Enterprise Data Architecture How To Navigate Its Landscape

Enterprise Architecture (EA) is an essential part of the fabric of a business; however, EA also transcends and transforms technology and moves it into the business space. Therefore, EA needs to be discussed in an integrated, holistic, and comprehensive manner. Only such an integrated approach to EA can provide the foundation for a transformation that readies the business for the myriad enterprise-wide challenges it will face. Highly disruptive technologies such as Big Data, Machine Learning, and Mobile and Cloud Computing require a fine balance between their business and technical aspects as an organization moves forward with its digital transformation. This book focuses on preparing all organizations – large and small – and those wishing to move into them for the impact of leveraging these emerging, disruptive, and innovative technologies within the EA framework.

"The book addresses a sorely missing set of considerations in the real world... This is a very timely book." -Peter Herzum, author of Business Component Factory and CEO of Herzum Software XML is a tremendous enabler for platform agnostic data and metadata exchanges. However, there are no clear processes and techniques specifically focused on the engineering of XML structures to support reuse and integration simplicity, which are of particular importance in the age of application integration and Web services. This book describes the challenges of using XML in a manner that promotes simplification of integration, and a high degree of schema reuse. It also describes the syntactical capabilities of XML and XML Schemas, and the similarities (and in some cases limitations) of XML DTDs. This book presents combinations of architectural and design approaches to using XML as well as numerous syntactical and working examples. * Designed to be read three different ways: skim the margin notes for quick information, or use tables in the appendix to locate sections relevant the to a particular issue, or read cover-to-cover for the in-depth treatment. * Contains numerous tables that describe datatypes supported by the most common DBMSs and map to XML Schema supported data types. * Unique focus on the value added role and processes of the data architect as they apply to enterprise use of XML.

Data has never been more important to your success than it is today, yet you are surrounded with data you can't trust, and the overwhelming burden of fixing it. Everyone deserves data that helps-not hurts-their organization.

Master a proven approach to create, implement, and sustain a data strategy. Pervasive, data is a unique organizational resource, and this distinction warrants its own strategy. Data, representing your single non-depletable, non-degradable, durable strategic asset, is likely also your most poorly leveraged and underutilized organizational asset. Lack of talent, barriers in organizational thinking, and seven specific data sins prevent most organizations from benefiting fully from their data asset investments. Solving these prerequisites will allow your organization to: Improve your organization’s data; Improve the way your people use data; and Improve the way your people use data to achieve your organizational strategy. This method better focuses data and thinking in direct support of strategic objectives. After eliminating necessary prerequisites, organizations can develop a disciplined and repeatable means of improving their data, literacy, standards, and controls using data governance practices. Once in place, the process (based on the theory of constraints) becomes a variant of lather, rinse, and repeat. Several complementary concepts covered include: An overview of data strategy prerequisites; A repeatable process for identifying and removing data constraints; Why data strategy is necessary for effective data governance; Balancing operational results with capability development; An objective definition of data-centric thinking; and Ways to monetize these efforts.

A Dell Technologies perspective on today's data landscape and the key ingredients for planning a modern, distributed data pipeline for your multicloud data-driven enterprise

While business functions such as manufacturing, operations, and marketing often utilize various software applications, they tend to operate without the ability to interact with each other and exchange data. This provides a challenge to gain an enterprise-wide view of a business and to assist real-time decision making. Service-Driven Approaches to Architecture and Enterprise Integration addresses the issues of integrating assorted software applications and systems by using a service driven approach. Supporting the dynamics of business needs, this book highlights the tools, techniques, and governance aspects of design, and implements cost-effective enterprise integration solutions. It is a valuable source of information for software architects, SOA practitioners, and software engineers as well as researchers and students in pursuit of extensible and agile software design.

