

architecting modern data platforms pdf

architecting modern data platforms pdf is a crucial resource for professionals and organizations aiming to design scalable, efficient, and flexible data infrastructures. This article explores the core principles and best practices detailed in such comprehensive guides, emphasizing the importance of a robust architecture in managing big data, cloud integration, and real-time analytics. Modern data platforms require a combination of advanced technologies, strategic planning, and a deep understanding of data workflows to support business intelligence and operational needs effectively. By examining the structure and components commonly addressed in architecting modern data platforms pdf documents, readers can gain insights into data ingestion, storage solutions, processing frameworks, and governance mechanisms. This article also highlights the role of cloud-native architectures, data lakes, and data warehouses in creating a unified and agile data ecosystem. The following sections break down these topics systematically, providing a detailed overview that aligns with the latest trends and technological advancements in the data platform domain.

- Understanding Modern Data Platform Architecture
- Key Components of Modern Data Platforms
- Design Principles for Scalable Data Platforms
- Cloud Integration and Deployment Strategies
- Data Governance and Security Considerations
- Case Studies and Practical Applications

Understanding Modern Data Platform Architecture

Modern data platform architectures are designed to handle the growing volume, variety, and velocity of data generated by enterprises today. These architectures integrate multiple data sources and technologies to provide a cohesive framework that supports advanced analytics, machine learning, and real-time data processing. Architecting modern data platforms pdf documents typically emphasize the shift from traditional monolithic data warehouses to flexible, modular platforms that can adapt to changing business requirements.

The Evolution of Data Platforms

Data platforms have evolved from simple storage solutions to complex

ecosystems involving data lakes, cloud services, and streaming technologies. Architecting modern data platforms pdf resources often trace this evolution to highlight how modern platforms address limitations such as scalability constraints, data silos, and latency issues found in legacy systems.

Core Architectural Layers

A typical modern data platform includes distinct layers for data ingestion, storage, processing, and consumption. Each layer is optimized to handle specific tasks, ensuring efficient data flow and management across the platform. Architecting modern data platforms pdf materials frequently outline these layers to guide architects in building robust infrastructures.

Key Components of Modern Data Platforms

The architecture of modern data platforms comprises several essential components that work together to enable comprehensive data management and analytics. These components are detailed extensively in architecting modern data platforms pdf guides, providing a blueprint for implementation.

Data Ingestion and Integration

Data ingestion involves collecting data from various sources such as databases, IoT devices, social media, and transactional systems. Modern platforms support batch and real-time ingestion to cater to diverse use cases. Tools and frameworks for data integration ensure seamless aggregation and transformation of raw data.

Data Storage Solutions

Storage is a critical component, often involving a combination of data lakes and data warehouses. Data lakes store raw, unstructured data at scale, while data warehouses hold structured, curated data optimized for querying. Architecting modern data platforms pdf documents emphasize selecting storage technologies based on performance, cost, and scalability.

Data Processing and Analytics

Processing frameworks such as Apache Spark, Flink, or cloud-native services enable batch and stream processing of data. These systems support analytical workloads, machine learning model training, and real-time monitoring. Architecting modern data platforms pdf resources highlight best practices for choosing and configuring processing engines.

Data Consumption and Visualization

Data consumption layers include business intelligence tools, dashboards, and APIs that deliver insights to end-users and applications. Ensuring accessibility and usability of data is a key focus within modern data platform design.

Design Principles for Scalable Data Platforms

Scalability is paramount in architecting modern data platforms pdf discussions, as platforms must accommodate increasing data volumes and user demands without performance degradation. Several design principles help achieve this goal.

Modularity and Microservices

Breaking the platform into modular components or microservices allows independent development, deployment, and scaling. This approach enhances flexibility and reduces system complexity.

Decoupling and Event-Driven Architecture

Decoupling components through event-driven mechanisms facilitates asynchronous processing and improves system resilience. Architecting modern data platforms pdf resources often recommend event buses or messaging systems like Kafka for this purpose.

Automation and Infrastructure as Code

Automating deployment and management using infrastructure as code tools ensures consistency and repeatability, crucial for scaling and maintaining modern data platforms.

Performance Optimization

Optimization techniques such as data partitioning, indexing, and caching are vital to achieving low latency and high throughput in data processing and querying.

Cloud Integration and Deployment Strategies

Cloud platforms provide the foundation for many modern data platforms, offering scalability, flexibility, and cost-efficiency. Architecting modern

data platforms pdf materials extensively cover cloud deployment models and integration strategies.

Public, Private, and Hybrid Clouds

Selecting between public, private, or hybrid cloud environments depends on factors like security requirements, data sovereignty, and cost considerations. Modern data platforms often leverage hybrid models to balance control and scalability.

Cloud-Native Services and Tools

Utilizing cloud-native services such as managed databases, serverless compute, and data pipelines simplifies platform architecture and accelerates development.

Multi-Cloud and Vendor Lock-In Avoidance

Designing platforms to operate across multiple cloud providers reduces dependency on a single vendor and enhances disaster recovery capabilities.

Data Governance and Security Considerations

Effective governance and security are critical in architecting modern data platforms pdf frameworks to ensure data integrity, compliance, and protection against threats.

Data Quality and Lineage

Implementing data quality checks and tracking data lineage helps maintain trustworthy data and supports audit requirements.

Access Control and Privacy

Robust access control mechanisms, including role-based access control (RBAC) and data masking, safeguard sensitive information and comply with privacy regulations.

Compliance and Regulatory Standards

Platforms must adhere to standards such as GDPR, HIPAA, or CCPA, necessitating comprehensive policies and technical controls outlined in

architecting modern data platforms pdf guidelines.

Case Studies and Practical Applications

Real-world implementations of modern data platforms illustrate the application of architectural principles and highlight challenges and solutions encountered in diverse industries.

Enterprise Data Platform Transformation

Many organizations undergo digital transformation by adopting modern data platforms to improve decision-making, operational efficiency, and customer engagement.

Analytics and Machine Learning Integration

Integrating analytics and machine learning workflows into data platforms enables predictive insights and automation, driving business value.

Industry-Specific Use Cases

Different sectors such as finance, healthcare, and retail apply tailored data platform architectures to meet unique data needs and compliance requirements.

- Improved data accessibility and collaboration
- Enhanced scalability and cost management
- Accelerated innovation through real-time insights
- Strengthened data security and compliance

Frequently Asked Questions

What is the 'Architecting Modern Data Platforms' PDF about?

The 'Architecting Modern Data Platforms' PDF provides comprehensive guidance on designing scalable, flexible, and efficient data platforms using modern technologies and best practices.

Who is the target audience for the 'Architecting Modern Data Platforms' PDF?

The target audience includes data architects, data engineers, IT professionals, and business stakeholders involved in building and managing data platforms.

What key components are covered in the 'Architecting Modern Data Platforms' PDF?

Key components include data ingestion, storage, processing, security, governance, and integration with analytics and machine learning tools.

Does the 'Architecting Modern Data Platforms' PDF discuss cloud-based solutions?

Yes, the PDF discusses various cloud-based architectures and services that enable modern data platform implementations, including hybrid and multi-cloud strategies.

Are best practices for data governance included in the 'Architecting Modern Data Platforms' PDF?

Yes, the document emphasizes data governance best practices to ensure data quality, compliance, and security throughout the data lifecycle.

Where can I download the 'Architecting Modern Data Platforms' PDF?

The PDF is typically available from official vendor websites, technology blogs, or platforms like Microsoft Docs or AWS Architecture Center depending on the source of the publication.

