an immersive trip through the future. . . . Eye-opening.”—Mark Cuban AI will be the defining development of the twenty-first century. Within two decades, aspects of daily human life will be unrecognizable. AI will generate unprecedented wealth, revolutionize medicine and education through human-machine symbiosis, and create brand-new forms of communication and entertainment. In liberating us from routine work, however, AI will also challenge the organizing principles of our economic and social order. Meanwhile, AI will bring new risks in the form of autonomous weapons and smart technology that inherits human bias. AI is at a tipping point, and people need to wake up—both to AI’s radiant pathways and its existential perils for life as we know it. In this provocative, utterly original work, Kai-Fu Lee, the former president of Google China and bestselling author of AI Superpowers, teams up with celebrated novelist Chen Qiufan to imagine our world in 2041 and how it will be shaped by AI. In ten gripping short stories, they introduce readers to an array of eye-opening 2041 settings, such as: • In San Francisco, the “job reallocation” industry emerges as deep learning AI causes widespread job displacement • In Tokyo, a music fan is swept up in an immersive form of celebrity worship based on virtual reality and mixed reality • In Mumbai, a teenage girl rebels when AI’s crunching of big data gets in the way of romance • In Seoul, virtual companions with perfected natural language processing (NLP) skills offer orphaned twins new ways to connect • In Munich, a rogue scientist draws on quantum computing, computer vision and other AI technologies in a revenge plot that imperils the world By gazing toward a not-so-distant horizon, AI 2041 offers urgent insights into our collective future—while reminding readers that, ultimately, humankind remains the author of its destiny.

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 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

Concepts, Methodologies and Techniques

A User’s Guide