Use this practical guide to successfully handle the challenges encountered when designing an enterprise data lake and learn industry best practices to resolve issues. When designing an enterprise data lake you often hit a roadblock when you must leave the comfort of the relational world and learn the nuances of handling non-relational data. Starting from sourcing data into the Hadoop ecosystem, you will go through stages that can bring up tough questions such as data processing, data querying, and security. Concepts such as change data capture and data streaming are covered. The book takes an end-to-end solution approach in a data lake environment that includes data security, high availability, data processing, data streaming, and more. Each chapter includes application of a concept, code snippets, and use case demonstrations to provide you with a practical approach. You will learn the concept, scope, application, and starting point. What You'll Learn Get to know data lake architecture and design principles Implement data capture and streaming strategies Implement data processing strategies in Hadoop Understand the data lake security framework and availability model Who This Book Is For Big data architects and solution architects

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"

Enterprise data is growing at a much faster rate than traditional technologies allow. New enterprise architectures combining existing technologies are desperately needed. This book suggests a way forward by applying new techniques of the World Wide Web to enterprise information systems. Linking Enterprise Data is an edited volume contributed by worldwide leaders in Semantic Web and Linked Data research, standards development and adoption. Linking enterprise data is the application of World Wide Web architecture principles to real-world information management issues faced by commercial, not-for-profit and government enterprises. This book is divided into four sections: Benefits of applying Linked Data principles in enterprise settings, enterprise approval and support of Linked Data projects, specific Linked Data techniques and a number of real-world success stories from early enterprise adopters. Linking Enterprise Data targets professionals working as CTOs, CIOs, enterprise architects, project managers and application developers in commercial, not-for-profit and government organizations concerned with scalability, flexibility and robustness of information management systems. Computer science graduate students and researchers focusing on enterprise information integration will also benefit.

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

The practice of enterprise application development has benefited from the emergence of many new enabling technologies. Multi-tiered object-oriented platforms, such as Java and .NET, have become commonplace. These new tools and technologies are capable of building powerful applications, but they are not easily implemented. Common failures in enterprise applications often occur because their developers do not understand the architectural lessons that experienced object developers have learned. Patterns of Enterprise Application Architecture is written in direct response to the stiff challenges that face enterprise application developers. The author, noted object-oriented designer Martin Fowler, noticed that despite changes in technology--from Smalltalk to CORBA to Java to .NET--the same basic design ideas can be adapted and applied to solve common problems. With the help of an expert group of contributors, Martin distills over forty recurring solutions into patterns. The result is an indispensable handbook of solutions that are applicable to any enterprise application platform. This book is actually two books in one. The first section is a short tutorial on developing enterprise applications, which you can read from start to finish to understand the scope of the book's lessons. The next section, the bulk of the book, is a detailed reference to the patterns themselves. Each pattern provides usage and implementation information, as well as detailed code examples in Java or C#. The entire book is also richly illustrated with UML diagrams to further explain the concepts. Armed with this book, you will have the knowledge necessary to make important architectural decisions about building an enterprise application and the proven patterns for use when building them. The topics covered include · Dividing an enterprise application into layers · The major approaches to organizing business logic · An in-depth treatment of mapping between objects and relational databases · Using Model-View-Controller to organize a Web presentation · Handling concurrency for data that spans multiple transactions · Designing distributed object interfaces

This book presents unique techniques to conquer different Big Data processing and analytics challenges using Hadoop. Practical examples are provided to boost your understanding of Big Data concepts and their implementation. By the end of the book, you will have all the knowledge and skills you need to become a true Big Data expert.

Defining a set of guiding principles for data management and describing how these principles can be applied within data management functional areas; Providing a functional framework for the implementation of enterprise data management practices; including widely adopted practices, methods and techniques, functions, roles, deliverables and metrics; Establishing a common vocabulary for data management concepts and serving as the basis for best practices for data management professionals. DAMA-DMBOK2 provides data management and IT professionals, executives, knowledge workers, educators, and researchers with a framework to manage their data and mature their information infrastructure, based on these principles: Data is an asset with unique properties; The value of data can be and should be expressed in economic terms; Managing data means managing the quality of data; It takes metadata to manage data; It takes planning to manage data; Data management is cross-functional and requires a range of skills and expertise; Data management requires an enterprise perspective; Data management must account for a range of perspectives; Data management is data lifecycle management; Different types of data have different lifecycle requirements; Managing data includes managing risks associated with data; Data management requirements must drive information technology decisions; Effective data management requires leadership commitment.

