Principles Of Data Integration Author Alon Halevy Jul 2012

Data as a Service shows how organizations can
leverage "data as a service" by providing real-life
case studies on the various and innovative
architectures and related patterns Comprehensive
approach to introducing data as a service in any
organization A reusable and flexible SOA based
architecture framework Roadmap to introduce 'big
data as a service' for potential clients Presents a
thorough description of each component in the DaaS
reference architecture so readers can implement
solutions
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 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 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 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 Data Mining: Concepts and Techniques provides the concepts and techniques in processing gathered data or information, which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD). It focuses on the feasibility, usefulness, effectiveness, and scalability of techniques of large data sets. After describing data mining, this edition explains the methods of knowing, preprocessing, processing, and warehousing data. It then presents information about data warehouses, online analytical processing (OLAP), and data cube technology. Then, the methods involved in mining frequent patterns, associations, and correlations for large data sets are described. The book details the methods for data classification and introduces the concepts and methods for data clustering. The remaining chapters discuss the outlier detection and the trends, applications, and research frontiers in data mining. This book is intended for Computer Science students, application developers, business professionals, and researchers who seek information

on data mining. Presents dozens of algorithms and implementation examples, all in pseudo-code and suitable for use in real-world, large-scale data mining projects Addresses advanced topics such as mining object-relational databases, spatial databases, multimedia databases, time-series databases, text databases, the World Wide Web, and applications in several fields Provides a comprehensive, practical look at the concepts and techniques you need to get the most out of your data
Applying Serivce Oriented Architecture (SOA) Principles in Informatica is an unique attempt to map SOA Design Principles to Informatica product architecture. This book does not provide any samples, but attempts map the SOA concepts to Informatica's ETL and Data Integration Concepts. This book assumes basic / no understanding of SOA principles, however it assumes that the reader has a basic understanding of SOA itself This book takes simple SOA concepts and discusses at beginner level on how these can be applied in the Informatica world. NOTE: Purchasing this book does not entitle you to a free PowerCenter software license. Principles of Big Data helps readers avoid the common mistakes that endanger all Big Data projects. By stressing simple, fundamental concepts, this book teaches readers how to organize large volumes of complex data, and how to achieve data permanence when the content of the data is

constantly changing. General methods for data verification and validation, as specifically applied to Big Data resources, are stressed throughout the book. The book demonstrates how adept analysts can find relationships among data objects held in disparate Big Data resources, when the data objects are endowed with semantic support (i.e., organized in classes of uniquely identified data objects). Readers will learn how their data can be integrated with data from other resources, and how the data extracted from Big Data resources can be used for purposes beyond those imagined by the data creators. Learn general methods for specifying Big Data in a way that is understandable to humans and to computers Avoid the pitfalls in Big Data design and analysis Understand how to create and use Big Data safely and responsibly with a set of laws, regulations and ethical standards that apply to the acquisition, distribution and integration of Big Data resources

How do you approach answering queries when your data is stored in multiple databases that were designed independently by different people? This is first comprehensive book on data integration and is written by three of the most respected experts in the field. This book provides an extensive introduction to the theory and concepts underlying today's data integration techniques, with detailed, instruction for their application using concrete examples throughout

to explain the concepts. Data integration is the problem of answering queries that span multiple data sources (e.g., databases, web pages). Data integration problems surface in multiple contexts, including enterprise information integration, query processing on the Web, coordination between government agencies and collaboration between scientists. In some cases, data integration is the key bottleneck to making progress in a field. The authors provide a working knowledge of data integration concepts and techniques, giving you the tools you need to develop a complete and concise package of algorithms and applications. Offers a range of data integration solutions enabling you to focus on what is most relevant to the problem at hand Enables you to build your own algorithms and implement your own data integration applications

How do you approach answering queries when your data is stored in multiple databases that were designed independently by different people? This is first comprehensive book on data integration and is written by three of the most respected experts in the field. This book provides an extensive introduction to the theory and concepts underlying today's data integration techniques, with detailed, instruction for their application using concrete examples throughout to explain the concepts. Data integration is the problem of answering queries that span multiple data sources (e.g., databases, web pages). Data

integration problems surface in multiple contexts, including enterprise information integration, query processing on the Web, coordination between government agencies and collaboration between scientists. In some cases, data integration is the key bottleneck to making progress in a field. The authors provide a working knowledge of data integration concepts and techniques, giving you the tools you need to develop a complete and concise package of algorithms and applications. *Offers a range of data integration solutions enabling you to focus on what is most relevant to the problem at hand. *Enables you to build your own algorithms and implement your own data integration applications *Companion website with numerous project-based exercises and solutions and slides. Links to commercially available software allowing readers to build their own algorithms and implement their own data integration applications. Facebook page for reader input during and after publication.

