Designing Data Intensive Applications The Big Ideas Behind Reliable Scalable And Maintainable Systems

Every day, companies struggle to scale critical applications. As traffic volume and data demands increase, these applications become more complicated and brittle, exposing risks and compromising availability. This practical guide shows IT, devops, and system reliability managers how to prevent an application from becoming slow, inconsistent, or downright unavailable as it grows. Scaling isn't just about handling more users; it's also about managing risk and ensuring availability. Author Lee Atchison provides basic techniques for building applications that can handle huge quantities of traffic, data, and demand without affecting the quality your customers expect. In five parts, this book explores: Availability: learn techniques for building highly available applications, and for tracking and improving availability going forward Risk management: identify, mitigate, and manage risks in your application, test your recovery/disaster plans, and build out systems that contain fewer risks Services and microservices: understand the value of services for building complicated applications that need to operate at higher scale Scaling applications: assign services to specific teams, label the criticalness of each service, and devise failure scenarios and recovery plans Cloud services: understand the structure of cloud-based services, resource allocation, and service distribution

Summary Grokking Deep Learning teaches you to build deep learning neural networks from scratch! In his engaging style, seasoned deep learning expert Andrew Trask shows you the science under the hood, so you grok for yourself every detail of training neural networks. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Deep learning, a branch of artificial intelligence, teaches computers to learn by using neural networks, technology inspired by the human brain. Online text translation, self-driving cars, personalized product recommendations, and virtual voice assistants are just a few of the exciting modern advancements possible thanks to deep learning. About the Book Grokking Deep Learning teaches you to build deep learning neural networks from scratch! In his engaging style, seasoned deep learning expert Andrew Trask shows you the science under the hood, so you grok for yourself every detail of training neural networks. Using only Python and its math-supporting library, NumPy, you'll train your own neural networks to see and understand images, translate text into different languages, and even write like Shakespeare! When you're done, you'll be fully prepared to move on to mastering deep learning frameworks. What's inside The science behind deep learning Building and training your own neural networks Privacy concepts, including federated learning Tips for continuing your pursuit of deep learning About the Reader For readers with high school-level math and intermediate programming skills. About the Author Andrew Trask is a PhD student at Oxford University and a research scientist at DeepMind. Previously, Andrew was a researcher and analytics product manager at Digital Reasoning, where he trained the world's largest artificial neural network and helped guide the analytics roadmap for the Synthesys cognitive computing platform. Table of Contents Introducing deep learning: why you should learn it Fundamental concepts: how do machines learn? Introduction to neural prediction: forward propagation Introduction to neural learning: gradient descent Learning multiple weights at a time: generalizing gradient descent Building your first deep neural network: introduction to backpropagation How to picture neural networks: in your head and on paper Learning signal and ignoring noise:introduction to regularization and batching Modeling probabilities and nonlinearities: activation functions Neural learning about edges and corners: intro to convolutional neural networks Neural networks that understand language: king - man + woman == ? Neural networks that write like Shakespeare: recurrent layers for variable-length data Introducing automatic optimization: let's build a deep learning framework Learning to write like Shakespeare: long short-term memory Deep learning on unseen data: introducing federated learning Where to go from here: a brief guide

Designing Data-Intensive ApplicationsThe Big Ideas Behind Reliable, Scalable, and Maintainable Systems"O'Reilly Media, Inc."

One of the biggest challenges for organizations that have adopted microservice architecture is the lack of architectural, operational, and organizational standardization. After splitting a monolithic application or building a microservice ecosystem from scratch, many engineers are left wondering what's next. In this practical book, author Susan Fowler presents a set of microservice standards in depth, drawing from her experience standardizing over a thousand microservices at Uber. You'll learn how to design microservices that are stable, reliable, scalable, fault tolerant, performant, monitored, documented, and prepared for any catastrophe. Explore production-readiness standards, including: Stability and Reliability: develop, deploy, introduce, and deprecate microservices; protect against dependency failures Scalability and Performance: learn essential components for achieving greater microservice efficiency Fault Tolerance and Catastrophe Preparedness: ensure availability by actively pushing microservices to fail in real time Monitoring: learn how to monitor, log, and display key metrics; establish alerting and on-call procedures Documentation and Understanding: mitigate tradeoffs that come with microservice adoption, including organizational sprawl and technical debt

