Site To Download Modern Enterprise Data Pipelines

If you ally craving such a referred **Modern Enterprise Data Pipelines** book that will meet the expense of you worth, get the categorically best seller from us currently from several preferred authors. If you desire to droll books, lots of novels, tale, jokes, and more fictions collections are next launched, from best seller to one of the most current released.

You may not be perplexed to enjoy every ebook collections Modern Enterprise Data Pipelines that we will utterly offer. It is not more or less the costs. Its not quite what you obsession currently. This Modern Enterprise Data Pipelines, as one of the most keen sellers here will very be accompanied by the best options to review.

KEY=ENTERPRISE - STARK NEAL

Modern Enterprise Data Pipelines

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

Architecting Modern Data Platforms

A Guide to Enterprise Hadoop at Scale

"O'Reilly Media, Inc." 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

Snowflake: The Definitive Guide

"O'Reilly Media, Inc." Snowflake's ability to eliminate data silos and run workloads from a single platform creates opportunities to democratize data analytics, allowing users at all levels within an organization to make data-driven decisions. Whether you're an IT professional working in data warehousing or data science, a business analyst or technical manager, or an aspiring data professional wanting to get more hands-on experience with the Snowflake platform, this book is for you. You'll learn how Snowflake users can build modern integrated data applications and develop new revenue streams based on data. Using hands-on SQL examples, you'll also discover how the Snowflake Data Cloud helps you accelerate data science by avoiding replatforming or migrating data unnecessarily. You'll be able to: Efficiently capture, store, and process large amounts of data at an amazing speed Ingest and transform real-time data feeds in both structured and semistructured formats and deliver meaningful data insights within minutes Use Snowflake Time Travel and zero-copy cloning to produce a sensible data recovery strategy that balances system resilience with ongoing storage costs Securely share data and reduce or eliminate data integration costs by accessing ready-to-query datasets available in the Snowflake Marketplace

Data Management at Scale

"O'Reilly Media, Inc." 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

Architecting Google Cloud Solutions

Learn to design robust and future-proof solutions with Google Cloud technologies

Packt Publishing Ltd Achieve your business goals and build highly available, scalable, and secure cloud infrastructure by designing robust and cost-effective solutions as a Google Cloud Architect. Key FeaturesGain hands-on experience in designing and managing high-performance cloud solutionsLeverage Google Cloud Platform to optimize technical and business processes using cutting-edge technologies and servicesUse Google Cloud Big Data, AI, and ML services to design scalable and intelligent data solutionsBook Description Google has been one of the top players in the public cloud domain thanks to its agility and performance capabilities. This book will help you design, develop, and manage robust, secure, and dynamic solutions to successfully meet your business needs. You'll learn how to plan and design network, compute, storage, and big data systems that incorporate security and compliance from the ground up. The chapters will cover simple to complex use cases for devising solutions to business problems, before focusing on how to leverage Google Cloud's Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) capabilities for designing modern no-operations platforms. Throughout this book, you'll discover how to design for scalability, resiliency, and high availability. Later, you'll find out how to use Google Cloud to design modern applications using microservices architecture, automation, and Infrastructure-as-Code (IaC) practices. The concluding chapters then demonstrate how to apply machine learning and artificial intelligence (AI) to derive insights from your data. Finally, you will discover best practices for operating and monitoring your cloud solutions, as well as performing troubleshooting and quality assurance. By the end of this Google Cloud book, you'll be able to design robust enterprise-grade solutions using Google Cloud Platform. What you will learnGet to grips with compute, storage, networking, data analytics, and pricingDiscover delivery models such as IaaS, PaaS, and SaaSExplore the underlying technologies and economics of cloud computingDesign for scalability, business continuity, observability, and resiliencySecure Google Cloud solutions and ensure complianceUnderstand operational best practices and learn how to architect a monitoring solutionGain insights into modern application design with Google CloudLeverage big data, machine learning, and AI with Google CloudWho this book is for This book is for cloud architects who are responsible for designing and managing cloud solutions with GCP. You'll also find the book useful if you're a system engineer or enterprise architect looking to learn how to design solutions with Google Cloud. Moreover, cloud architects who already have experience with other cloud providers and are now beginning to work with Google Cloud will benefit from the book. Although an intermediate-level understanding of cloud computing and distributed apps is required, prior experience of working in the public and hybrid cloud domain is not mandatory.

Introducing Microsoft SQL Server 2019

Reliability, scalability, and security both on premises and in the cloud

Packt Publishing Ltd Explore the impressive storage and analytic tools available with the in-cloud and on-premises versions of Microsoft SQL Server 2019. Key FeaturesGain insights into what's new in SQL Server 2019Understand use cases and customer scenarios that can be implemented with SQL Server 2019Discover new cross-platform tools that simplify management and analysisBook Description Microsoft SQL Server comes equipped with industry-leading features and the best online transaction processing capabilities. If you are looking to work with data processing and management, getting up to speed with Microsoft Server 2019 is key. Introducing SQL Server 2019 takes you through the latest features in SQL Server 2019 and their importance. You will learn to unlock faster querying speeds and understand how to leverage the new and improved security features to build robust data management solutions. Further chapters will assist you with integrating, managing, and analyzing all data, including relational, NoSQL, and unstructured big data using SQL Server 2019. Dedicated sections in the book will also demonstrate how you can use SQL Server 2019 to leverage data processing platforms, such as Apache Hadoop and Spark, and containerization technologies like Docker and Kubernetes to control your data and efficiently monitor it. By the end of this book, you'll be well versed with all the features of Microsoft SQL Server 2019 and understand how to use them confidently to build robust data management solutions. What you will learnBuild a custom container image with a DockerfileDeploy and run the SQL Server 2019 container imageUnderstand how to use SQL server on LinuxMigrate existing paginated reports to Power BI Report ServerLearn to query Hadoop Distributed File System (HDFS) data using Azure Data StudioUnderstand the benefits of In-Memory OLTPWho this book is for This book is for database administrators, architects, big data engineers, or anyone who has experience with SQL Server and wants to explore and implement the new features in SQL Server 2019. Basic working knowledge of SQL Server and relational database management system (RDBMS) is required.