The concept of a data lake is less than 10 years old, but they are already hugely implemented within large companies. Their goal is to efficiently deal with ever-growing volumes of heterogeneous data, while also facing various sophisticated user needs. However, defining and building a data lake is still a challenge, as no consensus has been reached so far. Data Lakes presents recent outcomes and trends in the field of data repositories. The main topics discussed

are the data-driven architecture of a data lake; the management of metadata – supplying key information about the stored data, master data and reference data; the roles of linked data and fog computing in a data lake ecosystem; and how gravity principles apply in the context of data lakes. A variety of case studies are also presented, thus providing the reader with practical examples of data lake management.

Developing High Quality Data Models provides an introduction to the key principles of data modeling. It explains the purpose of data models in both developing an Enterprise Architecture and in supporting Information Quality; common problems in data model development; and how to develop high quality data models, in particular conceptual, integration, and enterprise data models. The book is organized into four parts. Part 1 provides an overview of data models and data modeling including the basics of data model notation; types and uses of data models; and the place of data models in enterprise architecture. Part 2 introduces some general principles for data models, including principles for developing ontologically based data models; and applications of the principles for attributes, relationship types, and entity types. Part 3 presents an ontological framework for developing consistent data models. Part 4 provides the full data model that has been in development throughout the

book. The model was created using Jotne EPM Technologys EDMVisualExpress data modeling tool. This book was designed for all types of modelers: from those who understand data modeling basics but are just starting to learn about data modeling in practice, through to experienced data modelers seeking to expand their knowledge and skills and solve some of the more challenging problems of data modeling. Uses a number of common data model patterns to explain how to develop data models over a wide scope in a way that is consistent and of high quality Offers generic data model templates that are reusable in many applications and are fundamental for developing more specific templates Develops ideas for creating consistent approaches to high quality data models

Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between

business stakeholders and data scientists, but also
how participate intelligently in your company's data
science projects. You'll also discover how to think
data-analytically, and fully appreciate how data
science methods can support business decision-
making. Understand how data science fits in your
organization—and how you can use it for competitive
advantage Treat data as a business asset that
requires careful investment if you're to gain real
value Approach business problems data-analytically,
using the data-mining process to gather good data in
the most appropriate way Learn general concepts for
actually extracting knowledge from data Apply data
science principles when interviewing data science
job candidates
This book provides a systematic and comparative
description of the vast number of research issues
related to the quality of data and information. It does
so by delivering a sound, integrated and
comprehensive overview of the state of the art and
future development of data and information quality in
databases and information systems. To this end, it
presents an extensive description of the techniques
that constitute the core of data and information
quality research, including record linkage (also called
object identification), data integration, error
localization and correction, and examines the related
techniques in a comprehensive and original
methodological framework. Quality dimension

definitions and adopted models are also analyzed in detail, and differences between the proposed solutions are highlighted and discussed. Furthermore, while systematically describing data and information quality as an autonomous research area, paradigms and influences deriving from other areas, such as probability theory, statistical data analysis, data mining, knowledge representation, and machine learning are also included. Last not least, the book also highlights very practical solutions, such as methodologies, benchmarks for the most effective techniques, case studies, and examples. The book has been written primarily for researchers in the fields of databases and information management or in natural sciences who are interested in investigating properties of data and information that have an impact on the quality of experiments, processes and on real life. The material presented is also sufficiently self-contained for masters or PhD-level courses, and it covers all the fundamentals and topics without the need for other textbooks. Data and information system administrators and practitioners, who deal with systems exposed to data-quality issues and as a result need a systematization of the field and practical methods in the area, will also benefit from the combination of concrete practical approaches with sound theoretical formalisms.
In a world of fierce global competition and rapid