When it comes to choosing, using, and maintaining a database, understanding its internals is essential. But with so many distributed databases and tools available today, it's often difficult to understand what each one offers and how they differ. With this practical guide, Alex Petrov guides developers through the concepts behind modern database and storage engine internals. Throughout the book, you'll explore relevant material gleaned from numerous books, papers, blog posts, and the source code of several open source databases. These resources are listed at the end of parts one and two. You'll discover that the most significant distinctions among many modern databases reside in subsystems that determine how storage is organized and how data is distributed. This book examines: Storage engines: Explore storage classification and taxonomy, and dive into B-Tree-based and immutable Log Structured storage engines, with differences and use-cases for each Storage building blocks: Learn how database files are organized to build efficient storage, using auxiliary data structures such as Page Cache, Buffer Pool and Write-Ahead Log Distributed systems: Learn step-by-step how nodes and processes connect and build complex communication patterns Database clusters: Which consistency models are commonly used by modern databases and how distributed storage systems achieve consistency

Summary Kafka Streams in Action teaches you everything you need to know to implement stream processing on data flowing into your Kafka platform, allowing you to focus on getting more from your data without sacrificing time or effort. Foreword by Neha Narkhede, Cocreator of Apache Kafka Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Not all stream-based applications require a dedicated processing cluster. The lightweight Kafka Streams library provides exactly the power and simplicity you need for message handling in microservices and real-time event processing. With the Kafka Streams API, you filter and transform data streams with just Kafka and your application. About the Book Kafka Streams in Action teaches you to implement stream processing within the Kafka platform. In this easy-to-follow book, you'll explore real-world examples to collect, transform, and aggregate data, work with multiple processors, and handle real-time events. You'll even dive into streaming SQL with KSQL! Practical to the very end, it finishes with testing and operational aspects, such as monitoring and debugging. What's inside Using the KStreams API Filtering, transforming, and splitting data Working with the Processor API Integrating with external systems About the Reader Assumes some experience with distributed systems. No knowledge of Kafka or streaming applications required. About the Author Bill Bejeck is a Kafka Streams contributor and Confluent engineer with over 15 years of software development experience. Table of Contents PART 1 - GETTING STARTED WITH KAFKA STREAMS Welcome to Kafka Streams Kafka quicklyPART 2 - KAFKA STREAMS DEVELOPMENT Developing Kafka Streams Streams and state The KTable API The Processor APIPART 3 - ADMINISTERING KAFKA STREAMS Monitoring and performance Testing a Kafka Streams applicationPART 4 - ADVANCED CONCEPTS WITH KAFKA STREAMS Advanced applications with Kafka StreamsAPPENDIXES Appendix A - Additional configuration information Appendix B - Exactly once semantics

The ultimate guide for anyone wondering how President Joe Biden will respond to the COVID-19 pandemic—all his plans, goals, and executive orders in response to the coronavirus crisis. Shortly after being inaugurated as the 46th President of the United States, Joe Biden and his administration released this 200 page guide detailing his plans to respond to the coronavirus pandemic. The National Strategy for the COVID-19 Response and Pandemic Preparedness breaks down seven crucial goals of President Joe Biden's administration with regards to the coronavirus pandemic: 1. Restore trust with the American people. 2. Mount a safe, effective, and comprehensive vaccination campaign. 3. Mitigate spread through expanding masking, testing, data, treatments, health care workforce, and clear public health standards. 4. Immediately expand emergency relief and exercise the Defense Production Act. 5. Safely reopen schools, businesses, and travel while protecting workers. 6. Protect those most at risk and advance equity, including across racial, ethnic and rural/urban lines. 7. Restore U.S. leadership globally and build better preparedness for future threats. Each of these goals are explained and detailed in the book, with evidence about the current circumstances and how we got here, as well as plans and concrete steps to achieve each goal. Also included is the full text of the many Executive Orders that will be issued by President Biden to achieve each of these goals. The National Strategy for the COVID-19 Response and Pandemic Preparedness is required reading for anyone interested in or concerned about the COVID-19 pandemic and its effects on American society.