Unlock Complex and Streaming Data with Declarative Data Pipelines

How New Technologies Have Transformed the Way Modern Data is Processed for Analytics

Unlocking the value of modern data is critical for data-driven companies. This report provides a concise, practical guide to building a data architecture that efficiently delivers big, complex, and streaming data to both internal users and customers. Authors Ori Rafael, Roy Hasson, and Rick Bilodeau from Upsolver examine how modern data pipelines can improve business outcomes. Tech leaders and data engineers will explore the role these pipelines play in the data architecture and learn how to intelligently consider tradeoffs between different data architecture patterns and data pipeline development approaches.

Future And Fintech, The: Abcdi And Beyond

World Scientific The Future and FinTech examines the fundamental financial technologies and its growing impact on the Banking, Financial Services and Insurance (BFSI) sectors. With global investment amounting to more than $100 billion in 2020, the proliferation of FinTech has underpinned the direction payments, loans, wealth management, insurance, and cryptocurrencies are heading.This book presents FinTech from an industrial perspective in the context of architecture and its basic building blocks, e.g., Artificial Intelligence (AI), Blockchain, Cloud, Big Data, Internet of Things (IoT), and its connections to real-life applications at work. It provides a detailed guidance on how FinTech digitalizes business operations, improves productivity and efficiency, and optimizes resource management with the help of some new concepts, such as AIOps, MLOps and DevSecOps. Readers will also discover how FinTech Innovations connect BFSI to the rest of the world with growing interests in Open Banking, Banking-as-a-Service (BaaS) and FinTech-as-a-Service (FaaS).To help readers understand how FinTech has unlocked numerous opportunities for tapping into the massive substantial group of customers, this book illustrates the massive changes already underway and provides insights into changes yet to come through practical examples and applications with illustrative figures and summary tables, making this book a handy quick reference for all things of FinTech.Related Link(s)

Spring Data

Modern Data Access for Enterprise Java

"O'Reilly Media, Inc." You can choose several data access frameworks when building Java enterprise applications that work with relational databases. But what about big data? This hands-on introduction shows you how Spring Data makes it relatively easy to build applications across a wide range of new data access technologies such as NoSQL and Hadoop. Through several sample projects, you'll learn how Spring Data provides a consistent programming model that retains NoSQL-specific features and capabilities, and helps you develop Hadoop applications across a wide range of use-cases such as data analysis, event stream processing, and workflow. You'll also discover the features Spring Data adds to Spring's existing JPA and JDBC support for writing RDBMS-based data access layers. Learn about Spring's template helper classes to simplify the use ofdatabase-specific functionality Explore Spring Data's repository abstraction and advanced query functionality Use Spring Data with Redis (key/value store), HBase(column-family), MongoDB (document database), and Neo4j (graph database) Discover the GemFire distributed data grid solution Export Spring Data JPA-managed entities to the Web as RESTful web services Simplify the development of HBase applications, using a lightweight object-mapping framework Build example big-data pipelines with Spring Batch and Spring Integration

The Self-Service Data Roadmap

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

Manufacturing and Enterprise

An Integrated Systems Approach

CRC Press This book presents an integrated systems approach to manufacturing and business enterprise. Traditionally, these topics are treated as separate and independent subjects, but the practical fact is that the manufacturing and the business enterprises are intertwined. Currently, there is no book on the market that addresses both subjects from an integrated systems approach with a manufacturing engineering foundation. Topics covered include engineering process, systems modeling, business enterprise, forecasting, inventory management, product design, and project management.

DevOps for the Modern Enterprise

Winning Practices to Transform Legacy IT Organizations

IT Revolution Many organizations are facing the uphill battle of modernizing their legacy IT infrastructure. Most have evolved over the years by taking lessons from traditional or legacy manufacturing: creating a production process that puts the emphasis on the process instead of the people performing the tasks, allowing the organization to treat people like resources to try to achieve high-quality outcomes. But those practices and ideas are failing modern IT, where collaboration and creativeness are required to achieve high-performing, high-quality success. Mirco Hering, a thought leader in managing IT within legacy organizations, lays out a roadmap to success for IT managers, showing them how to create the right ecosystem, how to empower people to bring their best to work every day, and how to put the right technology in the driver's seat to propel their organization to success. But just having the right methods and tools will not magically transform an organization; the cultural change that is the hardest is also the most impactful. Using principles from Agile, Lean, and DevOps as well as first-hand examples from the enterprise world, Hering addresses the different challenges that legacy organizations face as they transform into modern IT departments.