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

"Regardless of the type of architecture, architecture itself is an organized accumulation of knowledge within a particular domain. While we generally conceive of its representation as a set of diagrams, containing specific notations and taxonomies of symbols and glossary terms, an architecture may actually be represented using anything that can be arranged in a pattern to record information. The earliest forms of architecture relate to architecting buildings, monuments, military disciplines, organized religion, music, storytelling, and various other forms within the arts. These early forms of architecture of course predate computer related architectures by thousands of years. That said, it is worth noting that there are a number of common elements among architectures irrespective of their relative age, such as forms of standardization, reusable structures, the accumulation of knowledge, and providing a context for understanding something. Needless to say, anyone can be an architect in a topic in which they have a deep understanding and appreciation of. While one obvious difference among architects is the amount and variety of pertinent experience, the less obvious difference is the degree to which an architect recognizes the potential forms of standardization, reusable structures, accumulation of knowledge, relationships among the components, and use of architecture as an accelerator to more rapidly understand the context and scope of a particular topic or to rapidly convey it to another. Architectures as a result must be easy to understand"--

Wouldn't it be great to understand all the data in your organisation? Just imagine being able to define, agree and manage information concepts that impact on business strategy? Then image that these information concepts can be linked to the physical database attributes that ultimately are used to create them. That's what this book is about. It focuses on the data model as the foundation for achieving this understanding. This book provides a framework for the enterprise data model, the business reasons behind it and the differences between conceptual, logical and physical data models. The question of how,

and why, to use a data model artifact as part of the data governance toolkit for the whole enterprise is also addressed. This publication is not an in-depth manual on how to model data for a new database system or your next design project. It instead focuses at a level above these implementation projects and addresses the issues that organisations typical struggling with such as: * How do we provide a framework within which we can manage our data assets? * How do we develop applications that adhere to a set of data standards; without creating a nightmare of administration and governance that is both unwieldy and unusable? * How can we get business value from our enterprise data? Chapter headings are: * Chapter 1 - Introduction * Chapter 2 - Information and Data * Chapter 3 - Pillars of Value * Chapter 4 - An Overview of Data Modelling * Chapter 5 - Data Architecture * Chapter 6 - The Enterprise Data Model * Chapter 7 - Build the Model one Project at a Time * Chapter 8 - Master Data * Chapter 9 - Data Governance * Chapter 10 - The Enterprise Data Framework This 2nd edition revises the original text to add extra details around key areas such as the enterprise data model framework and the pillars of value. It also improves the quality of the original text.

All the answers to your data science questions Over half of all businesses are using data science to generate insights and value from big data. How are they doing it? Data Science Strategy For Dummies answers all your questions about how to build a data science capability from scratch, starting with the “what” and the “why” of data science and covering what it takes to lead and nurture a top-notch team of data scientists. With this book, you’ll learn how to incorporate data science as a strategic function into any business, large or small. Find solutions to your real-life challenges as you uncover the stories and value hidden within data. Learn exactly what data science is and why it’s important Adopt a data-driven mindset as the foundation to success Understand the processes and common roadblocks behind data science Keep your data science program focused on generating business value Nurture a top-quality data science team In non-technical language, Data Science Strategy For Dummies outlines new perspectives and strategies to effectively lead analytics and data science functions to create real value.

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.