Additional Resources

1. *Architecting Modern Data Platforms: A Guide to Enterprise Hadoop at Scale*
This book offers a comprehensive overview of building scalable and efficient data platforms using Hadoop and related technologies. It covers architectural principles, design patterns, and best practices for managing big data workloads. Readers will learn how to integrate various tools to create a robust data ecosystem for enterprise needs.

2. Designing Data-Intensive Applications

Written by Martin Kleppmann, this book dives into the architecture of scalable and maintainable data systems. It explains the fundamentals of data storage, retrieval, and processing with a focus on distributed systems. The

book is essential for understanding how modern data platforms handle reliability, scalability, and fault tolerance.

3. Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success

This guide is tailored for researchers but provides valuable insights into data organization and management relevant to data platform architects. It discusses data lifecycle, metadata, and best practices for data sharing and preservation. The book helps readers appreciate the importance of structured data management in any platform.

4. Building Data Streaming Applications with Apache Kafka

This practical book focuses on architecting real-time data platforms using Apache Kafka. It explains event-driven architecture and how to design systems for processing streaming data efficiently. Readers will get hands-on examples and strategies for handling data ingestion, processing, and analytics in modern data environments.

5. Cloud Data Management

This title explores the principles and technologies behind managing data in cloud environments. It covers cloud storage, databases, and data processing frameworks that are essential components of modern data platforms. The book also discusses security, compliance, and cost optimization strategies for cloud-based data architectures.

6. Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump

This book addresses the challenges of building and maintaining effective data lakes. It offers architectural guidelines to ensure data quality, governance, and usability within large-scale data repositories. The author emphasizes the importance of metadata management and data lifecycle policies to avoid common pitfalls.

7. Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing

An in-depth exploration of streaming data architectures, this book covers the theory and practice behind processing continuous data streams. It discusses frameworks like Apache Flink, Apache Beam, and Spark Streaming, illustrating how to build resilient and low-latency data platforms. The content is suitable for architects aiming to leverage real-time insights.

8. Enterprise Data Architecture: How to Navigate Its Landscape

This book provides a strategic approach to designing enterprise-wide data architectures. It integrates concepts from data governance, integration, and analytics to build cohesive data platforms. Readers will find frameworks and methodologies for aligning data architecture with business goals and technology trends.

9. Big Data Fundamentals: Concepts, Drivers & Techniques

A foundational book that introduces the core concepts and technologies driving big data platforms. It explains data types, storage solutions,

processing techniques, and analytics tools pivotal to modern data ecosystems. The clear and concise explanations make it a great starting point for professionals entering the field of data platform architecture.

[Architecting Modern Data Platforms Pdf](#)

Find other PDF articles:

<https://a.comtex-nj.com/wwu3/pdf?dataid=kQL44-8408&title=cashier-math-test-pdf.pdf>

Architecting Modern Data Platforms PDF

Unleash the Power of Your Data: Build a Future-Proof Data Platform Today.

Are you drowning in data, struggling to extract meaningful insights, and facing escalating costs from inefficient data management? Building and maintaining a robust, scalable data platform can feel like navigating a labyrinth. Legacy systems are creaking under the strain of ever-increasing data volumes, cloud costs are spiraling, and your team lacks the tools and expertise to effectively manage it all. You need a clear roadmap to modernize your data infrastructure and unlock the true potential of your data assets.

This comprehensive guide, "Architecting Modern Data Platforms," provides the blueprint you need.

Inside, you'll discover:

Introduction: Setting the Stage for Modern Data Platforms

Chapter 1: Defining Your Data Strategy: Understanding your business needs and aligning them with your data architecture.

Chapter 2: Choosing the Right Cloud Provider: Navigating the AWS, Azure, and GCP landscape for optimal data solutions.

Chapter 3: Designing Your Data Lake and Data Warehouse: Building a robust and scalable foundation for data storage and processing.

Chapter 4: Implementing Data Ingestion and Transformation: Efficiently moving and transforming data for analysis.

Chapter 5: Data Governance and Security: Ensuring data quality, compliance, and protecting sensitive information.

Chapter 6: Building a Modern Data Pipeline: Automating data workflows for increased efficiency and reliability.

Chapter 7: Data Visualization and Business Intelligence: Transforming data into actionable insights.

Chapter 8: Monitoring and Optimization: Maintaining performance and proactively addressing issues.

Conclusion: The Future of Modern Data Platforms and Next Steps

Architecting Modern Data Platforms: A Comprehensive Guide

Introduction: Setting the Stage for Modern Data Platforms

The modern business landscape is defined by data. Organizations of all sizes are grappling with unprecedented volumes of data from diverse sources – transactions, social media, IoT devices, and more. Effectively managing, processing, and analyzing this data is no longer a luxury; it's a necessity for survival and growth. Traditional data warehouses and data lakes are struggling to keep pace. This necessitates a shift toward modern data platforms – flexible, scalable, and cost-effective solutions designed to handle the demands of the big data era. This guide provides a roadmap for architecting a modern data platform tailored to your specific business needs.

Chapter 1: Defining Your Data Strategy: Aligning Business Needs with Data Architecture

Before diving into technical details, a clear data strategy is paramount. This involves understanding your business objectives and translating them into concrete data requirements. Key questions to ask include:

What are your key business goals? Understanding these goals (e.g., improving customer retention, optimizing supply chains, launching new products) dictates the type of data you need to collect and analyze.

What data do you need to achieve these goals? Identify the sources, types, and volume of data required.

Who are your stakeholders? Understand who needs access to the data and what insights they require.

What are your data governance requirements? Consider compliance regulations (GDPR, HIPAA, etc.) and data security policies.

A well-defined data strategy serves as the foundation for your entire data platform architecture. It ensures that your technical choices align with your business objectives, avoiding costly misalignments down the line. This stage involves significant stakeholder engagement and requires a clear understanding of both business and technical considerations.

Chapter 2: Choosing the Right Cloud Provider:

Navigating the AWS, Azure, and GCP Landscape

The cloud offers unparalleled scalability, flexibility, and cost-effectiveness for modern data platforms. However, choosing the right cloud provider (AWS, Azure, or GCP) requires careful evaluation based on your specific needs.

AWS (Amazon Web Services): Known for its extensive range of services, mature ecosystem, and large community support. Its strengths lie in its breadth of offerings, making it suitable for complex data architectures.

Azure (Microsoft Azure): Offers strong integration with Microsoft products and services, making it attractive for organizations already invested in the Microsoft ecosystem. Its focus on hybrid cloud deployments makes it a good choice for organizations with on-premises infrastructure.

GCP (Google Cloud Platform): Known for its powerful data analytics capabilities and innovative technologies like BigQuery. Its strengths lie in its scalability and performance for large-scale data processing.

The decision should consider factors like existing infrastructure, budget, technical expertise, and specific service requirements. Often, a hybrid approach combining multiple cloud providers or a combination of cloud and on-premises solutions is the optimal strategy.

Chapter 3: Designing Your Data Lake and Data Warehouse: Building a Robust Foundation

Your data platform needs a robust foundation for storing and processing data. This typically involves a combination of a data lake and a data warehouse.

Data Lake: A centralized repository for storing raw data in its native format. It provides the flexibility to explore and analyze diverse data types without upfront schema definition. Popular choices include AWS S3, Azure Data Lake Storage, and Google Cloud Storage.

Data Warehouse: A structured repository designed for analytical processing. Data is organized into a schema-defined format, optimized for querying and reporting. Popular choices include Snowflake, Amazon Redshift, Azure Synapse Analytics, and Google BigQuery.

The choice of technologies will depend on your data volume, velocity, variety, and veracity (the 4 Vs of big data), as well as your analytical needs. Consider factors like query performance, scalability, and cost when selecting your data lake and data warehouse solutions.