Modern Big Data Processing with Hadoop

Expert techniques for architecting end-to-end big data solutions to get valuable insights

Packt Publishing Ltd A comprehensive guide to design, build and execute effective Big Data strategies using Hadoop Key Features -Get an in-depth view of the Apache Hadoop ecosystem and an overview of the architectural patterns pertaining to the popular Big Data platform -Conquer different data processing and analytics challenges using a multitude of tools such as Apache Spark, Elasticsearch, Tableau and more -A comprehensive, step-by-step guide that will teach you everything you need to know, to be an expert Hadoop Architect Book Description The complex structure of data these days requires sophisticated solutions for data transformation, to make the information more accessible to the users.This book empowers you to build such solutions with relative ease with the help of Apache Hadoop, along with a host of other Big Data tools. This book will give you a complete understanding of the data lifecycle management with Hadoop, followed by modeling of structured and unstructured data in Hadoop. It will also show you how to design real-time streaming pipelines by leveraging tools such as Apache Spark, and build efficient enterprise search solutions using Elasticsearch. You will learn to build enterprise-grade analytics solutions on Hadoop, and how to visualize your data using tools such as Apache Superset. This book also covers techniques for deploying your Big Data solutions on the cloud Apache Ambari, as well as expert techniques for managing and administering your Hadoop cluster. By the end of this book, you will have all the knowledge you need to build expert Big Data systems. What you will learn Build an efficient enterprise Big Data strategy centered around Apache Hadoop Gain a thorough understanding of using Hadoop with various Big Data frameworks such as Apache Spark, Elasticsearch and more Set up and deploy your Big Data environment on premises or on the cloud with Apache Ambari Design effective streaming data pipelines and build your own enterprise search solutions Utilize the historical data to build your analytics solutions and visualize them using popular tools such as Apache Superset Plan, set up and administer your Hadoop cluster efficiently Who this book is for This book is for Big Data professionals who want to fast-track their career

in the Hadoop industry and become an expert Big Data architect. Project managers and mainframe professionals looking forward to build a career in Big Data Hadoop will also find this book to be useful. Some understanding of Hadoop is required to get the best out of this book.

Performance Dashboards

Measuring, Monitoring, and Managing Your Business

John Wiley & Sons 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.

Snowflake Cookbook

Techniques for building modern cloud data warehousing solutions

Packt Publishing Ltd Develop modern solutions with Snowflake's unique architecture and integration capabilities; process bulk and real-time data into a data lake; and leverage time travel, cloning, and data-sharing features to optimize data operations Key FeaturesBuild and scale modern data solutions using the all-in-one Snowflake platformPerform advanced cloud analytics for implementing big data and data science solutionsMake quicker and better-informed business decisions by uncovering key insights from your dataBook Description Snowflake is a unique cloud-based data warehousing platform built from scratch to perform data management on the cloud. This book introduces you to Snowflake's unique architecture, which places it at the forefront of cloud data warehouses. You'll explore the compute model available with Snowflake, and find out how Snowflake allows extensive scaling through the virtual warehouses. You will then learn how to configure a virtual warehouse for optimizing cost and performance. Moving on, you'll get to grips with the data ecosystem and discover how Snowflake integrates with other technologies for staging and loading data. As you progress through the chapters, you will leverage Snowflake's capabilities to process a series of SQL statements using tasks to build data pipelines and find out how you can create modern data solutions and pipelines designed to provide high performance and scalability. You will also get to grips with creating role hierarchies, adding custom roles, and setting default roles for users before covering advanced topics such as data sharing, cloning, and performance optimization. By the end of this Snowflake book, you will be well-versed in Snowflake's architecture for building modern analytical solutions and understand best practices for solving commonly faced problems using practical recipes. What you will learnGet to grips with data warehousing techniques aligned with Snowflake's cloud architectureBroaden your skills as a data warehouse designer to cover the Snowflake ecosystemTransfer skills from on-premise data warehousing to the Snowflake cloud analytics platformOptimize performance and costs associated with a Snowflake solutionStage data on object stores and load it into SnowflakeSecure data and share it efficiently for accessManage transactions and extend Snowflake using stored proceduresExtend cloud data applications using Spark ConnectorWho this book is for This book is for data warehouse developers, data analysts, database administrators, and anyone involved in designing, implementing, and optimizing a Snowflake data warehouse. Knowledge of data warehousing and database and cloud concepts will be useful. Basic familiarity with Snowflake is beneficial, but not necessary.

Designing Cloud Data Platforms

Simon and Schuster In Designing Cloud Data Platforms, Danil Zburivsky and Lynda Partner reveal a six-layer approach that increases flexibility and reduces costs. Discover patterns for ingesting data from a variety of sources, then learn to harness pre-built services provided by cloud vendors. Summary Centralized data warehouses, the long-time defacto standard for housing data for analytics, are rapidly giving way to multi-faceted cloud data platforms. Companies that embrace modern cloud data platforms benefit from an integrated view of their business using all of their data and can take advantage of advanced analytic practices to drive predictions and as yet unimagined data services. Designing Cloud Data Platforms is a hands-on guide to envisioning and designing a modern scalable data platform that takes full advantage of the flexibility of the cloud. As you read, you'll learn the core components of a cloud data platform design, along with the role of key technologies like Spark and Kafka Streams. You'll also explore setting up processes to manage cloud-based data, keep it secure, and using advanced analytic and BI tools to analyze it. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Well-designed pipelines, storage systems, and APIs eliminate the complicated scaling and maintenance required with on-prem data centers. Once you learn the patterns for designing cloud data platforms, you'll maximize performance no matter which cloud vendor you use. About the book In Designing Cloud Data Platforms, Danil Zburivsky and Lynda Partner reveal a six-layer approach that increases flexibility and reduces costs. Discover patterns for ingesting data from a variety of sources, then learn to harness pre-built services provided by cloud vendors. What's inside Best practices for structured and unstructured data sets Cloud-ready machine learning tools Metadata and real-time analytics Defensive architecture, access, and security About the reader For data professionals familiar with the basics of cloud computing, and Hadoop or Spark. About the author Danil Zburivsky has over 10 years of experience designing and supporting large-scale data infrastructure for enterprises across the globe. Lynda Partner is the VP of Analytics-as-a-Service at Pythian, and has been on the business side of data for over 20 years. Table of Contents 1 Introducing the data platform 2 Why a data platform and not just a data warehouse 3 Getting bigger and leveraging the Big 3: Amazon, Microsoft Azure, and Google 4 Getting data into the platform 5 Organizing and processing data 6 Real-time data processing and analytics 7 Metadata layer architecture 8 Schema management 9 Data access and security 10 Fueling business value with data platforms