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

There’s a lot of information about big data technologies, but splicing these technologies into an end-to-end enterprise data platform is a daunting task not widely covered. With this practical book, you’ll learn how to build big data infrastructure both on-premises and in the cloud and successfully architect a modern data platform. Ideal for enterprise architects, IT managers, application architects, and data engineers, this book shows you how to overcome the many challenges that emerge during Hadoop projects. You’ll explore the vast landscape of tools available in the Hadoop and big data realm in a thorough technical primer before diving into: Infrastructure: Look at all component layers in a modern data platform, from the server to the data center, to establish a solid foundation for data in your enterprise Platform: Understand aspects of deployment, operation, security, high availability, and disaster recovery, along with everything you need to know to integrate your platform with the rest of your enterprise IT Taking Hadoop to the cloud: Learn the important architectural aspects of running a big data platform in the cloud while maintaining enterprise security and high availability

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

processes, and technology together with MDM and data governance Best practices in MDM deployment, including data warehouse and SAP integration

Building upon his earlier book that detailed agile data warehousing programming techniques for the Scrum master, Ralph's latest work illustrates the agile interpretations of the remaining software engineering disciplines: Requirements management benefits from streamlined templates that not only define projects quickly, but ensure nothing essential is overlooked. Data engineering receives two new "hyper modeling" techniques, yielding data warehouses that can be easily adapted when requirements change without having to invest in ruinously expensive data-conversion programs. Quality assurance advances with not only a stereoscopic top-down and bottom-up planning method, but also the incorporation of the latest in automated test engines. Use this step-by-step guide to deepen your own application development skills through self-study, show your teammates the world's fastest and most reliable techniques for creating business intelligence systems, or ensure that the IT department working for you is building your next decision support system the right way. Learn how to quickly define scope and architecture before programming starts Includes techniques of process and data engineering that enable iterative and incremental delivery Demonstrates how to plan and execute quality assurance plans and includes a guide to continuous integration and automated regression testing Presents program management strategies for coordinating multiple agile data mart projects so that over time an enterprise data warehouse emerges Use the provided 120-day road map to establish a robust, agile data warehousing program

Driven by the need and desire to reduce costs, organizations are faced with a set of decisions that require analytical scrutiny. Enterprise Architecture A to Z: Frameworks, Business Process Modeling, SOA, and Infrastructure Technology examines cost-saving trends in architecture planning, administration, and management. To establish a framework for discussion, this book begins by evaluating the role of Enterprise Architecture Planning and Service-Oriented Architecture (SOA) modeling. It provides an extensive review of the most widely deployed architecture framework models. In particular, the book discusses The Open Group Architecture Framework (TOGAF) and the Zachman Architectural Framework (ZAF) in detail, as well as formal architecture standards and all four layers of these models: the business architecture, the information architecture, the solution architecture, and the technology architecture. The first part of the text focuses on the upper layers of the architecture framework, while the second part focuses on the technology architecture. In this second section, the author presents an assessment of storage technologies and networking and addresses regulatory and security issues. Additional coverage includes high-speed communication mechanisms such as Ethernet, WAN and Internet communication technologies, broadband communications, and chargeback models. Daniel Minoli has written a number of columns and books on the high-tech industry and has many years of technical hands-on and managerial experience at top financial companies and telecom/networking providers. He brings a wealth of knowledge and practical experience to these pages. By reviewing the strategies in this book, CIOs, CTOs, and senior managers are empowered by a set of progressive approaches to designing state-of-the-art IT data centers.

Adopting the latest technological and data related innovations has caused many organisations to realise they don’t have a firm grasp on their basic operational data. This is a problem that Logical Data Models are uniquely qualified to help them solve. The realisation of the need to define a Logical Data Model may be driven by any number of reasons including; trying to link Big Data Analytics to operational data, plunging into Digital Marketing, choosing the best SaaS solution, carrying out a core Data Migration, developing a Data Warehouse, enhancing Data Governance processes, or even just trying to get everyone to agree on their Product specifications! This book will provide you with the skills required to start to answer these and many similar types of questions. It is not written with a focus on IT development, so you don’t need a technical background to get the most from it. But for any professional working in an organisation’s data landscape, this book will provide the skills they need to define high quality and beneficial data models quickly and easily. It does this using a wealth of practical examples, tips and techniques, as well as providing checklists and templates. It is structured into three parts: The Foundations: What are the solid foundations necessary for building effective data models? The Tools: What Tools are required to enable you to specify clear, precise and accurate data model definitions? The Deliverables: What processes will you need to successfully define the models, what will they deliver, and how can we make them beneficial to the organisation? “In this data-rich era, it is even more critical for organisations to answer the question of what their data means and the value it can bring. Those who can, will gain a competitive advantage through their use of data to streamline their operations and energise their strategies. Core to revealing this meaning, is the data model that is now, more than ever, the lynchpin of success. The Data Model Toolkit provides the essential knowledge and skills that will ensure this success.” – Reem Zahran, Global IT Platform Director, TNS “We work with many enterprise customers to help them transform their technology and it always starts with data. The key is a clear definition of their data quality, completeness and governance. This book shows you step by step how to define and use Data Models as powerful tools to define an organisation’s data and maximise its business benefit.” – John Casserly, CEO, Xceed Group