Chapter 4: Implementing Data Ingestion and Transformation: Efficiently Moving and Transforming Data

Efficiently moving data from various sources into your data lake and data warehouse is crucial. This involves implementing robust data ingestion pipelines that handle diverse data formats and volumes. Key considerations include:

Data Sources: Identify all data sources (databases, APIs, files, streaming platforms, etc.).

Data Integration Tools: Choose tools to handle data ingestion (e.g., Apache Kafka, Apache NiFi, AWS Glue, Azure Data Factory, Google Cloud Dataflow).

Data Transformation: Implement processes to cleanse, transform, and enrich data before loading it into your data warehouse. This might involve ETL (Extract, Transform, Load) or ELT (Extract, Load, Transform) processes using tools like Apache Spark, SQL, or cloud-based services.

A well-designed ingestion and transformation pipeline ensures data quality and consistency, paving the way for reliable analytics. Automation is key to optimizing efficiency and reducing manual intervention.

Chapter 5: Data Governance and Security: Ensuring Data Quality, Compliance, and Protection

Data governance and security are paramount in a modern data platform. This involves implementing policies and procedures to ensure data quality, compliance with regulations, and protection against unauthorized access.

Data Quality: Implement processes to monitor and improve data accuracy, completeness, and consistency.

Data Security: Implement robust security measures such as access control, encryption, and data loss prevention (DLP) to protect sensitive information.

Compliance: Ensure compliance with relevant regulations (GDPR, HIPAA, CCPA, etc.).

Data governance requires a well-defined framework, clear roles and responsibilities, and the implementation of appropriate tools and technologies.

Chapter 6: Building a Modern Data Pipeline: Automating Data Workflows

Automating data workflows using a modern data pipeline is crucial for efficiency and scalability. This involves orchestrating the entire data lifecycle, from ingestion to transformation to loading and analysis. Consider using tools like Apache Airflow, Prefect, or cloud-native services such as AWS Step Functions, Azure Logic Apps, or Google Cloud Composer. A well-designed pipeline enables:

Increased Efficiency: Automating repetitive tasks reduces manual effort and human error.

Scalability: Handles increasing data volumes and processing demands.

Improved Reliability: Ensures data consistency and reduces downtime.

Reproducibility: Easily reproduce data processing steps for consistency.

Chapter 7: Data Visualization and Business Intelligence: Transforming Data into Actionable Insights

The ultimate goal of a modern data platform is to transform raw data into actionable insights. This requires implementing robust data visualization and business intelligence (BI) tools. Popular choices include Tableau, Power BI, Qlik Sense, or cloud-native solutions like Amazon QuickSight, Azure Analysis Services, or Google Looker. Effective data visualization enables:

Improved Decision-Making: Provides clear and concise insights that inform strategic decisions.

Enhanced Communication: Communicates complex data effectively to stakeholders.

Increased Collaboration: Enables teams to collaborate on data analysis and insights.

Chapter 8: Monitoring and Optimization: Maintaining Performance and Proactively Addressing Issues

Ongoing monitoring and optimization are essential for maintaining the performance and stability of your data platform. This involves implementing monitoring tools to track key metrics, identify bottlenecks, and proactively address issues. Key areas to monitor include:

Data Ingestion: Monitor data volume, velocity, and latency.

Data Processing: Track processing time, resource utilization, and error rates.

Data Storage: Monitor storage capacity, costs, and performance.

Data Quality: Track data accuracy, completeness, and consistency.

Proactive monitoring enables timely identification and resolution of problems, preventing disruptions and ensuring the smooth operation of your data platform.

Conclusion: The Future of Modern Data Platforms and Next Steps

Modern data platforms are evolving rapidly, driven by innovations in cloud computing, big data technologies, and artificial intelligence. The future will likely see increased adoption of serverless architectures, real-time data processing, and advanced analytics techniques like machine learning. The key takeaway is to build a flexible and scalable platform that can adapt to the ever-changing needs of your business. By following the steps outlined in this guide, you can build a foundation for data-driven decision-making and unlock the full potential of your data assets.

FAQs:

1. What is the difference between a data lake and a data warehouse? A data lake stores raw data in its native format, while a data warehouse stores structured data optimized for analysis.
2. Which cloud provider is best for my business? The best provider depends on your specific needs, existing infrastructure, and budget.
3. How can I ensure data security in my modern data platform? Implement robust security measures such as access control, encryption, and data loss prevention (DLP).
4. What are the key benefits of automating data workflows? Automation increases efficiency, scalability, reliability, and reproducibility.
5. What tools can I use for data visualization and BI? Popular options include Tableau, Power BI, Qlik Sense, and cloud-native solutions.
6. How can I monitor the performance of my data platform? Implement monitoring tools to track key metrics and proactively address issues.
7. What are the common challenges in architecting a modern data platform? Challenges include data volume, velocity, variety, veracity, cost, security, and integration.
8. How can I ensure data quality in my platform? Implement processes for data cleansing, transformation, and validation.
9. What are the future trends in modern data platforms? Serverless architectures, real-time processing, and AI-powered analytics are key trends.

Related Articles:

1. Building a Serverless Data Lake on AWS: A guide to building a scalable and cost-effective data lake using serverless technologies on AWS.
2. Data Governance Best Practices for Modern Data Platforms: A deep dive into establishing effective data governance policies and procedures.
3. Optimizing Cloud Costs for Your Data Platform: Strategies for reducing cloud spending while maintaining performance.
4. Implementing Real-Time Data Processing with Apache Kafka: A tutorial on using Apache Kafka for real-time data ingestion and processing.
5. Choosing the Right Data Warehouse for Your Business: A comparison of popular data warehouse solutions and how to choose the best one for your needs.
6. Data Security and Compliance in the Cloud: A comprehensive guide to securing your data in the

cloud and meeting compliance requirements.

7. Advanced Analytics with Machine Learning on BigQuery: Exploring the use of machine learning for advanced analytics on Google BigQuery.

8. Automating Data Pipelines with Apache Airflow: A practical guide to building automated data pipelines using Apache Airflow.

9. Data Lakehouse Architecture: Combining the Best of Data Lakes and Data Warehouses: An exploration of the data lakehouse architectural pattern.

architecting modern data platforms pdf: Architecting Modern Data Platforms Jan

Kunigk, Ian Buss, Paul Wilkinson, Lars George, 2018-12-05 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

architecting modern data platforms pdf: Architecting Modern Data Platforms Jan Kunigk,

Ian Buss, Paul Wilkinson, Lars George, 2018 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.

architecting modern data platforms pdf: *Designing Cloud Data Platforms* Danil Zburivsky,

Lynda Partner, 2021-04-20 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 an 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 analyse it. about the technology Access to affordable, dependable, serverless cloud services has revolutionized the way organizations can approach data management, and companies both big and small are raring to migrate to the cloud. But without a properly designed data platform, data in the cloud can remain just as siloed and inaccessible as it is today for most organizations. Designing Cloud Data Platforms

lays out the principles of a well-designed platform that uses the scalable resources of the public cloud to manage all of an organization's data, and present it as useful business insights. about the book In Designing Cloud Data Platforms, you'll learn how to integrate data from multiple sources into a single, cloud-based, modern data platform. Drawing on their real-world experiences designing cloud data platforms for dozens of organizations, cloud data experts Danil Zburivsky and Lynda Partner take you through a six-layer approach to creating cloud data platforms that maximizes flexibility and manageability and reduces costs. Starting with foundational principles, you'll learn how to get data into your platform from different databases, files, and APIs, the essential practices for organizing and processing that raw data, and how to best take advantage of the services offered by major cloud vendors. As you progress past the basics you'll take a deep dive into advanced topics to get the most out of your data platform, including real-time data management, machine learning analytics, schema management, and more. what's inside The tools of different public cloud for implementing data platforms Best practices for managing structured and unstructured data sets Machine learning tools that can be used on top of the cloud Cost optimization techniques about the reader For data professionals familiar with the basics of cloud computing and distributed data processing systems like Hadoop and Spark. about the authors Danil Zburivsky has over 10 years 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.