Data Pipelines Pocket Reference

O'Reilly Media Data pipelines are the foundation for success in data analytics. Moving data from numerous diverse sources and transforming it to provide context is the difference between having data and actually gaining value from it. This pocket reference defines data pipelines and explains how they work in today's modern data stack. You'll learn common considerations and key decision points when implementing pipelines, such as batch versus streaming data ingestion and build versus buy. This book addresses the most common decisions made by data professionals and discusses foundational concepts that apply to open source frameworks, commercial products, and homegrown solutions. You'll learn: What a data pipeline is and how it works How data is moved and processed on modern data infrastructure, including cloud platforms Common tools and products used by data engineers to build pipelines How pipelines support analytics and reporting needs Considerations for pipeline maintenance, testing, and alerting

Data Pipelines with Apache Airflow

Simon and Schuster "An Airflow bible. Useful for all kinds of users, from novice to expert." - Rambabu Posa, Sai Aashika Consultancy Data Pipelines with Apache Airflow teaches you how to build and maintain effective data pipelines. A successful pipeline moves data efficiently, minimizing pauses and blockages between tasks, keeping every process along the way operational. Apache Airflow provides a single customizable environment for building and managing data pipelines, eliminating the need for a hodgepodge collection of tools, snowflake code, and homegrown processes. Using real-world scenarios and examples, Data Pipelines with Apache Airflow teaches you how to simplify and automate data pipelines, reduce operational overhead, and smoothly integrate all the technologies in your stack. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Data pipelines manage the flow of data from initial collection through consolidation, cleaning, analysis, visualization, and more. Apache Airflow provides a single platform you can use to design, implement, monitor, and maintain your pipelines. Its easy-to-use UI, plug-and-play options, and flexible Python scripting make Airflow perfect for any data management task. About the book Data Pipelines with Apache Airflow teaches you how to build and maintain effective data pipelines. You'll explore the most common usage patterns, including aggregating multiple data sources, connecting to and from data lakes, and cloud deployment. Part reference and part tutorial, this practical guide covers every aspect of the directed acyclic graphs (DAGs) that power Airflow, and how to customize them for your pipeline's needs. What's inside Build, test, and deploy Airflow pipelines as DAGs Automate moving and transforming data Analyze historical datasets using backfilling Develop custom components Set up Airflow in production environments About the reader For DevOps, data engineers, machine learning engineers, and sysadmins with intermediate Python skills. About the author Bas Harenslak and Julian de Ruiter are data engineers with extensive experience using Airflow to develop pipelines for major companies. Bas is also an Airflow committer. Table of Contents PART 1 - GETTING STARTED 1 Meet Apache Airflow 2 Anatomy of an Airflow DAG 3 Scheduling in Airflow 4 Templating tasks using the Airflow context 5 Defining dependencies between tasks PART 2 - BEYOND THE BASICS 6 Triggering workflows 7 Communicating with external systems 8 Building custom components 9 Testing 10 Running tasks in containers PART 3 - AIRFLOW IN PRACTICE 11 Best practices 12 Operating Airflow in production 13 Securing Airflow 14 Project: Finding the fastest way to get around NYC PART 4 - IN THE CLOUDS 15 Airflow in the clouds 16 Airflow on AWS 17 Airflow on Azure 18 Airflow in GCP

Data Engineering with AWS

Learn how to design and build cloud-based data transformation pipelines using AWS

Packt Publishing Ltd The missing expert-led manual for the AWS ecosystem — go from foundations to building data engineering pipelines effortlessly Key FeaturesLearn about common data architectures and modern approaches to generating value from big dataExplore AWS tools for ingesting, transforming, and consuming data, and for orchestrating pipelinesLearn how to architect and implement data lakes and data lakehouses for big data analytics from a data lakes expertBook Description Written by a Senior Data Architect with over twenty-five years of experience in the business, Data Engineering for AWS is a book whose sole aim is to make you proficient in using the AWS ecosystem. Using a thorough and hands-on approach to data, this book will give aspiring and new data engineers a solid theoretical and practical foundation to succeed with AWS. As you progress,