Enterprise Information Architecture for a New Age: Big Data and The Internet of Things, provides guidance in designing an information architecture to accommodate increasingly large amounts of data, massively large amounts of data, not only from traditional sources, but also from novel sources such everyday objects that are fast becoming wired into global Internet. No business can afford to be caught out by missing the value to be mined from the increasingly large amounts of available data generated by everyday devices. The text provides background as to how analytical solutions and enterprise architecture methodologies and concepts have evolved (including the roles of data warehouses, business intelligence tools, predictive analytics, data discovery, Big Data, and the impact of the Internet of Things). Then you’re taken through a series of steps by which to define a future state architecture and create a plan for how to reach that future state. Enterprise Information Architecture for a New Age: Big Data and The Internet of Things helps you gain an understanding of the following: Implications of Big Data from a variety of new data sources (including data from sensors that are part of the Internet of Things) upon an information architecture How establishing a vision for data usage by defining a roadmap that aligns IT with line-of-business needs is a key early step The importance and details of taking a step-by-step approach when dealing with shifting business challenges and changing technology capabilities How to mitigate risk when evaluating existing infrastructure and designing and deploying new infrastructure Enterprise Information Architecture for a New Age: Big Data and The Internet of Things combines practical advice with technical considerations. Author Robert Stackowiak and his team are recognized worldwide for their expertise in large data solutions, including analytics. Don’t miss your chance to read this book and gain the benefit of their advice as you look forward in thinking through your own choices and designing your own architecture to accommodate the burgeoning explosion in data that can be analyzed and converted into valuable information to drive your business forward toward success.

Enterprise Data Architecture: How to navigate its landscapeParagon Publishing

Data Architecture: From Zen to Reality explains the principles underlying data architecture, how data evolves with organizations, and the challenges organizations face in structuring and managing their data.

Using a holistic approach to the field of data architecture, the book describes proven methods and technologies to solve the complex issues dealing with data. It covers the various applied areas of data, including data modelling and data model management, data quality, data governance, enterprise information management, database design, data warehousing, and warehouse design. This text is a core resource for anyone customizing or aligning data management systems, taking the Zen-like idea of data architecture to an attainable reality. The book presents fundamental concepts of enterprise architecture with definitions and real-world applications and scenarios. It teaches data managers and planners about the challenges of building a data architecture roadmap, structuring the right team, and building a long term set of solutions. It includes the detail needed to illustrate how the fundamental principles are used in current business practice. The book is divided into five sections, one of which addresses the software-application development process, defining tools, techniques, and methods that ensure repeatable results. Data Architecture is intended for people in business management involved with corporate data issues and information technology decisions, ranging from data architects to IT consultants, IT auditors, and data administrators. It is also an ideal reference tool for those in a higher-level education process involved in data or information technology management. Presents fundamental concepts of enterprise architecture with definitions and real-world applications and scenarios Teaches data managers and planners about the challenges of building a data architecture roadmap, structuring the right team, and building a long term set of solutions Includes the detail needed to illustrate how the fundamental principles are used in current business practice

The data lakehouse is the next generation of the data warehouse and data lake, designed to meet today's complex and ever-changing analytics, machine learning, and data science requirements. Learn about the features and architecture of the data lakehouse, along with its powerful analytical infrastructure. Appreciate how the universal common connector blends structured, textual, analog, and IoT data. Maintain the lakehouse for future generations through Data Lakehouse Housekeeping and Data Future-proofing. Know how to incorporate the lakehouse into an existing data governance strategy. Incorporate data catalogs, data lineage tools, and open source software into your architecture to ensure your data scientists, analysts, and end users live happily ever after.