technologicalchange, traditional strategies for gaining market share andachieving efficiencies no longer yield the returns they once did.How can companies drive consumer preference and secure sustainablegrowth in this digital, social, and mobile age? The answer is through functional integration. Some of theworld's most highly valued companies—including Amazon, Appleand Google—have harnessed this new business model to buildhighly interactive ecosystems of interrelated products and digitalservices, gaining new levels of customer engagement. Functionalintegration offers forward-looking brands a unique competitive edgeby using transformative digital technologies to deliver high-valuecustomer experiences, generate repeat business, and unlocklucrative new business-to-business revenue streams. Connected By Design is the first book to show businessleaders and marketers exactly how to use functional integration toachieve transformative growth within any type of company. Based onR/GA's pioneering work with firms at the forefront of functionalintegration, Barry Wacksman and Chris Stutzman identify sevenprinciples companies must follow in order to create and deliver newvalue for customers and capture new revenues. Connected ByDesign explains how functional integration drove thetransformation of market-leading companies as diverse as Nike,General Motors, McCormick & Co.,

and Activision to establishauthentic brand relationships with their customers, enter newcategories, and develop new sources of income. With Connected byDesign, any company can leverage technological disruption toredefine its mission and foster greater brand loyalty andengagement.

Large-Scale 3D Data Integration: Challenges and Opportunities examines the fundamental aspects of 3D geo-information, focusing on the latest developments in 3D GIS (geographic information) and AEC (architecture, engineering, construction) systems. This book addresses policy makers, designers and engineers, and individuals that need to overco

Stable, predictive biomarkers and interpretable disease signatures are seen as a significant step towards personalized medicine. In this perspective, integration of multi-omic data coming from genomics, transcriptomics, glycomics, proteomics, metabolomics is a powerful strategy to reconstruct and analyse complex multi-dimensional interactions, enabling deeper mechanistic and medical insight. At the same time, there is a rising concern that much of such different omic data –although often publicly and freely available- lie in databases and repositories underutilised or not used at all. Issues coming from lack of standardisation and shared biological identities are also well-known. From these

considerations, a novel, pressing request arises from the life sciences to design methodologies and approaches that allow for these data to be interpreted as a whole, i.e. as intertwined molecular signatures containing genes, proteins, mRNAs and miRNAs, able to capture inter-layers connections and complexity. Papers discuss data integration approaches and methods of several types and extents, their application in understanding the pathogenesis of specific diseases or in identifying candidate biomarkers to exploit the full benefit of multi-omic datasets and their intrinsic information content. Topics of interest include, but are not limited to: • Methods for the integration of layered data, including, but not limited to, genomics, transcriptomics, glycomics, proteomics, metabolomics; • Application of multi-omic data integration approaches for diagnostic biomarker discovery in any field of the life sciences; • Innovative approaches for the analysis and the visualization of multi-omic datasets; • Methods and applications for systematic measurements from single/undivided samples (comprising genomic, transcriptomic, proteomic, metabolomic measurements, among others); • Multi-scale approaches for integrated dynamic modelling and simulation; • Implementation of applications, computational resources and repositories devoted to data integration including, but not limited to, data

warehousing, database federation, semantic integration, service-oriented and/or wiki integration; • Issues related to the definition and implementation of standards, shared identities and semantics, with particular focus on the integration problem. Research papers, reviews and short communications on all topics related to the above issues were welcomed.

Expertly written by top experts, this book provides an in-depth analysis of the CASE tool integration problem, and describes practical approaches that can be used with current CASE technology.

"This book presents cutting-edge research and analysis of the most recent advancements in the fields of database systems and software development"--Provided by publisher.

Managing Data in Motion describes techniques that have been developed for significantly reducing the complexity of managing system interfaces and enabling scalable architectures. Author April Reeve brings over two decades of experience to present a vendor-neutral approach to moving data between computing environments and systems. Readers will learn the techniques, technologies, and best practices for managing the passage of data between computer systems and integrating disparate data together in an enterprise environment. The average enterprise's computing environment is comprised of hundreds to thousands computer systems that have