you'll be taken through the services and the skills you need to architect and implement data pipelines on AWS. You'll begin by reviewing important data engineering concepts and some of the core AWS services that form a part of the data engineer's toolkit. You'll then architect a data pipeline, review raw data sources, transform the data, and learn how the transformed data is used by various data consumers. You'll also learn about populating data marts and data warehouses along with how a data lakehouse fits into the picture. Later, you'll be introduced to AWS tools for analyzing data, including those for ad-hoc SQL queries and creating visualizations. In the final chapters, you'll understand how the power of machine learning and artificial intelligence can be used to draw new insights from data. By the end of this AWS book, you'll be able to carry out data engineering tasks and implement a data pipeline on AWS independently. What you will learnUnderstand data engineering concepts and emerging technologiesIngest streaming data with Amazon Kinesis Data FirehoseOptimize, denormalize, and join datasets with AWS Glue StudioUse Amazon S3 events to trigger a Lambda process to transform a fileRun complex SQL queries on data lake data using Amazon AthenaLoad data into a Redshift data warehouse and run queriesCreate a visualization of your data using Amazon QuickSightExtract sentiment data from a dataset using Amazon ComprehendWho this book is for This book is for data engineers, data analysts, and data architects who are new to AWS and looking to extend their skills to the AWS cloud. Anyone new to data engineering who wants to learn about the foundational concepts while gaining practical experience with common data engineering services on AWS will also find this book useful. A basic understanding of big data-related topics and Python coding will help you get the most out of this book but it's not a prerequisite. Familiarity with the AWS console and core services will also help you follow along.

Distributed Data Systems with Azure Databricks

Create, deploy, and manage enterprise data pipelines

Packt Publishing Ltd Quickly build and deploy massive data pipelines and improve productivity using Azure Databricks Key FeaturesGet to grips with the distributed training and deployment of machine learning and deep learning modelsLearn how ETLs are integrated with Azure Data Factory and Delta LakeExplore deep learning and machine learning models in a distributed computing infrastructureBook Description Microsoft Azure Databricks helps you to harness the power of distributed computing and apply it to create robust data pipelines, along with training and deploying machine learning and deep learning models. Databricks' advanced features enable developers to process, transform, and explore data. Distributed Data Systems with Azure Databricks will help you to put your knowledge of Databricks to work to create big data pipelines. The book provides a hands-on approach to implementing Azure Databricks and its associated methodologies that will make you productive in no time. Complete with detailed explanations of essential concepts, practical examples, and self-assessment questions, you'll begin with a quick introduction to Databricks core functionalities, before performing distributed model training and inference using TensorFlow and Spark MLlib. As you advance, you'll explore MLflow Model Serving on Azure Databricks and implement distributed training pipelines using HorovodRunner in Databricks. Finally, you'll discover how to transform, use, and obtain insights from massive amounts of data to train predictive models and create entire fully working data pipelines. By the end of this MS Azure book, you'll have gained a solid understanding of how to work with Databricks to create and manage an entire big data pipeline. What you will learnCreate ETLs for big data in Azure DatabricksTrain, manage, and deploy machine learning and deep learning modelsIntegrate Databricks with Azure Data Factory for extract, transform, load (ETL) pipeline creationDiscover how to use Horovod for distributed deep learningFind out how to use Delta Engine to query and process data from Delta LakeUnderstand how to use Data Factory in combination with DatabricksUse Structured Streaming in a production-like environmentWho this book is for This book is for software engineers, machine learning engineers, data scientists, and data engineers who are new to Azure Databricks and want to build high-quality data pipelines without worrying about infrastructure. Knowledge of Azure Databricks basics is required to learn the concepts covered in this book more effectively. A basic understanding of machine learning concepts and beginner-level Python programming knowledge is also recommended.

Amazon Redshift Cookbook

Recipes for building modern data warehousing solutions

Packt Publishing Ltd Discover how to build a cloud-based data warehouse at petabyte-scale that is burstable and built to scale for end-to-end analytical solutions Key FeaturesDiscover how to translate familiar data warehousing concepts into Redshift implementationUse impressive Redshift features to optimize development, productionizing, and operations processesFind out how to use advanced features such as concurrency scaling, Redshift Spectrum, and federated queriesBook Description Amazon Redshift is a fully managed, petabyte-scale AWS cloud data warehousing service. It enables you to build new data warehouse workloads on AWS and migrate on-premises traditional data warehousing platforms to Redshift. This book on Amazon Redshift starts by focusing on Redshift architecture, showing you how to perform database administration tasks on Redshift. You'll then learn how to optimize your data warehouse to quickly execute complex analytic queries against very large datasets. Because of the massive amount of data involved in data warehousing, designing your database for analytical processing lets you take full advantage of Redshift's columnar architecture and managed services. As you advance, you'll discover how to deploy fully automated and highly scalable extract, transform, and load (ETL) processes, which help minimize the operational efforts that you have to invest in managing regular ETL pipelines and ensure the timely and accurate refreshing of your data warehouse. Finally, you'll gain a clear understanding of Redshift use cases, data ingestion, data management, security, and scaling so that you can build a scalable data warehouse platform. By the end of this Redshift book, you'll be able to implement a Redshift-based data analytics solution and have understood the best practice solutions to commonly faced problems. What you will learnUse Amazon Redshift to build petabyte-scale data warehouses that are agile at scaleIntegrate your data warehousing solution with a data lake using purpose-built features and services on AWSBuild end-to-end analytical solutions from data sourcing to consumption with the help of useful recipesLeverage Redshift's comprehensive security capabilities to meet the most demanding business requirementsFocus on architectural insights and rationale when using analytical recipesDiscover best practices for working with big data to operate a fully managed solutionWho this book is for This book is for anyone involved in architecting, implementing, and optimizing an Amazon Redshift data warehouse, such as data warehouse developers, data analysts, database administrators, data engineers, and data scientists. Basic knowledge of data warehousing, database systems, and cloud concepts and familiarity with Redshift will be beneficial.