Over the past 5 years, the concept of big data has matured, data science has grown exponentially, and data architecture has become a standard part of organizational decision-making. Throughout all this change, the basic principles that shape the architecture of data have remained the same. There remains a need for people to take a look at the "bigger picture" and to understand where their data fit into the grand scheme of things. Data Architecture: A Primer for the Data Scientist, Second Edition addresses the larger architectural picture of how big data fits within the existing information infrastructure or data warehousing systems. This is an essential topic not only for data scientists, analysts, and managers but also for researchers and engineers who increasingly need to deal with large and complex sets of data. Until data are gathered and can be placed into an existing framework or architecture, they cannot be used to their full potential. Drawing upon years of practical experience and using numerous examples and case studies from across various industries, the authors seek to explain this larger picture into which big data fits, giving data scientists the necessary context for how pieces of the puzzle should fit together. New case studies include expanded coverage of textual management and analytics New chapters on visualization and big data Discussion of new visualizations of the end-state architecture

Are you looking to make better use of data captured within your organisation or want to learn more about how Data Architecture can transform your operations? Answering these questions is at the very heart of Navigating the Data Architecture Landscape. By reading this book you will learn how to: Introduce or improve the Data Architecture function of your organisation Enhance your skills in this domain to deliver more from your data. You may be wondering how a book can do this if it knows nothing about where you are now, or where you want to be? It can, because by leveraging its principles you will discover how to create optimised potential routes to achieve your own Data Architectural objectives. Basic building blocks, concepts and models are defined, enabling you to create new or adapt existing frameworks appropriate for any data landscape. Practical tips and suggestions are also detailed throughout, helping you gain immediate improvements from the way you work and enhance the benefits your organisation can derive from its data. So if you are a Data Architect or deal with data in your organisation and want to learn how to transform the positive yield from its data, then this book is a must read for you! “David has been there and dealt with the issues, which is why this book is an outstanding resource for Data Architects and indeed anyone dealing with the serious challenges of an enterprise data landscape.” – Richard Rendell, Technical Services Director, AgeSmart “An essential read for anyone wishing to practically achieve more benefit from data for their organisation within today’s constraints.” – Reem Zahran - Director, Offering Development, IMS Health “This book provides a comprehensive set of tools enabling you to improve the business outcomes from your organisation’s use of data.” – Andrew Rowland, Global Head Database Engineering, UBS This book is an essential read for Data Architects or indeed anyone wanting to improve the benefit that their organisation can derive from its data usage. It does this by providing principles and models that are appropriate to use within any framework, or even the absence of one. The book is designed to be practical and contains many tips and suggestions as well as examples that can be used as the basis for the reader's own Data Architectural definitions. The breadth of the book covers contemporary themes for Data Architecture and the chapters include; Data Modelling, Enterprise Data Models, Data Governance, Master Data Management and Big Data

Enterprise Architecture Planning (EAP) is a high-level blueprint for data, applications, and technology that is a cost-effective long-term solution. The authors give you a common-sense approach to EAP, supported by examples of architectures, procedures, checklists, and useful guidelines.