architecting modern data platforms pdf: Foundations for Architecting Data Solutions Ted Malaska, Jonathan Seidman, 2018-08-29 While many companies ponder implementation details such as distributed processing engines and algorithms for data analysis, this practical book takes a much wider view of big data development, starting with initial planning and moving diligently toward execution. Authors Ted Malaska and Jonathan Seidman guide you through the major components necessary to start, architect, and develop successful big data projects. Everyone from CIOs and COOs to lead architects and developers will explore a variety of big data architectures and applications, from massive data pipelines to web-scale applications. Each chapter addresses a piece of the software development life cycle and identifies patterns to maximize long-term success throughout the life of your project. Start the planning process by considering the key data project types Use guidelines to evaluate and select data management solutions Reduce risk related to technology, your team, and vague requirements Explore system interface design using APIs, REST, and pub/sub systems Choose the right distributed storage system for your big data system Plan and implement metadata collections for your data architecture Use data pipelines to ensure data integrity from source to final storage Evaluate the attributes of various engines for processing the data you collect

architecting modern data platforms pdf: Data Mesh Zhamak Dehghani, 2022-03-08 Many enterprises are investing in a next-generation data lake, hoping to democratize data at scale to provide business insights and ultimately make automated intelligent decisions. In this practical book, author Zhamak Dehghani reveals that, despite the time, money, and effort poured into them, data warehouses and data lakes fail when applied at the scale and speed of today's organizations. A distributed data mesh is a better choice. Dehghani guides architects, technical leaders, and decision makers on their journey from monolithic big data architecture to a sociotechnical paradigm that draws from modern distributed architecture. A data mesh considers domains as a first-class concern, applies platform thinking to create self-serve data infrastructure, treats data as a product, and introduces a federated and computational model of data governance. This book shows you why and how. Examine the current data landscape from the perspective of business and organizational needs, environmental challenges, and existing architectures Analyze the landscape's underlying characteristics and failure modes Get a complete introduction to data mesh principles and its constituents Learn how to design a data mesh architecture Move beyond a monolithic data lake to a distributed data mesh.

architecting modern data platforms pdf: Designing Data-Intensive Applications Martin

Kleppmann, 2017-03-16 Data is at the center of many challenges in system design today. Difficult issues need to be figured out, such as scalability, consistency, reliability, efficiency, and maintainability. In addition, we have an overwhelming variety of tools, including relational databases, NoSQL datastores, stream or batch processors, and message brokers. What are the right choices for your application? How do you make sense of all these buzzwords? In this practical and comprehensive guide, author Martin Kleppmann helps you navigate this diverse landscape by examining the pros and cons of various technologies for processing and storing data. Software keeps changing, but the fundamental principles remain the same. With this book, software engineers and architects will learn how to apply those ideas in practice, and how to make full use of data in modern applications. Peer under the hood of the systems you already use, and learn how to use and operate them more effectively Make informed decisions by identifying the strengths and weaknesses of different tools Navigate the trade-offs around consistency, scalability, fault tolerance, and complexity Understand the distributed systems research upon which modern databases are built Peek behind the scenes of major online services, and learn from their architectures

architecting modern data platforms pdf: Modern Data Architecture on AWS Behram Irani, 2023-08-31 Discover all the essential design and architectural patterns in one place to help you rapidly build and deploy your modern data platform using AWS services Key Features Learn to build modern data platforms on AWS using data lakes and purpose-built data services Uncover methods of applying security and governance across your data platform built on AWS Find out how to operationalize and optimize your data platform on AWS Purchase of the print or Kindle book includes a free PDF eBook Book Description Many IT leaders and professionals are adept at extracting data from a particular type of database and deriving value from it. However, designing and implementing an enterprise-wide holistic data platform with purpose-built data services, all seamlessly working in tandem with the least amount of manual intervention, still poses a challenge. This book will help you explore end-to-end solutions to common data, analytics, and AI/ML use cases by leveraging AWS services. The chapters systematically take you through all the building blocks of a modern data platform, including data lakes, data warehouses, data ingestion patterns, data consumption patterns, data governance, and AI/ML patterns. Using real-world use cases, each chapter highlights the features and functionalities of numerous AWS services to enable you to create a scalable, flexible, performant, and cost-effective modern data platform. By the end of this book, you'll be equipped with all the necessary architectural patterns and be able to apply this knowledge to efficiently build a modern data platform for your organization using AWS services. What you will learn Familiarize yourself with the building blocks of modern data architecture on AWS Discover how to create an end-to-end data platform on AWS Design data architectures for your own use cases using AWS services Ingest data from disparate sources into target data stores on AWS Build data pipelines, data sharing mechanisms, and data consumption patterns using AWS services Find out how to implement data governance using AWS services Who this book is for This book is for data architects, data engineers, and professionals creating data platforms. The book's use case-driven approach helps you conceptualize possible solutions to specific use cases, while also providing you with design patterns to build data platforms for any organization. It's beneficial for technical leaders and decision makers to understand their organization's data architecture and how each platform component serves business needs. A basic understanding of data & analytics architectures and systems is desirable along with beginner's level understanding of AWS Cloud.

architecting modern data platforms pdf: Architecting Modern Java EE Applications Sebastian Daschner, 2017-10-09 Find out how to craft effective, business-oriented Java EE 8 applications that target customer's demands in the age of Cloud platforms and container technology. About This Book Understand the principles of modern Java EE and how to realize effective architectures Gain knowledge of how to design enterprise software in the age of automation, Continuous Delivery and Cloud platforms Learn about the reasoning and motivations behind state-of-the-art enterprise Java technology, that focuses on business Who This Book Is For This book is for experienced Java EE developers who are aspiring to become the architects of enterprise-grade

applications, or software architects who would like to leverage Java EE to create effective blueprints of applications. What You Will Learn What enterprise software engineers should focus on Implement applications, packages, and components in a modern way Design and structure application architectures Discover how to realize technical and cross-cutting aspects Get to grips with containers and container orchestration technology Realize zero-dependency, 12-factor, and Cloud-native applications Implement automated, fast, reliable, and maintainable software tests Discover distributed system architectures and their requirements In Detail Java EE 8 brings with it a load of features, mainly targeting newer architectures such as microservices, modernized security APIs, and cloud deployments. This book will teach you to design and develop modern, business-oriented applications using Java EE 8. It shows how to structure systems and applications, and how design patterns and Domain Driven Design aspects are realized in the age of Java EE 8. You will learn about the concepts and principles behind Java EE applications, and how to effect communication, persistence, technical and cross-cutting concerns, and asynchronous behavior. This book covers Continuous Delivery, DevOps, infrastructure-as-code, containers, container orchestration technologies, such as Docker and Kubernetes, and why and especially how Java EE fits into this world. It also covers the requirements behind containerized, zero-dependency applications and how modern Java EE application servers support these approaches. You will also learn about automated, fast, and reliable software tests, in different test levels, scopes, and test technologies. This book covers the prerequisites and challenges of distributed systems that lead to microservice, shared-nothing architectures. The challenges and solutions of consistency versus scalability will further lead us to event sourcing, event-driven architectures, and the CQRS principle. This book also includes the nuts and bolts of application performance as well as how to realize resilience, logging, monitoring and tracing in a modern enterprise world. Last but not least the demands of securing enterprise systems are covered. By the end, you will understand the ins and outs of Java EE so that you can make critical design decisions that not only live up to, but also surpass your clients' expectations. Style and approach This book focuses on solving business problems and meeting customer demands in the enterprise world. It covers how to create enterprise applications with reasonable technology choices, free of cargo-cult and over-engineering. The aspects shown in this book not only demonstrate how to realize a certain solution, but also explain its motivations and reasoning.