The Enterprise Big Data Lake

Delivering the Promise of Big Data and Data Science

"O'Reilly Media, Inc." 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

The Informed Company

How to Build a Cloud-Based Data Stack to Explore and Understand Data

John Wiley & Sons Learn how to manage a modern data stack and get the most out of data in your organization! Thanks to the emergence of new technologies and the explosion of data in recent years, we need new practices for managing and getting value out of data. In the modern, data driven competitive landscape the "best guess" approach—reading blog posts here and there and patching together data practices without any real visibility—is no longer going to hack it. The Informed Company provides definitive direction on how best to leverage the modern data stack, including cloud computing, columnar storage, cloud ETL tools, and cloud BI tools. You'll learn how to work with Agile methods and set up processes that's right for your company to use your data as a key weapon for your success . . . You'll discover best practices for every stage, from querying production databases at a small startup all the way to setting up data marts for different business lines of an enterprise. In their work at Chartio, authors Fowler and David have learned that most businesspeople are almost completely self-taught when it comes to data. If they are using resources, those resources are outdated, so they're missing out on the latest cloud technologies and advances in data analytics. This book will firm up your understanding of data and bring you into the present with knowledge around what works and what doesn't. Discover the data stack strategies that are working for today's successful small, medium, and enterprise companies Learn the different Agile stages of data organization, and the right one for your team Learn how to maintain Data Lakes and Data Warehouses for effective, accessible data storage Gain the knowledge you need to architect Data Warehouses and Data Marts Understand your business's level of data sophistication and the steps you can take to get to "level up" your data The Informed Company is the definitive data book for anyone who wants to work faster and more nimbly, armed with actionable decision-making data.

Azure for Architects

Create secure, scalable, high-availability applications on the cloud, 3rd Edition

Packt Publishing Ltd Build and design multiple types of applications that are cross-language, platform, and cost-effective by understanding core Azure principles and foundational concepts Key Features Get familiar with the different design patterns available in Microsoft Azure Develop Azure cloud architecture and a pipeline management system Get to know the security best practices for your Azure deployment Book Description Thanks to its support for high availability, scalability, security, performance, and disaster recovery, Azure has been widely adopted to create and deploy different types of application with ease. Updated for the latest developments, this third edition of Azure for Architects helps you get to grips with the core concepts of designing serverless architecture, including containers, Kubernetes deployments, and big data solutions. You'll learn how to architect solutions such as serverless functions, you'll discover deployment patterns for containers and Kubernetes, and you'll explore large-scale big data processing using Spark and Databricks. As you advance, you'll implement DevOps using Azure DevOps, work with intelligent solutions using Azure Cognitive Services, and integrate security, high availability, and scalability into each solution. Finally, you'll delve into Azure security concepts such as OAuth, OpenConnect, and managed identities. By the end of this book, you'll have gained the confidence to design intelligent Azure solutions based on containers and serverless functions. What you will learn Understand the components of the Azure cloud platform Use cloud design patterns Use enterprise security guidelines for your Azure deployment Design and implement serverless and integration solutions Build efficient data solutions on Azure Understand container services on Azure Who this book is for If you are a cloud architect, DevOps engineer, or a developer looking to learn about the key architectural aspects of the Azure cloud platform, this book is for you. A basic understanding of the Azure cloud platform will help you grasp the concepts covered in this book more effectively.

Data Science on the Google Cloud Platform

"O'Reilly Media, Inc." Learn how easy it is to apply sophisticated statistical and machine learning methods to real-world problems when you build using Google Cloud Platform (GCP). This hands-on guide shows data engineers and data scientists how to implement an end-to-end data pipeline with cloud native tools on GCP. Throughout this updated second edition, you'll work through a sample business decision by employing a variety of data science approaches. Follow along by building a data pipeline in your own project on GCP, and discover how to solve data science problems in a transformative and more collaborative way. You'll learn how to: Employ best practices in building highly scalable data and ML pipelines on Google Cloud Automate and schedule data ingest using Cloud Run Create and populate a dashboard in Data Studio Build a real-time analytics pipeline using Pub/Sub, Dataflow, and BigQuery Conduct interactive data exploration with BigQuery Create a Bayesian model with Spark on Cloud Dataproc Forecast time series and do anomaly detection with BigQuery ML Aggregate within time windows with Dataflow Train explainable machine learning models with Vertex AI Operationalize ML with Vertex AI Pipelines

Data Lakes For Dummies

John Wiley & Sons Take a dive into data lakes "Data lakes" is the latest buzz word in the world of data storage, management, and analysis. Data Lakes For Dummies decodes and demystifies the concept and helps you get a straightforward answer the question: "What exactly is a data lake and do I need one for my business?" Written for an audience of technology decision makers tasked with keeping up with the latest and greatest data options, this book provides the perfect introductory survey of these novel and growing features of the information landscape. It explains how they can help your business, what they can (and can't) achieve, and what you need to do to create the lake that best suits your particular needs. With a minimum of jargon, prolific tech author and business intelligence consultant Alan Simon explains how data lakes differ from other data storage paradigms. Once you've got the background picture, he maps out ways you can add a data lake to your business systems; migrate existing information and switch on the fresh data supply; clean up the product; and open channels to the best intelligence software for to interpreting what you've stored. Understand and build data lake architecture Store, clean, and synchronize new and existing data Compare the best data lake vendors Structure raw data and produce usable analytics Whatever your business, data lakes are going to form ever more prominent parts of the information universe every business should have access to. Dive into this book to start exploring the deep competitive advantage they make possible—and make sure your business isn't left standing on the shore.