Architect and design data-intensive applications and, in the process, learn how to collect, process, store, govern, and expose data for a variety of use cases Key Features Integrate the data-intensive approach into your application architecture Create a robust application layout with effective messaging and data querying architecture Enable smooth data flow and make the data of your application intensive and fast Book Description Are you an architect or a developer who looks at your own applications gingerly while browsing through Facebook and applauding it silently for its data-intensive, yet ?uent and efficient, behaviour? This book is your gateway to build smart data-intensive systems by incorporating the core data-intensive architectural principles, patterns, and techniques directly into your application architecture. This book starts by taking you through the primary design challenges involved with architecting data-intensive applications. You will learn how to implement data curation and data dissemination, depending on the volume of your data. You will then implement your application architecture one step at a time. You will get to grips with implementing the correct message delivery protocols and creating a data layer that doesn't fail when running high traffic. This book will show you how you can divide your application into layers, each of which adheres to the single responsibility principle. By the end of this book, you will learn to streamline your thoughts and make the right choice in terms of technologies and architectural principles based on the problem at hand. What you will learn Understand how to envision a data-intensive system Identify and compare the non-functional requirements of a data collection component Understand patterns involving data processing, as well as technologies that help to speed up the development of data processing systems Understand how to implement Data Governance policies at design time using various Open Source Tools Recognize the anti-patterns to avoid while designing a data store for applications Understand the different data dissemination technologies available to query the data in an efficient manner Implement a simple data governance policy that can be extended using Apache Falcon Who this book is for This book is for developers and data architects who have to code, test, deploy, and/or maintain large-scale, high data volume applications. It is also useful for system architects who need to understand various non-functional aspects revolving around Data Intensive Systems.

Data is getting bigger and more complex by the day, and so are your choices in handling it. Explore some of the most cutting-edge databases available - from a traditional relational database to newer NoSQL approaches - and make informed decisions about challenging data storage problems. This is the only comprehensive guide to the world of NoSQL databases, with in-depth practical and conceptual introductions to seven different technologies: Redis, Neo4J, CouchDB, MongoDB, HBase, Postgres, and DynamoDB. This second edition includes a new chapter on DynamoDB and updated content for each chapter. While relational databases such as MySQL remain as relevant as ever, the alternative, NoSQL paradigm has opened up new horizons in performance and scalability and changed the way we approach data-centric problems. This book presents the essential concepts behind each database alongside hands-on examples that make each technology come alive. With each database, tackle a real-world problem that highlights the concepts and features that make it shine. Along the way, explore five database models - relational, key/value, columnar, document, and graph - from the perspective of challenges faced by real applications. Learn how MongoDB and CouchDB are strikingly different, make your applications faster with Redis and more connected with Neo4J, build a cluster of HBase servers using cloud services such as Amazon's Elastic MapReduce, and more. This new edition brings a brand new chapter on DynamoDB, updated code samples and exercises, and a more up-to-date account of each database's feature set. Whether you're a programmer building the next big thing, a data scientist seeking solutions to thorny problems, or a technology enthusiast venturing into new territory, you will find something to inspire you in this book. What You Need: You'll need a *nix shell (Mac OS or Linux preferred, Windows users will need Cygwin), Java 6 (or greater), and Ruby 1.8.7 (or greater). Each chapter will list the downloads required for that database.

You want increased customer satisfaction, faster development cycles, and less wasted work. Domain-driven design (DDD) combined with functional programming is the innovative combo that will get you there. In this pragmatic, down-to-earth guide, you'll see how applying the core principles of functional programming can result in software designs that model real-world requirements both elegantly and concisely - often more so than an object-oriented approach. Practical examples in the open-source F# functional language, and examples from familiar business domains, show you how to apply these techniques to build software that is business-focused, flexible, and high quality. Domain-driven design is a well-established approach to designing software that ensures that domain experts and developers work together effectively to create high-quality software. This book is the first to combine DDD with techniques from statically typed functional programming. This book is perfect for newcomers to DDD or functional programming - all the techniques you need will be introduced and explained. Model a complex domain accurately using the F# type system, creating compilable code that is also readable documentation---ensuring that the code and design never get out of sync. Encode business rules in the design so that you have "compile-time unit tests," and eliminate many potential bugs by making illegal states unrepresentable. Assemble a series of small, testable functions into a complete use case, and compose these individual scenarios into a large-scale design. Discover why the combination of functional programming and DDD leads naturally to service-oriented and hexagonal architectures. Finally, create a functional domain model that works with traditional databases, NoSQL, and event stores, and safely expose your domain via a website or API. Solve real problems by focusing on real-world requirements for your software. What You Need: The code in this book is designed to be run interactively on Windows, Mac and Linux.You will need a recent version of F# (4.0 or greater), and the appropriate .NET runtime for your platform.Full installation instructions for all platforms at fsharp.org.