architecting modern data platforms pdf: BUILDING a MODERN DATA CENTER Principles and Strategies of Design Scott D. Lowe, 2016-01-10

architecting modern data platforms pdf: The Art of Systems Architecting Mark W. Maier, 2009-01-06 If engineering is the art and science of technical problem solving, systems architecting happens when you don't yet know what the problem is. The third edition of a highly respected bestseller, The Art of Systems Architecting provides in-depth coverage of the least understood part of systems design: moving from a vague concept and limited resources

architecting modern data platforms pdf: *Data Lake Architecture* Bill Inmon, 2016 Data Lake Architecture will explain how to build a useful data lake, where data scientists and data analysts can solve business challenges and identify new business opportunities

architecting modern data platforms pdf: **HBase in Action** Amandeep Khurana, Nick Dimiduk, 2012-11-01 Summary HBase in Action has all the knowledge you need to design, build, and run applications using HBase. First, it introduces you to the fundamentals of distributed systems and large scale data handling. Then, you'll explore real-world applications and code samples with just enough theory to understand the practical techniques. You'll see how to build applications with HBase and take advantage of the MapReduce processing framework. And along the way you'll learn patterns and best practices. About the Technology HBase is a NoSQL storage system designed for fast, random access to large volumes of data. It runs on commodity hardware and scales smoothly from modest datasets to billions of rows and millions of columns. About this Book HBase in Action is an experience-driven guide that shows you how to design, build, and run applications using HBase. First, it introduces you to the fundamentals of handling big data. Then, you'll explore HBase with the

help of real applications and code samples and with just enough theory to back up the practical techniques. You'll take advantage of the MapReduce processing framework and benefit from seeing HBase best practices in action. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. What's Inside When and how to use HBase Practical examples Design patterns for scalable data systems Deployment, integration, and design Written for developers and architects familiar with data storage and processing. No prior knowledge of HBase, Hadoop, or MapReduce is required. Table of Contents PART 1 HBASE FUNDAMENTALS Introducing HBase Getting started Distributed HBase, HDFS, and MapReduce PART 2 ADVANCED CONCEPTS HBase table design Extending HBase with coprocessors Alternative HBase clients PART 3 EXAMPLE APPLICATIONS HBase by example: OpenTSDB Scaling GIS on HBase PART 4 OPERATIONALIZING HBASE Deploying HBase Operations

architecting modern data platforms pdf: Cloud Computing Thomas Erl, Ricardo Puttini, Zaigham Mahmood, 2013 This book describes cloud computing as a service that is highly scalable and operates in a resilient environment. The authors emphasize architectural layers and models - but also business and security factors.

architecting modern data platforms pdf: Data Lakehouse in Action Pradeep Menon, 2022-03-17 Propose a new scalable data architecture paradigm, Data Lakehouse, that addresses the limitations of current data architecture patterns Key Features Understand how data is ingested, stored, served, governed, and secured for enabling data analytics Explore a practical way to implement Data Lakehouse using cloud computing platforms like Azure Combine multiple architectural patterns based on an organization's needs and maturity level Book Description The Data Lakehouse architecture is a new paradigm that enables large-scale analytics. This book will guide you in developing data architecture in the right way to ensure your organization's success. The first part of the book discusses the different data architectural patterns used in the past and the need for a new architectural paradigm, as well as the drivers that have caused this change. It covers the principles that govern the target architecture, the components that form the Data Lakehouse architecture, and the rationale and need for those components. The second part deep dives into the different layers of Data Lakehouse. It covers various scenarios and components for data ingestion, storage, data processing, data serving, analytics, governance, and data security. The book's third part focuses on the practical implementation of the Data Lakehouse architecture in a cloud computing platform. It focuses on various ways to combine the Data Lakehouse pattern to realize macro-patterns, such as Data Mesh and Data Hub-Spoke, based on the organization's needs and maturity level. The frameworks introduced will be practical and organizations can readily benefit from their application. By the end of this book, you'll clearly understand how to implement the Data Lakehouse architecture pattern in a scalable, agile, and cost-effective manner. What you will learn Understand the evolution of the Data Architecture patterns for analytics Become well versed in the Data Lakehouse pattern and how it enables data analytics Focus on methods to ingest, process, store, and govern data in a Data Lakehouse architecture Learn techniques to serve data and perform analytics in a Data Lakehouse architecture Cover methods to secure the data in a Data Lakehouse architecture Implement Data Lakehouse in a cloud computing platform such as Azure Combine Data Lakehouse in a macro-architecture pattern such as Data Mesh Who this book is for This book is for data architects, big data engineers, data strategists and practitioners, data stewards, and cloud computing practitioners looking to become well-versed with modern data architecture patterns to enable large-scale analytics. Basic knowledge of data architecture and familiarity with data warehousing concepts are required.

architecting modern data platforms pdf: Software Architecture: The Hard Parts Neal Ford, Mark Richards, Pramod Sadalage, Zhamak Dehghani, 2021-09-23 There are no easy decisions in software architecture. Instead, there are many hard parts--difficult problems or issues with no best practices--that force you to choose among various compromises. With this book, you'll learn how to think critically about the trade-offs involved with distributed architectures. Architecture veterans and practicing consultants Neal Ford, Mark Richards, Pramod Sadalage, and Zhamak

Dehghani discuss strategies for choosing an appropriate architecture. By interweaving a story about a fictional group of technology professionals--the Sysops Squad--they examine everything from how to determine service granularity, manage workflows and orchestration, manage and decouple contracts, and manage distributed transactions to how to optimize operational characteristics, such as scalability, elasticity, and performance. By focusing on commonly asked questions, this book provides techniques to help you discover and weigh the trade-offs as you confront the issues you face as an architect. Analyze trade-offs and effectively document your decisions Make better decisions regarding service granularity Understand the complexities of breaking apart monolithic applications Manage and decouple contracts between services Handle data in a highly distributed architecture Learn patterns to manage workflow and transactions when breaking apart applications

architecting modern data platforms pdf: Just Enough Software Architecture George Fairbanks, 2010-08-30 This is a practical guide for software developers, and different than other software architecture books. Here's why: It teaches risk-driven architecting. There is no need for meticulous designs when risks are small, nor any excuse for sloppy designs when risks threaten your success. This book describes a way to do just enough architecture. It avoids the one-size-fits-all process tar pit with advice on how to tune your design effort based on the risks you face. It democratizes architecture. This book seeks to make architecture relevant to all software developers. Developers need to understand how to use constraints as guiderails that ensure desired outcomes, and how seemingly small changes can affect a system's properties. It cultivates declarative knowledge. There is a difference between being able to hit a ball and knowing why you are able to hit it, what psychologists refer to as procedural knowledge versus declarative knowledge. This book will make you more aware of what you have been doing and provide names for the concepts. It emphasizes the engineering. This book focuses on the technical parts of software development and what developers do to ensure the system works not job titles or processes. It shows you how to build models and analyze architectures so that you can make principled design tradeoffs. It describes the techniques software designers use to reason about medium to large sized problems and points out where you can learn specialized techniques in more detail. It provides practical advice. Software design decisions influence the architecture and vice versa. The approach in this book embraces drill-down/pop-up behavior by describing models that have various levels of abstraction, from architecture to data structure design.