Understanding Modern Business Data Processing

New York : Gregg Division, McGraw-Hill

Data Engineering with Apache Spark, Delta Lake, and Lakehouse

Create scalable pipelines that ingest, curate, and aggregate complex data in a timely and secure way

Packt Publishing Ltd Understand the complexities of modern-day data engineering platforms and explore strategies to deal with them with the help of use case scenarios led by an industry expert in big data Key Features Become well-versed with the core concepts of Apache Spark and Delta Lake for building data platforms Learn how to ingest, process, and analyze data that can be later used for training machine learning models Understand how to operationalize data models in production using curated data Book Description In the world of ever-changing data and schemas, it is important to build data pipelines that can auto-adjust to changes. This book will help you build scalable data platforms that managers, data scientists, and data analysts can rely on. Starting with an introduction to data engineering, along with its key concepts and architectures, this book will show you how to use Microsoft Azure Cloud services effectively for data engineering. You'll cover data lake design patterns and the different stages through which the data needs to flow in a typical data lake. Once you've explored the main features of Delta Lake to build data lakes with fast performance and governance in mind, you'll advance to implementing the lambda architecture using Delta Lake. Packed with practical examples and code snippets, this book takes you through real-world examples based on production scenarios faced by the author in his 10 years of experience working with big data. Finally, you'll cover data lake deployment strategies that play an important role in provisioning the cloud resources and deploying the data pipelines in a repeatable and continuous way. By the end of this data engineering book, you'll know how to effectively deal with ever-changing data and create scalable data pipelines to streamline data science, ML, and artificial intelligence (AI) tasks. What you will learn Discover the challenges you may face in the data engineering world Add ACID transactions to Apache Spark using Delta Lake Understand effective design strategies to build enterprise-grade data lakes Explore architectural and design patterns for building efficient data ingestion pipelines Orchestrate a data pipeline for preprocessing data using Apache Spark and Delta Lake APIs Automate deployment and monitoring of data pipelines in production Get to grips with securing, monitoring, and managing data pipelines models efficiently Who this book is for This book is for aspiring data engineers and data analysts who are new to the world of data engineering and are looking for a practical guide to building scalable data platforms. If you already work with PySpark and want to use Delta Lake for data engineering, you'll find this book useful. Basic knowledge of Python, Spark, and SQL is expected.

Data Lake for Enterprises

Packt Publishing Ltd 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.

Data Warehousing For Dummies

John Wiley & Sons Data warehousing is one of the hottest business topics, and there's more to understanding data warehousing technologies than you might think. Find out the basics of data warehousing and how it facilitates data mining and business intelligence with Data Warehousing For Dummies, 2nd Edition. Data is probably your company's most important asset, so your data warehouse should serve your needs. The fully updated Second Edition of Data Warehousing For Dummies helps you understand, develop, implement, and use data warehouses, and offers a sneak peek into their future. You'll learn to: Analyze top-down and bottom-up data warehouse designs Understand the structure and technologies of data warehouses, operational data stores, and data marts Choose your project team and apply best development practices to your data warehousing projects Implement a data warehouse, step by step, and involve end-users in the process Review and upgrade existing data storage to make it serve your needs Comprehend OLAP, column-wise databases, hardware assisted databases, and middleware Use data mining intelligently and find what you need Make informed choices about consultants and data warehousing products Data Warehousing For Dummies, 2nd Edition also shows you how to involve users in the testing process and gain valuable feedback, what it takes to successfully manage a data warehouse project, and how to tell if your project is on track. You'll find it's the most useful source of data on the topic!

Enterprise Architecture for Digital Business

"O'Reilly Media, Inc." Digital transformation has accelerated nearly tenfold in recent years as both a business and technology journey. Yet, most white papers and how-to guides still focus solely on the business side, rather than include methods for optimizing the technology behind it. This handbook shows CIOs, IT directors, and architects how to balance these two concerns successfully. You'll explore current technology trends and shifts required to build a digital business, including how enterprise architecture should evolve if it's to sustain and grow your business. A CIO who can handle digital transformation along with business interests is a rare find. This is the ideal guide to modernizing IT. You'll examine: The latest trends and technologies driving the need for a digital enterprise architecture New components, layers, and concepts that comprise a framework for digital enterprise architecture Skills and technologies you need to modernize an enterprise architecture for a digital business Domains and characteristics of a digital enterprise architecture How to map digital enterprise technologies to the appropriate teams

Business in Real-Time Using Azure IoT and Cortana Intelligence Suite

Driving Your Digital Transformation

Apress Learn how today's businesses can transform themselves by leveraging real-time data and advanced machine learning analytics. This book provides prescriptive guidance for architects and developers on the design and development of modern Internet of Things (IoT) and Advanced Analytics solutions. In addition, Business in Real-Time Using Azure IoT and Cortana Intelligence Suite offers patterns and practices for those looking to engage their customers and partners through Software-as-a-Service solutions that work on any device. Whether you're working in Health & Life Sciences, Manufacturing, Retail, Smart Cities and Buildings or Process Control, there exists a common platform from which you can create your targeted vertical solutions. Business in Real-Time Using Azure IoT and Cortana Intelligence Suite uses a reference architecture as a road map. Building on Azure's PaaS services, you'll see how a solution architecture unfolds that demonstrates a complete end-to-end IoT and Advanced Analytics scenario. What You'll Learn: Automate your software product life cycle using PowerShell, Azure Resource Manager Templates, and Visual Studio Team Services Implement smart devices using Node.JS and C# Use Azure Streaming Analytics to ingest millions of events Provide both "Hot" and "Cold" path outputs for real-time alerts, data transformations, and aggregation analytics Implement batch processing using Azure Data Factory Create a new form of Actionable Intelligence (AI) to drive mission critical business processes Provide rich Data Visualizations across a wide variety of mobile and web devices Who This Book is For: Solution Architects, Software Developers, Data Architects, Data Scientists, and CIO/CTA Technical Leadership Professionals