been built, purchased, and acquired over time. The data from these various systems needs to be integrated for reporting and analysis, shared for business transaction processing, and converted from one format to another when old systems are replaced and new systems are acquired. The management of the "data in motion" in organizations is rapidly becoming one of the biggest concerns for business and IT management. Data warehousing and conversion, real-time data integration, and cloud and "big data" applications are just a few of the challenges facing organizations and businesses today. Managing Data in Motion tackles these and other topics in a style easily understood by business and IT managers as well as programmers and architects. Presents a vendor-neutral overview of the different technologies and techniques for moving data between computer systems including the emerging solutions for unstructured as well as structured data types Explains, in non-technical terms, the architecture and components required to perform data integration Describes how to reduce the complexity of managing system interfaces and enable a scalable data architecture that can handle the dimensions of "Big Data"
Principles of Data IntegrationElsevier
Between the high-level concepts of business intelligence and the nitty-gritty instructions for using vendors' tools lies the essential, yet poorly-

understood layer of architecture, design and process. Without this knowledge, Big Data is belittled – projects flounder, are late and go over budget. Business Intelligence Guidebook: From Data Integration to Analytics shines a bright light on an often neglected topic, arming you with the knowledge you need to design rock-solid business intelligence and data integration processes. Practicing consultant and adjunct BI professor Rick Sherman takes the guesswork out of creating systems that are cost-effective, reusable and essential for transforming raw data into valuable information for business decision-makers. After reading this book, you will be able to design the overall architecture for functioning business intelligence systems with the supporting data warehousing and data-integration applications. You will have the information you need to get a project launched, developed, managed and delivered on time and on budget – turning the deluge of data into actionable information that fuels business knowledge. Finally, you'll give your career a boost by demonstrating an essential knowledge that puts corporate BI projects on a fast-track to success. Provides practical guidelines for building successful BI, DW and data integration solutions. Explains underlying BI, DW and data integration design, architecture and processes in clear, accessible language. Includes the complete project

development lifecycle that can be applied at large enterprises as well as at small to medium-sized businesses Describes best practices and pragmatic approaches so readers can put them into action. Companion website includes templates and examples, further discussion of key topics, instructor materials, and references to trusted industry sources.

Adaptive Query Processing surveys the fundamental issues, techniques, costs, and benefits of adaptive query processing. It begins with a broad overview of the field, identifying the dimensions of adaptive techniques. It then looks at the spectrum of approaches available to adapt query execution at runtime - primarily in a non-streaming context. The emphasis is on simplifying and abstracting the key concepts of each technique, rather than reproducing the full details available in the papers. The authors identify the strengths and limitations of the different techniques, demonstrate when they are most useful, and suggest possible avenues of future research. Adaptive Query Processing serves as a valuable reference for students of databases, providing a thorough survey of the area. Database researchers will benefit from a more complete point of view, including a number of approaches which they may not have focused on within the scope of their own research.

This book is an informative reference, or working

textbook, on the mathematics, and general physical and chemical processes behind remote sensor measurements.

The practical application of structural geology in industry is varied and diverse; it is relevant at all scales, from plate-wide screening of new exploration areas down to fluid-flow behaviour along individual fractures. From an industry perspective, good structural practice is essential since it feeds into the quantification and recovery of reserves and ultimately underpins commercial investment choices. Many of the fundamental structural principles and techniques used by industry can be traced back to the academic community, and this volume aims to provide insights into how structural theory translates into industry practice. Papers in this publication describe case studies and workflows that demonstrate applied structural geology, covering a spread of topics including trap definition, fault seal, fold-and-thrust belts, fractured reservoirs, fluid flow and geomechanics. Against a background of evolving ideas, new data types and advancing computational tools, the volume highlights the need for structural geologists to constantly re-evaluate the role they play in solving industrial challenges.

Principles of Transaction Processing is a comprehensive guide to developing applications, designing systems, and evaluating engineering products. The book provides detailed discussions of

the internal workings of transaction processing systems, and it discusses how these systems work and how best to utilize them. It covers the architecture of Web Application Servers and transactional communication paradigms. The book is divided into 11 chapters, which cover the following: Overview of transaction processing application and system structure Software abstractions found in transaction processing systems Architecture of multitier applications and the functions of transactional middleware and database servers Queued transaction processing and its internals, with IBM's Websphere MQ and Oracle's Stream AQ as examples Business process management and its mechanisms Description of the two-phase locking function, B-tree locking and multigranularity locking used in SQL database systems and nested transaction locking System recovery and its failures Two-phase commit protocol Comparison between the tradeoffs of replicating servers versus replication resources Transactional middleware products and standards Future trends, such as cloud computing platforms, composing scalable systems using distributed computing components, the use of flash storage to replace disks and data streams from sensor devices as a source of transaction requests. The text meets the needs of systems professionals, such as IT application programmers who construct TP applications, application analysts, and product