Cloud Enterprise Architecture examines enterprise architecture (EA) in the context of the surging popularity of Cloud computing. It explains the different kinds of desired transformations the architectural blocks of EA undergo in light of this strategically significant convergence. Chapters cover each of the contributing architectures of EA—business, information, application, integration, security, and technology—illustrating the current and impending implications of the Cloud on each. Discussing the implications of the Cloud paradigm on EA, the book details the perceptible and positive changes that will affect EA design, governance, strategy, management, and sustenance. The author ties these topics together with chapters on Cloud integration and composition architecture. He also examines the Enterprise Cloud, Federated Clouds, and the vision to establish the InterCloud. Laying out a comprehensive strategy for planning and executing Cloud-inspired transformations, the book: Explains how the Cloud changes and affects enterprise architecture design, governance, strategy, management, and sustenance Presents helpful information on next-generation Cloud computing Describes additional architectural types such as enterprise-scale integration, security, management, and governance architectures This book is an ideal resource for enterprise architects, Cloud evangelists and enthusiasts, and Cloud application and service architects. Cloud center administrators, Cloud business executives, managers, and analysts will also find the book helpful and inspirational while formulating appropriate mechanisms and schemes for sound modernization and migration of traditional applications to Cloud infrastructures and platforms.

A practical guide to implementing your enterprise data lake using Lambda Architecture as the base About This Book Build a full-fledged data lake for your organization with popular big data technologies using the Lambda architecture as the base Delve into the big data technologies required to meet modern day business strategies A highly practical guide to implementing enterprise data lakes with lots of examples and real-world use-cases Who This Book Is For Java developers and architects who would like to implement a data lake for their enterprise will find this book useful. If you want to get hands-on experience with the Lambda Architecture and big data technologies by implementing a practical solution using these technologies, this book will also help you. What You Will Learn Build an enterprise-level data lake using the relevant big data technologies Understand the core of the Lambda architecture and how to apply it in an enterprise Learn the technical details around Sqoop and its functionalities Integrate Kafka

with Hadoop components to acquire enterprise data Use flume with streaming technologies for stream-based processing Understand stream- based processing with reference to Apache Spark Streaming Incorporate Hadoop components and know the advantages they provide for enterprise data lakes Build fast, streaming, and high-performance applications using ElasticSearch Make your data ingestion process consistent across various data formats with configurability Process your data to derive intelligence using machine learning algorithms In Detail The term "Data Lake" has recently emerged as a prominent term in the big data industry. Data scientists can make use of it in deriving meaningful insights that can be used by businesses to redefine or transform the way they operate. Lambda architecture is also emerging as one of the very eminent patterns in the big data landscape, as it not only helps to derive useful information from historical data but also correlates real-time data to enable business to take critical decisions. This book tries to bring these two important aspects — data lake and lambda architecture—together. This book is divided into three main sections. The first introduces you to the concept of data lakes, the importance of data lakes in enterprises, and getting you up-to-speed with the Lambda architecture. The second section delves into the principal components of building a data lake using the Lambda architecture. It introduces you to popular big data technologies such as Apache Hadoop, Spark, Sqoop, Flume, and ElasticSearch. The third section is a highly practical demonstration of putting it all together, and shows you how an enterprise data lake can be implemented, along with several real-world use-cases. It also shows you how other peripheral components can be added to the lake to make it more efficient. By the end of this book, you will be able to choose the right big data technologies using the lambda architectural patterns to build your enterprise data lake. Style and approach The book takes a pragmatic approach, showing ways to leverage big data technologies and lambda architecture to build an enterprise-level data lake.

Tips, techniques, and trends on how to use dashboard technology tooptimize business performance Business performance management is a hot new management disciplinethat delivers tremendous value when supported by informationtechnology. Through case studies and industry research, this bookshows how leading companies are using performance dashboards toexecute strategy, optimize business processes, and improveperformance. Wayne W. Eckerson (Hingham, MA) is the Director of Research for TheData Warehousing Institute (TDWI), the leading association ofbusiness intelligence and data warehousing professionals worldwidethat provide high-quality, in-depth education, training, andresearch. He is a columnist for SearchCIO.com, DM Review,Application Development Trends, the Business Intelligence Journal,and TDWI Case Studies & Solution.