Handbook of Research on Foundations and Applications of Intelligent Business Analytics

IGI Global Intelligent business analytics is an emerging technology that has become a mainstream market adopted broadly across industries, organizations, and geographic regions. Intelligent business analytics is a current focus for research and development across academia and industries and must be examined and considered thoroughly so businesses can apply the technology appropriately. The Handbook of Research on Foundations and Applications of Intelligent Business Analytics examines the technologies and applications of intelligent business analytics and discusses the foundations of intelligent analytics such as intelligent mining, intelligent statistical modeling, and machine learning. Covering topics such as augmented analytics and artificial intelligence systems, this major reference work is ideal for scholars, engineers, professors, practitioners, researchers, industry professionals, academicians, and students.

The Modern Data Warehouse in Azure

Building with Speed and Agility on Microsoft's Cloud Platform

Apress Build a modern data warehouse on Microsoft's Azure Platform that is flexible, adaptable, and fast—fast to snap together, reconfigure, and fast at delivering results to drive good decision making in your business. Gone are the days when data warehousing projects were lumbering dinosaur-style projects that took forever, drained budgets, and produced business intelligence (BI) just in time to tell you what to do 10 years ago. This book will show you how to assemble a data warehouse solution like a jigsaw puzzle by connecting specific Azure technologies that address your own needs and bring value to your business. You will see how to implement a range of architectural patterns using batches, events, and streams for both data lake technology and SQL databases. You will discover how to manage metadata and automation to accelerate the development of your warehouse while establishing resilience at every level. And you will know how to feed downstream analytic solutions such as Power BI and Azure Analysis Services to empower data-driven decision making that drives your business forward toward a pattern of success. This book teaches you how to employ the Azure platform in a strategy to dramatically improve implementation speed and flexibility of data warehousing systems. You will know how to make correct decisions in design, architecture, and infrastructure such as choosing which type of SQL engine (from at least three options) best meets the needs of your organization. You also will learn about ETL/ELT structure and the vast number of accelerators and patterns that can be used to aid implementation and ensure resilience. Data warehouse developers and architects will find this book a tremendous resource for moving their skills into the future through cloud-based implementations. What You Will LearnChoose the appropriate Azure SQL engine for implementing a given data warehouse Develop smart, reusable ETL/ELT processes that are resilient and easily maintained Automate mundane development tasks through tools such as PowerShell Ensure consistency of data by creating and enforcing data contracts Explore streaming and event-driven architectures for data ingestionCreate advanced staging layers using Azure Data Lake Gen 2 to feed your data warehouse Who This Book Is For Data warehouse or ETL/ELT developers who wish to implement a data warehouse project in the Azure cloud, and developers currently working in on-premise environments who want to move to the cloud, and for developers with Azure experience looking to tighten up their implementation and consolidate their knowledge

Managing Data in Motion

Data Integration Best Practice Techniques and Technologies

Newnes 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"

Pro Hadoop Data Analytics

Designing and Building Big Data Systems using the Hadoop Ecosystem

Apress Learn advanced analytical techniques and leverage existing tool kits to make your analytic applications more powerful, precise, and efficient. This book provides the right combination of architecture, design, and implementation information to create analytical systems that go beyond the basics of classification, clustering, and recommendation. Pro Hadoop Data Analytics emphasizes best practices to ensure coherent, efficient development. A complete example system will be developed using standard third-party components that consist of the tool kits, libraries, visualization and reporting code, as well as support glue to provide a working and extensible end-to-end system. The book also highlights the importance of end-to-end, flexible, configurable, high-performance data pipeline systems with analytical components as well as appropriate visualization results. You'll discover the importance of mix-and-match or hybrid systems, using different analytical components in one application. This hybrid approach will be prominent in the examples. What You'll Learn Build big data analytic systems with the Hadoop ecosystem Use libraries, tool kits, and algorithms to make development easier and more effective Apply metrics to measure performance and efficiency of components and systems Connect to standard relational databases, noSQL data sources, and more Follow case studies with example components to create your own systems Who This Book Is For Software engineers, architects, and data scientists with an interest in the design and implementation of big data analytical systems using Hadoop, the Hadoop ecosystem, and other associated technologies.

Rise of the Data Cloud

AuthorHouse The rise of the Data Cloud is ushering in a new era of computing. The world's digital data is mass migrating to the cloud, where it can be more effectively integrated, managed, and mobilized. The data cloud eliminates data siloes and enables data sharing with business partners, capitalizing on data network effects. It democratizes data analytics, making the most sophisticated data science tools accessible to organizations of all sizes. Data exchanges enable businesses to discover, explore, and easily purchase or sell data—opening up new revenue streams. Business leaders have long dreamed of data driving their organizations. Now, thanks to the Data Cloud, nothing stands in their way.

Practical Enterprise Data Lake Insights

Handle Data-Driven Challenges in an Enterprise Big Data Lake

Apress 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

Proceedings of International Conference on Deep Learning, Computing and Intelligence

ICDCI 2021

Springer Nature

DAMA-DMBOK

Data Management Body of Knowledge

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.