developers. The book will also be invaluable to students and novices in application programming. Complete revision of the classic "non mathematical" transaction processing reference for systems professionals. Updated to focus on the needs of transaction processing via the Internet-- the main focus of business data processing investments, via web application servers, SOA, and important new TP standards. Retains the practical, non-mathematical, but thorough conceptual basis of the first edition. A brand new textbook with an innovative and exciting approach to marketing strategy. Moving away from the outdated 4Ps model to a new approach that reflects real-world companies responding to a differing and dynamic customer base. Research-based and action-orientated, it equips students with the tools to succeed in today's competitive markets. Introductory, theory-practice balanced text teaching the fundamentals of databases to advanced undergraduates or graduate students in information systems or computer science. Useful attributes capture and quantify key components of the seismic amplitude and texture for subsequent integration with well log, microseismic, and production data through either interactive visualization or machine learning. Although both approaches can accelerate and facilitate the interpretation process, they can by no means

replace the interpreter. Interpreter "grayware" includes the incorporation and validation of depositional, diagenetic, and tectonic deformation models, the integration of rock physics systematics, and the recognition of unanticipated opportunities and hazards. This book is written to accompany and complement the 2018 SEG Distinguished Instructor Short Course that provides a rapid overview of how 3D seismic attributes provide a framework for data integration over the life of the oil and gas field. Key concepts are illustrated by example, showing modern workflows based on interactive interpretation and display as well as those aided by machine learning.
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, 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.

Designing a complete visualization system involves many subtle decisions. When designing a complex, real-world visualization system, such decisions involve many types of constraints, such as performance, platform (in)dependence, available programming languages and styles, user-interface toolkits, input/output data format constraints, integration with third-party code, and more. Focusing on those techniques and methods with the broadest applicability across fields, the second edition of Data Visualization: Principles and Practice provides a streamlined introduction to various visualization techniques. The book illustrates a wide variety of

applications of data visualizations, illustrating the range of problems that can be tackled by such methods, and emphasizes the strong connections between visualization and related disciplines such as imaging and computer graphics. It covers a wide range of sub-topics in data visualization: data representation; visualization of scalar, vector, tensor, and volumetric data; image processing and domain modeling techniques; and information visualization. See What's New in the Second Edition: Additional visualization algorithms and techniques New examples of combined techniques for diffusion tensor imaging (DTI) visualization, illustrative fiber track rendering, and fiber bundling techniques Additional techniques for point-cloud reconstruction Additional advanced image segmentation algorithms Several important software systems and libraries Algorithmic and software design issues are illustrated throughout by (pseudo)code fragments written in the C++ programming language. Exercises covering the topics discussed in the book, as well as datasets and source code, are also provided as additional online resources.
This book discusses the evolution of future-generation technologies through the Internet of things, bringing together all the related technologies on a single platform to offer valuable insights for undergraduate and postgraduate students, researchers, academics and industry practitioners.

The book uses data, network engineering and intelligent decision- support system-by-design principles to design a reliable IoT-enabled ecosystem and to implement cyber-physical pervasive infrastructure solutions. It takes readers on a journey that begins with understanding the insight paradigm of IoT-enabled technologies and how it can be applied. It walks readers through engaging with real-time challenges and building a safe infrastructure for IoT-based, future-generation technologies. The book helps researchers and practitioners to understand the design architecture through IoT and the state of the art in IoT countermeasures. It also highlights the differences between heterogeneous platforms in IoT-enabled infrastructure and traditional ad hoc or infrastructural networks, and provides a comprehensive discussion on functional frameworks for IoT, object identification, IoT domain model, RFID technology, wearable sensors, WBAN, IoT semantics, knowledge extraction, and security and privacy issues in IoT-based ecosystems. Written by leading international experts, it explores IoT-enabled insight paradigms, which are utilized for the future benefit of humans. It also includes references to numerous works. Divided into stand-alone chapters, this highly readable book is intended for specialists, researchers, graduate students, designers, experts, and engineers involved in research on healthcare-

related issues.