architecting modern data platforms pdf: Architecting Modern Data Platforms Jan Kunigk, Ian Buss, Paul Wilkinson, Lars George, 2018-12-05 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

architecting modern data platforms pdf: Architecting Google Cloud Solutions Victor Dantas, 2021-05-14 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 learn Get to grips with compute, storage, networking, data analytics, and pricing Discover delivery models such as IaaS, PaaS, and SaaS Explore the underlying technologies and economics of cloud computing Design for scalability, business continuity, observability, and resiliency Secure Google Cloud solutions and ensure compliance Understand operational best practices and learn how to architect a monitoring solution Gain insights into modern application design with Google Cloud Leverage big data, machine learning, and AI with Google Cloud Who 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.

architecting modern data platforms pdf: Data Lake for Enterprises Tomcy John, Pankaj Misra, 2017-05-31 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.

architecting modern data platforms pdf: Flow Architectures James Urquhart, 2021-01-06 Software development today is embracing events and streaming data, which optimizes not only how technology interacts but also how businesses integrate with one another to meet customer needs. This phenomenon, called flow, consists of patterns and standards that determine which activity and related data is communicated between parties over the internet. This book explores critical implications of that evolution: What happens when events and data streams help you discover new activity sources to enhance existing businesses or drive new markets? What technologies and architectural patterns can position your company for opportunities enabled by flow? James Urquhart, global field CTO at VMware, guides enterprise architects, software developers, and product managers through the process. Learn the benefits of flow dynamics when businesses, governments, and other institutions integrate via events and data streams Understand the value chain for flow integration through Wardley mapping visualization and promise theory modeling Walk through basic concepts behind today's event-driven systems marketplace Learn how today's integration patterns will influence the real-time events flow in the future Explore why companies should architect and build software today to take advantage of flow in coming years

architecting modern data platforms pdf: Fundamentals of Software Architecture Mark Richards, Neal Ford, 2020-01-28 Salary surveys worldwide regularly place software architect in the top 10 best jobs, yet no real guide exists to help developers become architects. Until now. This book provides the first comprehensive overview of software architecture's many aspects. Aspiring and existing architects alike will examine architectural characteristics, architectural patterns, component determination, diagramming and presenting architecture, evolutionary architecture, and many other topics. Mark Richards and Neal Ford—hands-on practitioners who have taught software architecture classes professionally for years—focus on architecture principles that apply across all technology stacks. You'll explore software architecture in a modern light, taking into account all the innovations of the past decade. This book examines: Architecture patterns: The technical basis for many architectural decisions Components: Identification, coupling, cohesion, partitioning, and granularity Soft skills: Effective team management, meetings, negotiation, presentations, and more Modernity: Engineering practices and operational approaches that have changed radically in the past few years Architecture as an engineering discipline: Repeatable results, metrics, and concrete valuations that add rigor to software architecture

architecting modern data platforms pdf: Architecture Francis D. K. Ching, 2012-07-16 A superb visual reference to the principles of architecture Now including interactive CD-ROM! For more than thirty years, the beautifully illustrated *Architecture: Form, Space, and Order* has been the classic introduction to the basic vocabulary of architectural design. The updated Third Edition features expanded sections on circulation, light, views, and site context, along with new considerations of environmental factors, building codes, and contemporary examples of form, space, and order. This classic visual reference helps both students and practicing architects understand the basic vocabulary of architectural design by examining how form and space are ordered in the built environment.? Using his trademark meticulous drawing, Professor Ching shows the relationship between fundamental elements of architecture through the ages and across cultural boundaries. By looking at these seminal ideas, *Architecture: Form, Space, and Order* encourages the reader to look

critically at the built environment and promotes a more evocative understanding of architecture. In addition to updates to content and many of the illustrations, this new edition includes a companion CD-ROM that brings the book's architectural concepts to life through three-dimensional models and animations created by Professor Ching.

architecting modern data platforms pdf: Modern Big Data Processing with Hadoop V
Naresh Kumar, Prashant Shindgikar, 2018-03-30 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.

architecting modern data platforms pdf: The Enterprise Big Data Lake Alex Gorelik, 2019-02-21 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

architecting modern data platforms pdf: Semantic Software Design Eben Hewitt, 2019-09-25 With this practical book, architects, CTOs, and CIOs will learn a set of patterns for the practice of architecture, including analysis, documentation, and communication. Author Eben Hewitt shows you how to create holistic and thoughtful technology plans, communicate them clearly, lead people toward the vision, and become a great architect or Chief Architect. This book covers each key

aspect of architecture comprehensively, including how to incorporate business architecture, information architecture, data architecture, application (software) architecture together to have the best chance for the system's success. Get a practical set of proven architecture practices focused on shipping great products using architecture Learn how architecture works effectively with development teams, management, and product management teams through the value chain Find updated special coverage on machine learning architecture Get usable templates to start incorporating into your teams immediately Incorporate business architecture, information architecture, data architecture, and application (software) architecture together

architecting modern data platforms pdf: Azure Modern Data Architecture Anouar BEN ZAHRA, Key Features Discover the key drivers of successful Azure architecture Practical guidance Focus on scalability and performance Expert authorship Book Description This book presents a guide to design and implement scalable, secure, and efficient data solutions in the Azure cloud environment. It provides Data Architects, developers, and IT professionals who are responsible for designing and implementing data solutions in the Azure cloud environment with the knowledge and tools needed to design and implement data solutions using the latest Azure data services. It covers a wide range of topics, including data storage, data processing, data analysis, and data integration. In this book, you will learn how to select the appropriate Azure data services, design a data processing pipeline, implement real-time data processing, and implement advanced analytics using Azure Databricks and Azure Synapse Analytics. You will also learn how to implement data security and compliance, including data encryption, access control, and auditing. Whether you are building a new data architecture from scratch or migrating an existing on premises solution to Azure, the Azure Data Architecture Guidelines are an essential resource for any organization looking to harness the power of data in the cloud. With these guidelines, you will gain a deep understanding of the principles and best practices of Azure data architecture and be equipped to build data solutions that are highly scalable, secure, and cost effective. What You Need to Use this Book? To use this book, it is recommended that readers have a basic understanding of data architecture concepts and data management principles. Some familiarity with cloud computing and Azure services is also helpful. The book is designed for data architects, data engineers, data analysts, and anyone involved in designing, implementing, and managing data solutions on the Azure cloud platform. It is also suitable for students and professionals who want to learn about Azure data architecture and its best practices.

architecting modern data platforms pdf: Clean Architecture Robert C. Martin, 2017-09-12 Practical Software Architecture Solutions from the Legendary Robert C. Martin ("Uncle Bob") By applying universal rules of software architecture, you can dramatically improve developer productivity throughout the life of any software system. Now, building upon the success of his best-selling books Clean Code and The Clean Coder, legendary software craftsman Robert C. Martin ("Uncle Bob") reveals those rules and helps you apply them. Martin's Clean Architecture doesn't merely present options. Drawing on over a half-century of experience in software environments of every imaginable type, Martin tells you what choices to make and why they are critical to your success. As you've come to expect from Uncle Bob, this book is packed with direct, no-nonsense solutions for the real challenges you'll face—the ones that will make or break your projects. Learn what software architects need to achieve—and core disciplines and practices for achieving it Master essential software design principles for addressing function, component separation, and data management See how programming paradigms impose discipline by restricting what developers can do Understand what's critically important and what's merely a "detail" Implement optimal, high-level structures for web, database, thick-client, console, and embedded applications Define appropriate boundaries and layers, and organize components and services See why designs and architectures go wrong, and how to prevent (or fix) these failures Clean Architecture is essential reading for every current or aspiring software architect, systems analyst, system designer, and software manager—and for every programmer who must execute someone else's designs. Register your product for convenient access to downloads, updates, and/or corrections as they become

available.