There is an easier way to build Hadoop applications. With this hands-on book, you'll learn how to use Cascading, the open source abstraction framework for Hadoop that lets you easily create and manage powerful enterprise-grade data processing applications—without having to learn the intricacies of MapReduce. Working with sample apps based on Java and other JVM languages, you'll quickly learn Cascading's streamlined approach to data processing, data filtering, and workflow optimization. This book demonstrates how this framework can help your business extract meaningful information from large amounts of distributed data. Start working on Cascading example projects right away Model and analyze unstructured data in any format, from any source Build and test applications with familiar constructs and reusable components Work with the Scalding and Cascalog Domain-Specific Languages Easily deploy applications to Hadoop, regardless of cluster location or data size Build workflows that integrate several big data frameworks and processes Explore common use cases for Cascading, including features and tools that support them Examine a case study that uses a dataset from the Open Data Initiative Learn everything you need to become a successful data architect on the Salesforce platform Key Features Adopt best practices relating to data governance and learn how to implement them Learn how to work with data in Salesforce while maintaining scalability and security of an instance Gain insights into managing large data volumes in Salesforce Book Description As Salesforce orgs mature over time, data management and integrations are becoming more challenging than ever. Salesforce Data Architecture and Management follows a hands-on approach to managing data and tracking the performance of your Salesforce org. You'll start by understanding the role and skills required to become a successful data architect. The book focuses on data modeling concepts, how to apply them in Salesforce, and how they relate to objects and fields in Salesforce. You'll learn the intricacies of managing data in Salesforce, starting from understanding why Salesforce has chosen to optimize for read rather than write operations. After developing a solid foundation, you'll explore examples and best practices for managing your data. You'll understand how to manage your master data and discover what the Golden Record is and why it is important for organizations. Next, you'll learn how to align your MDM and CRM strategy with a discussion on Salesforce's Customer 360 and its key components. You'll also cover data governance, its multiple facets, and how GDPR compliance can be achieved with Salesforce. Finally, you'll discover Large Data Volumes (LDVs) and best practices for migrating data using APIs. By the end of this book, you'll be well-versed with data management, data backup, storage, and archiving in Salesforce. What you will learn Understand the Salesforce data architecture Explore various data backup and archival strategies Understand how the Salesforce platform is designed and how it is different from other relational databases Uncover tools that can help in data management that minimize data trust issues in your Salesforce org Focus on the Salesforce Customer 360 platform, its key components, and how it can help organizations in connecting with customers Discover how Salesforce can be used for GDPR compliance Measure and monitor the performance of your Salesforce org Who this book is for This book is for aspiring architects, Salesforce admins, and developers. You will also find the book useful if you're preparing for the Salesforce Data Architecture and Management exam. A basic understanding of Salesforce is assumed.

Today, the world is trying to create and educate data scientists because of the phenomenon of Big Data. And everyone is looking deeply into this technology. But no one is looking at the larger architectural picture of how Big Data needs to fit within the existing systems (data warehousing systems). Taking a look at the larger picture into which Big Data fits gives the data scientist the necessary context for how pieces of the puzzle should fit together. Most references on Big Data look at only one tiny part of a much larger whole. Until data gathered can be put into an existing framework or architecture it can't be used to its full potential. Data Architecture a Primer for the Data Scientist addresses the larger architectural picture of how Big Data fits with the existing information infrastructure, an essential topic for the data scientist. Drawing upon years of practical experience and using numerous examples and an easy to understand framework. W.H. Inmon, and Daniel Linstedt define the importance of data architecture and how it can be used effectively to harness big data within existing systems. You'll be able to: Turn textual information into a form that can be analyzed by standard tools. Make the connection between analytics and Big Data Understand how Big Data fits within an existing systems environment Conduct analytics on repetitive and non-repetitive data Discusses the value in Big Data that is often overlooked, non-repetitive data, and why there is significant business value in using it Shows how to turn textual information into a form that can be analyzed by standard tools Explains how Big Data fits within an existing systems environment Presents new opportunities that are afforded by the advent of Big Data Demystifies the murky waters of repetitive and non-repetitive data in Big Data

Enterprise architecture requires an understanding of all technologies, strategies, and data consumption throughout the enterprise. To this end, one must strive to always broaden knowledge of existing, as well as emerging trends and solutions. As a trade, this role demands an understanding beyond the specificities of technologies and vendor products