Forecasting is required in many situations. Stocking an inventory may require forecasts of demand months in advance. Telecommunication routing requires traffic forecasts a few minutes ahead. Whatever the circumstances or time horizons involved, forecasting is an important aid in effective and efficient planning. This textbook provides a comprehensive introduction to forecasting methods and presents enough information about each method for readers to use them sensibly.

This third edition of a classic textbook can be used to teach at the senior undergraduate and graduate levels. The material concentrates on fundamental theories as well as techniques and algorithms. The advent of the Internet and the World Wide Web, and, more recently, the emergence of cloud computing and streaming data applications, has forced a renewal of interest in distributed and parallel data management, while, at the same time, requiring a rethinking of some of the traditional techniques. This book covers the breadth and depth of this re-emerging field. The coverage consists of two parts. The first part discusses the fundamental principles of distributed data management and includes distribution design, data integration, distributed query processing and optimization, distributed transaction management, and replication. The second part focuses on more advanced topics and includes

discussion of parallel database systems, distributed object management, peer-to-peer data management, web data management, data stream systems, and cloud computing. New in this Edition: • New chapters, covering database replication, database integration, multidatabase query processing, peer-to-peer data management, and web data management. • Coverage of emerging topics such as data streams and cloud computing • Extensive revisions and updates based on years of class testing and feedback Ancillary teaching materials are available. Data Stewardship for Open Science: Implementing FAIR Principles has been written with the intention of making scientists, funders, and innovators in all disciplines and stages of their professional activities broadly aware of the need, complexity, and challenges associated with open science, modern science communication, and data stewardship. The FAIR principles are used as a guide throughout the text, and this book should leave experimentalists consciously incompetent about data stewardship and motivated to respect data stewards as representatives of a new profession, while possibly motivating others to consider a career in the field. The ebook, avalable for no additional cost when you buy the paperback, will be updated every 6 months on average (providing that significant updates are needed or avaialble). Readers will have the opportunity to contribute material towards these

updates, and to develop their own data management plans, via the free Data Stewardship Wizard. Making Data Integration Work: How to Systematically Reduce Cost, Improve Quality, and Enhance Effectiveness Today's enterprises are investing massive resources in data integration. Many possess thousands of point-to-point data integration applications that are costly, undocumented, and difficult to maintain. Data integration now accounts for a major part of the expense and risk of typical data warehousing and business intelligence projects--and, as businesses increasingly rely on analytics, the need for a blueprint for data integration is increasing now more than ever. This book presents the solution: a clear, consistent approach to defining, designing, and building data integration components to reduce cost, simplify management, enhance quality, and improve effectiveness. Leading IBM data management expert Tony Giordano brings together best practices for architecture, design, and methodology, and shows how to do the disciplined work of getting data integration right. Mr. Giordano begins with an overview of the "patterns" of data integration, showing how to build blueprints that smoothly handle both operational and analytic data integration. Next, he walks through the entire project lifecycle, explaining each phase, activity, task, and deliverable through a complete case study. Finally, he shows

how to integrate data integration with other information management disciplines, from data governance to metadata. The book's appendices bring together key principles, detailed models, and a complete data integration glossary. Coverage includes Implementing repeatable, efficient, and well-documented processes for integrating data Lowering costs and improving quality by eliminating unnecessary or duplicative data integrations Managing the high levels of complexity associated with integrating business and technical data Using intuitive graphical design techniques for more effective process and data integration modeling Building end-to-end data integration applications that bring together many complex data sources This comprehensive textbook teaches the fundamentals of database design, modeling, systems, data storage, and the evolving world of data warehousing, governance and more. Written by experienced educators and experts in big data, analytics, data quality, and data integration, it provides an up-to-date approach to database management. This full-color, illustrated text has a balanced theory-practice focus, covering essential topics, from established database technologies to recent trends, like Big Data, NoSQL, and more. Fundamental concepts are supported by real-world examples, query and code walkthroughs, and figures, making it perfect for introductory courses for

advanced undergraduates and graduate students in information systems or computer science. These examples are further supported by an online playground with multiple learning environments, including MySQL; MongoDB; Neo4j Cypher; and tree structure visualization. This combined learning approach connects key concepts throughout the text to the important, practical tools to get started in database management.