architecting modern data platforms pdf: *Building Evolutionary Architectures* Neal Ford, Rebecca Parsons, Patrick Kua, 2017-09-18 The software development ecosystem is constantly changing, providing a constant stream of new tools, frameworks, techniques, and paradigms. Over the past few years, incremental developments in core engineering practices for software development have created the foundations for rethinking how architecture changes over time, along with ways to protect important architectural characteristics as it evolves. This practical guide ties those parts together with a new way to think about architecture and time.

architecting modern data platforms pdf: *Big Data* James Warren, Nathan Marz, 2015-04-29 Summary Big Data teaches you to build big data systems using an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data. It describes a scalable, easy-to-understand approach to big data systems that can be built and run by a small team. Following a realistic example, this book guides readers through the theory of big data systems, how to implement them in practice, and how to deploy and operate them once they're built. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book Web-scale applications like social networks, real-time analytics, or e-commerce sites deal with a lot of data, whose volume and velocity exceed the limits of traditional database systems. These applications require architectures built around clusters of machines to store and process data of any size, or speed. Fortunately, scale and simplicity are not mutually exclusive. Big Data teaches you to build big data systems using an architecture designed specifically to capture and analyze web-scale data. This book presents the Lambda Architecture, a scalable, easy-to-understand approach that can be built and run by a small team. You'll explore the theory of big data systems and how to implement them in practice. In addition to discovering a general framework for processing big data, you'll learn specific technologies like Hadoop, Storm, and NoSQL databases. This book requires no previous exposure to large-scale data analysis or NoSQL tools. Familiarity with traditional databases is helpful. What's Inside Introduction to big data systems Real-time processing of web-scale data Tools like Hadoop, Cassandra, and Storm Extensions to traditional database skills About the Authors Nathan Marz is the creator of Apache Storm and the originator of the Lambda Architecture for big data systems. James Warren is an analytics architect with a background in machine learning and scientific computing. Table of Contents A new paradigm for Big Data PART 1 BATCH LAYER Data model for Big Data Data model for Big Data: Illustration Data storage on the batch layer Data storage on the batch layer: Illustration Batch layer Batch layer: Illustration An example batch layer: Architecture and algorithms An example batch layer: Implementation PART 2 SERVING LAYER Serving layer Serving layer: Illustration PART 3 SPEED LAYER Realtime views Realtime views: Illustration Queuing and stream processing Queuing and stream processing: Illustration Micro-batch stream processing Micro-batch stream processing: Illustration Lambda Architecture in depth

architecting modern data platforms pdf: *Building Digital Experience Platforms* Shailesh Kumar Shivakumar, Sourabh Sethii, 2019-01-31 Use digital experience platforms (DXP) to improve your development productivity and release timelines. Leverage the pre-integrated feature sets of DXPs in your organization's digital transformation journey to quickly develop a personalized, secure, and robust enterprise platform. In this book the authors examine various features of DXPs and provide rich insights into building each layer in a digital platform. Proven best practices are presented with examples for designing and building layers. A special focus is provided on security and quality attributes needed for business-critical enterprise applications. The authors cover modern and emerging digital trends such as Blockchain, IoT, containers, chatbots, artificial intelligence, and more. The book is divided into five parts related to requirements/design, development, security, infrastructure, and case study. The authors employ proven real-world methods, best practices, and security and integration techniques derived from their rich experience. An elaborate digital transformation case study for a banking application is included. What You'll Learn Develop a digital experience platform from end to end Understand best practices and proven methods for designing

overall architecture, user interface and integration components, security, and infrastructureStudy real-world cases, including an elaborate digital transformation building an enterprise platform for a banking applicationKnow the open source tools and technology frameworks that can be used to build DXPs Who This Book Is For Web developers, full stack developers, digital enthusiasts, digital project managers, and architects

architecting modern data platforms pdf: The Software Architect Elevator Gregor Hohpe, 2020-04-08 As the digital economy changes the rules of the game for enterprises, the role of software and IT architects is also transforming. Rather than focus on technical decisions alone, architects and senior technologists need to combine organizational and technical knowledge to effect change in their company's structure and processes. To accomplish that, they need to connect the IT engine room to the penthouse, where the business strategy is defined. In this guide, author Gregor Hohpe shares real-world advice and hard-learned lessons from actual IT transformations. His anecdotes help architects, senior developers, and other IT professionals prepare for a more complex but rewarding role in the enterprise. This book is ideal for: Software architects and senior developers looking to shape the company's technology direction or assist in an organizational transformation Enterprise architects and senior technologists searching for practical advice on how to navigate technical and organizational topics CTOs and senior technical architects who are devising an IT strategy that impacts the way the organization works IT managers who want to learn what's worked and what hasn't in large-scale transformation

architecting modern data platforms pdf: Embedded Systems Architecture Tammy Noergaard, 2012-12-31 Embedded Systems Architecture is a practical and technical guide to understanding the components that make up an embedded system's architecture. This book is perfect for those starting out as technical professionals such as engineers, programmers and designers of embedded systems; and also for students of computer science, computer engineering and electrical engineering. It gives a much-needed 'big picture' for recently graduated engineers grappling with understanding the design of real-world systems for the first time, and provides professionals with a systems-level picture of the key elements that can go into an embedded design, providing a firm foundation on which to build their skills. - Real-world approach to the fundamentals, as well as the design and architecture process, makes this book a popular reference for the daunted or the inexperienced: if in doubt, the answer is in here! - Fully updated with new coverage of FPGAs, testing, middleware and the latest programming techniques in C, plus complete source code and sample code, reference designs and tools online make this the complete package - Visit the companion web site at <http://booksite.elsevier.com/9780123821966/> for source code, design examples, data sheets and more - A true introductory book, provides a comprehensive get up and running reference for those new to the field, and updating skills: assumes no prior knowledge beyond undergrad level electrical engineering - Addresses the needs of practicing engineers, enabling it to get to the point more directly, and cover more ground. Covers hardware, software and middleware in a single volume - Includes a library of design examples and design tools, plus a complete set of source code and embedded systems design tutorial materials from companion website

architecting modern data platforms pdf: Human Dimension and Interior Space Julius Panero, Martin Zelnik, 2014-01-21 The study of human body measurements on a comparative basis is known as anthropometrics. Its applicability to the design process is seen in the physical fit, or interface, between the human body and the various components of interior space. Human Dimension and Interior Space is the first major anthropometrically based reference book of design standards for use by all those involved with the physical planning and detailing of interiors, including interior designers, architects, furniture designers, builders, industrial designers, and students of design. The use of anthropometric data, although no substitute for good design or sound professional judgment should be viewed as one of the many tools required in the design process. This comprehensive overview of anthropometrics consists of three parts. The first part deals with the theory and application of anthropometrics and includes a special section dealing with physically disabled and

elderly people. It provides the designer with the fundamentals of anthropometrics and a basic understanding of how interior design standards are established. The second part contains easy-to-read, illustrated anthropometric tables, which provide the most current data available on human body size, organized by age and percentile groupings. Also included is data relative to the range of joint motion and body sizes of children. The third part contains hundreds of dimensioned drawings, illustrating in plan and section the proper anthropometrically based relationship between user and space. The types of spaces range from residential and commercial to recreational and institutional, and all dimensions include metric conversions. In the Epilogue, the authors challenge the interior design profession, the building industry, and the furniture manufacturer to seriously explore the problem of adjustability in design. They expose the fallacy of designing to accommodate the so-called average man, who, in fact, does not exist. Using government data, including studies prepared by Dr. Howard Stoudt, Dr. Albert Damon, and Dr. Ross McFarland, formerly of the Harvard School of Public Health, and Jean Roberts of the U.S. Public Health Service, Panero and Zelnik have devised a system of interior design reference standards, easily understood through a series of charts and situation drawings. With *Human Dimension and Interior Space*, these standards are now accessible to all designers of interior environments.