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

Using research in neurobiology, cognitive science and learning theory, this text loads patterns into your brain in a way that lets you put them to work immediately, makes you better at solving software design problems, and improves your ability to speak the language of patterns with others on your team.

Data-intensive science has the potential to transform scientific research and quickly translate scientific progress into complete solutions, policies, and economic success. But this collaborative science is still lacking the effective access and exchange of knowledge among scientists, researchers, and policy makers across a range of disciplines. Bringing together leaders from multiple scientific disciplines, Data-Intensive Science shows how a comprehensive integration of various techniques and technological advances can effectively harness the vast amount of data being generated and significantly accelerate scientific progress to address some of the world's most challenging problems. In the book, a diverse cross-section of application, computer, and data scientists explores the impact of data-intensive science on current research and describes emerging technologies that will enable future scientific breakthroughs. The book identifies best practices used to tackle challenges facing data-intensive science as well as gaps in these approaches. It also focuses on the integration of data-intensive science into standard research practice, explaining how components in the data-intensive science environment need to work together to provide the necessary infrastructure for community-scale scientific collaborations. Organizing the material based on a high-level, data-intensive science workflow, this book provides an understanding of the scientific problems that would benefit from collaborative research, the current capabilities of data-intensive science, and the solutions to enable the next round of scientific advancements.

PEEK "UNDER THE HOOD" OF BIG DATA ANALYTICS The world of big data analytics grows ever more complex. And while many people can work superficially with specific frameworks, far fewer understand the fundamental principles of large-scale, distributed data processing systems and how they operate. In Foundations of Data Intensive Applications: Large Scale Data Analytics under the Hood, renowned big-data experts and computer scientists Drs. Supun Kamburugamuve and Saliya Ekanayake deliver a practical guide to applying the principles of big data to software development for optimal performance. The authors discuss foundational components of large-scale data systems and walk readers through the major software design decisions that define performance, application type, and usability. You???ll learn how to recognize problems in your applications resulting in performance and distributed operation issues, diagnose them, and effectively eliminate them by relying on the bedrock big data principles explained within. Moving beyond individual frameworks and APIs for data processing, this book unlocks the theoretical ideas that operate under the hood of every big data processing system. Ideal for data scientists, data architects, dev-ops engineers, and developers, Foundations of Data Intensive Applications: Large Scale Data Analytics under the Hood shows readers how to: Identify the foundations of large-scale, distributed data processing systems Make major software design decisions that optimize performance Diagnose performance problems and distributed operation issues Understand state-of-the-art research in big data Explain and use the major big data frameworks and understand what underpins them Use big data analytics in the real world to solve practical problems A surprisingly simple way for students to master any subject--based on one of the world's most popular online courses and the bestselling book A Mind for Numbers A Mind for Numbers and its wildly popular online companion course "Learning How to Learn" have empowered more than two million learners of all ages from around the world to master subjects that they once struggled with. Fans often wish they'd discovered these learning strategies earlier and ask how they can help their kids master these skills as well. Now in this new book for kids and teens, the authors reveal how to make the most of time spent studying. We all have the tools to learn what might not seem to come naturally to us at first--the secret is to understand how the brain works so we can unlock its power. This book explains: • Why sometimes letting your mind wander is an important part of the learning process • How to avoid "rut think" in order to think outside the box • Why having a poor memory can be a good thing • The value of metaphors in developing understanding • A simple, yet powerful, way to stop procrastinating Filled with illustrations, application questions, and exercises, this book makes learning easy and fun.