This book reviews and summarizes the development and achievement in cartography and geographic information engineering in China over the past 60 years after the founding of the People's Republic of China. It comprehensively reflects cartography, as a traditional discipline, has almost the same long history with the world's first culture and has experienced extraordinary and great changes. The book consists of nineteen thematic chapters. Each chapter is in accordance with the unified directory structure, introduction, development process, major study achievements, problem and prospect, representative works, as well as a lot of references. It is useful as a reference both for scientists and technicians who are engaged in teaching, researching and engineering of cartography and geographic information engineering.

This book constitutes revised selected papers from the 13th International Conference on Data Integration in the Life Sciences, DILS 2018, held in

Hannover, Germany, in November 2018. The 5 full, 8 short, 3 poster and 4 demo papers presented in this volume were carefully reviewed and selected from 22 submissions. The papers are organized in topical sections named: big biomedical data integration and management; data exploration in the life sciences; biomedical data analytics; and big biomedical applications.

As data management and integration continue to evolve rapidly, storing all your data in one place, such as a data warehouse, is no longer scalable. In the very near future, data will need to be distributed and available for several technological solutions. With this practical book, you'll learnhow to migrate your enterprise from a complex and tightly coupled data landscape to a more flexible architecture ready for the modern world of data consumption. Executives, data architects, analytics teams, and compliance and governance staff will learn how to build a modern scalable data landscape using the Scaled Architecture, which you can introduce incrementally without a large upfront investment. Author Piethein Strengholt provides blueprints, principles, observations, best practices, and patterns to get you up to speed. Examine data management trends, including technological developments, regulatory requirements, and privacy concerns Go deep into the Scaled Architecture and learn how the pieces fit together Explore data governance and data

security, master data management, self-service data marketplaces, and the importance of metadata Migrate your data to Salesforce and build low-maintenance and high-performing data integrations to get the most out of Salesforce and make it a "go-to" place for all your organization's customer information. When companies choose to roll out Salesforce, users expect it to be the place to find any and all Information related to a customer—the coveted Client 360° view. On the day you go live, users expect to see all their accounts, contacts, and historical data in the system. They also expect that data entered in other systems will be exposed in Salesforce automatically and in a timely manner. This book shows you how to migrate all your legacy data to Salesforce and then design integrations to your organization's mission-critical systems. As the Salesforce platform grows more powerful, it also grows in complexity. Whether you are migrating data to Salesforce, or integrating with Salesforce, it is important to understand how these complexities need to be reflected in your design. Developing Data Migrations and Integrations with Salesforce covers everything you need to know to migrate your data to Salesforce the right way, and how to design low-maintenance, high-performing data integrations with Salesforce. This book is written by a practicing Salesforce integration architect with dozens of Salesforce projects under his belt. The patterns and

practices covered in this book are the results of the
lessons learned during those projects. What You'll
Learn Know how Salesforce's data engine is
architected and why Use the Salesforce Data APIs
to load and extract data Plan and execute your data
migration to Salesforce Design low-maintenance,
high-performing data integrations with Salesforce
Understand common data integration patterns and
the pros and cons of each Know real-time integration
options for Salesforce Be aware of common pitfalls
Build reusable transformation code covering
commonly needed Salesforce transformation
patterns Who This Book Is For Those tasked with
migrating data to Salesforce or building ongoing data
integrations with Salesforce, regardless of the ETL
tool or middleware chosen; project sponsors or
managers nervous about data tracks putting their
projects at risk; aspiring Salesforce integration
and/or migration specialists; Salesforce developers
or architects looking to expand their skills and take
on new challenges
The Principles of Integrated Technology in Avionics
Systems describes how integration can improve
flight operations, enhance system processing
efficiency and equip resource integration. The title
provides systematic coverage of avionics system
architecture and ground system integration. Looking
beyond hardware resource sharing alone, it guides
the reader through the benefits and scope of a

modern integrated avionics system. Integrated technology enhances the performance of organizations by improving system capacity and boosting efficiency. Avionics systems are the functional center of aircraft systems. System integration technology plays a vital role in the complex world of avionics and an integrated avionics system will fully-address systems, information and processes. Introduces integration technology in complex avionics systems Guides the reader through the scope and benefits of avionic system integration Gives practical guidance on using integration to optimize an avionics system Describes the basis of avionics system architecture and ground system integration Presents modern avionics as a system that is becoming increasingly integrated

Copyright: 43239766ac203ae6702b06be3af60c06