architecting modern data platforms pdf: *Architecting Mobile Solutions for the Enterprise* Dino Esposito, 2012-05-15 Your guide to planning and executing a complete mobile web strategy Revisit your approach to the mobile web—and deliver effective solutions that reach customers and clients on a variety of mobile devices. In this practical guide, web development luminary Dino Esposito shows you how to develop a solid mobile strategy for the enterprise, starting with an effective mobile website. You'll receive essential architectural and implementation guidance, as well as mobile-specific design patterns for building cross-platform and native applications. Discover how to: Architect a website accessible from many different mobile devices Implement design patterns specific to mobile app development Examine tools that enable you to write one codebase for many platforms Use technologies for building Windows Phone, iPhone, and Android apps Develop cross-platform app features, such as localization and offline behavior

architecting modern data platforms pdf: *Microservices* Eberhard Wolff, 2016-10-03 The Most Complete, Practical, and Actionable Guide to Microservices Going beyond mere theory and marketing hype, Eberhard Wolff presents all the knowledge you need to capture the full benefits of this emerging paradigm. He illuminates microservice concepts, architectures, and scenarios from a technology-neutral standpoint, and demonstrates how to implement them with today's leading technologies such as Docker, Java, Spring Boot, the Netflix stack, and Spring Cloud. The author fully explains the benefits and tradeoffs associated with microservices, and guides you through the entire project lifecycle: development, testing, deployment, operations, and more. You'll find best practices for architecting microservice-based systems, individual microservices, and nanoservices, each illuminated with pragmatic examples. The author supplements opinions based on his experience with concise essays from other experts, enriching your understanding and illuminating areas where experts disagree. Readers are challenged to experiment on their own the concepts explained in the book to gain hands-on experience. Discover what microservices are, and how they differ from other forms of modularization Modernize legacy applications and efficiently build new systems Drive more value from continuous delivery with microservices Learn how microservices differ from SOA Optimize the microservices project lifecycle Plan, visualize, manage, and evolve architecture Integrate and communicate among microservices Apply advanced architectural techniques, including CQRS and Event Sourcing Maximize resilience and stability Operate and monitor microservices in production Build a full implementation with Docker, Java, Spring Boot, the Netflix stack, and Spring Cloud Explore nanoservices with Amazon Lambda, OSGi, Java EE, Vert.x, Erlang, and Seneca Understand microservices' impact on teams, technical leaders, product owners, and stakeholders Managers will discover better ways to support microservices, and learn how adopting the method affects the entire organization. Developers will master the technical skills and concepts they need to be effective. Architects will gain a deep understanding of key issues in creating or

migrating toward microservices, and exactly what it will take to transform their plans into reality.

architecting modern data platforms pdf: Documenting Software Architectures Paul Clements, Felix Bachmann, Len Bass, David Garlan, James Ivers, Reed Little, Paulo Merson, Robert Nord, Judith Stafford, 2010-10-05 Software architecture—the conceptual glue that holds every phase of a project together for its many stakeholders—is widely recognized as a critical element in modern software development. Practitioners have increasingly discovered that close attention to a software system’s architecture pays valuable dividends. Without an architecture that is appropriate for the problem being solved, a project will stumble along or, most likely, fail. Even with a superb architecture, if that architecture is not well understood or well communicated the project is unlikely to succeed. *Documenting Software Architectures*, Second Edition, provides the most complete and current guidance, independent of language or notation, on how to capture an architecture in a commonly understandable form. Drawing on their extensive experience, the authors first help you decide what information to document, and then, with guidelines and examples (in various notations, including UML), show you how to express an architecture so that others can successfully build, use, and maintain a system from it. The book features rules for sound documentation, the goals and strategies of documentation, architectural views and styles, documentation for software interfaces and software behavior, and templates for capturing and organizing information to generate a coherent package. New and improved in this second edition: Coverage of architectural styles such as service-oriented architectures, multi-tier architectures, and data models Guidance for documentation in an Agile development environment Deeper treatment of documentation of rationale, reflecting best industrial practices Improved templates, reflecting years of use and feedback, and more documentation layout options A new, comprehensive example (available online), featuring documentation of a Web-based service-oriented system Reference guides for three important architecture documentation languages: UML, AADL, and SysML

architecting modern data platforms pdf: Modern Architecture in Africa Antoni S. Folkers, Belinda A. C. van Buiten, 2019-07-22 This book offers unique insights into modern African architecture, influenced by modern European architecture, and at the same time a natural successor to existing site-specific and traditional architecture. It brings together the worlds of traditional site-specific architecture with the Modernist Project in Africa, which to date have only been considered in isolation. The book covers the four architectural disciplines: urban planning, building technology, building physics, and conservation. It includes an introduction with a historical outline and an analysis and comparison of a number of projects in various countries in Africa. On the basis of examples drawn from practice, the author documents and describes the hybrid architectural forms that have emerged from the confrontation and fusion with (pre)modern Western architecture and urban planning, and in so doing he also narrates the history of African architecture.

architecting modern data platforms pdf: Cloud Application Architectures George Reese, 2009-04-01 If you're involved in planning IT infrastructure as a network or system architect, system administrator, or developer, this book will help you adapt your skills to work with these highly scalable, highly redundant infrastructure services. While analysts hotly debate the advantages and risks of cloud computing, IT staff and programmers are left to determine whether and how to put their applications into these virtualized services. *Cloud Application Architectures* provides answers -- and critical guidance -- on issues of cost, availability, performance, scaling, privacy, and security. With *Cloud Application Architectures*, you will: Understand the differences between traditional deployment and cloud computing Determine whether moving existing applications to the cloud makes technical and business sense Analyze and compare the long-term costs of cloud services, traditional hosting, and owning dedicated servers Learn how to build a transactional web application for the cloud or migrate one to it Understand how the cloud helps you better prepare for disaster recovery Change your perspective on application scaling To provide realistic examples of the book's principles in action, the author delves into some of the choices and operations available on Amazon Web Services, and includes high-level summaries of several of the other services available on the market today. *Cloud Application Architectures* provides best practices that apply to every available

cloud service. Learn how to make the transition to the cloud and prepare your web applications to succeed.

architecting modern data platforms pdf: Computer Organization and Design RISC-V Edition David A. Patterson, John L. Hennessy, 2017-05-12 The new RISC-V Edition of Computer Organization and Design features the RISC-V open source instruction set architecture, the first open source architecture designed to be used in modern computing environments such as cloud computing, mobile devices, and other embedded systems. With the post-PC era now upon us, Computer Organization and Design moves forward to explore this generational change with examples, exercises, and material highlighting the emergence of mobile computing and the Cloud. Updated content featuring tablet computers, Cloud infrastructure, and the x86 (cloud computing) and ARM (mobile computing devices) architectures is included. An online companion Web site provides advanced content for further study, appendices, glossary, references, and recommended reading. - Features RISC-V, the first such architecture designed to be used in modern computing environments, such as cloud computing, mobile devices, and other embedded systems - Includes relevant examples, exercises, and material highlighting the emergence of mobile computing and the cloud

architecting modern data platforms pdf: Data Architecture Charles Tupper, 2011-05-09 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

Back to Home: <https://a.comtex-nj.com>