This invaluable roadmap for startup engineers reveals how to successfully handle web application scalability challenges to meet increasing product and traffic demands. Web Scalability for Startup Engineers shows engineers working at startups and small companies how to plan and implement a comprehensive scalability strategy. It presents broad and holistic view of infrastructure and architecture of a scalable web application. Successful startups often face the challenge of scalability, and the core concepts driving a scalable architecture are language and platform agnostic. The book covers scalability of HTTP-based systems (websites, REST APIs, SaaS, and mobile application backends), starting with a high-level perspective before taking a deep dive into common challenges and issues. This approach builds a holistic view of the problem, helping you see the big picture, and then introduces different technologies and best practices for solving the problem at hand. The book is enriched with the author's real-world experience and expert advice, saving you precious time and effort by learning from others' mistakes and successes. Language-agnostic approach addresses universally challenging concepts in Web development/scalability—does not require knowledge of a particular language Fills the gap for engineers in startups and smaller companies who have limited means for getting to the next level in terms of accomplishing scalability Strategies presented help to decrease time to market and increase the efficiency of web applications

Learn how easy it is to apply sophisticated statistical and machine learning methods to real-world problems when you build on top of the Google Cloud Platform (GCP). This hands-on guide shows developers entering the data science field how to implement an end-to-end data pipeline, using statistical and machine learning methods and tools on GCP. Through the course of the book, you'll work through a sample business decision by employing a variety of data science approaches. Follow along by implementing these statistical and machine learning solutions in your own project on GCP, and discover how this platform provides a transformative and more collaborative way of doing data science. You'll learn how to: Automate and schedule data ingest, using an App Engine application Create and populate a dashboard in Google Data Studio Build a real-time analysis pipeline to carry out streaming analytics Conduct interactive data exploration with Google BigQuery Create a Bayesian model on a Cloud Dataproc cluster Build a logistic regression machine-learning model with Spark Compute time-aggregate features with a Cloud Dataflow pipeline Create a high-performing prediction model with TensorFlow Use your deployed model as a microservice you can access from both batch and real-time pipelines

In the race to compete in today's fast-moving markets, large enterprises are busy adopting new technologies for creating new products, processes, and business models. But one obstacle on the road to digital transformation is placing too much emphasis on technology, and not enough on the types of processes technology enables. What if different lines of business could build their own services and applications—and decision-making was distributed rather than centralized? This report explores the concept of a digital business platform as a way of empowering individual business sectors to act on data in real time. Much innovation in a digital enterprise will increasingly happen at the edge, whether it involves business users (from marketers to data scientists) or IoT devices. To facilitate the process, your core IT team can provide these sectors with the digital tools they need to innovate quickly. This report explores: Key cultural and organizational changes for developing business capabilities through cross-functional product teams A platform for integrating applications, data sources, business partners, clients, mobile apps, social networks, and IoT devices Creating internal API programs for building innovative edge services in low-code or no-code environments Tools including Integration Platform as a Service, Application Platform as a Service, and Integration Software as a Service The challenge of

integrating microservices and serverless architectures Event-driven architectures for processing and reacting to events in real time You'll also learn about a complete pervasive integration solution as a core component of a digital business platform to serve every audience in your organization.

How do you detangle a monolithic system and migrate it to a microservice architecture? How do you do it while maintaining business-as-usual? As a companion to Sam Newman's extremely popular Building Microservices, this new book details a proven method for transitioning an existing monolithic system to a microservice architecture. With many illustrative examples, insightful migration patterns, and a bevy of practical advice to transition your monolith enterprise into a microservice operation, this practical guide covers multiple scenarios and strategies for a successful migration, from initial planning all the way through application and database decomposition. You'll learn several tried and tested patterns and techniques that you can use as you migrate your existing architecture. Ideal for organizations looking to transition to microservices, rather than rebuild Helps companies determine whether to migrate, when to migrate, and where to begin Addresses communication, integration, and the migration of legacy systems Discusses multiple migration patterns and where they apply Provides database migration examples, along with synchronization strategies Explores application decomposition, including several architectural refactoring patterns Delves into details of database decomposition, including the impact of breaking referential and transactional integrity, new failure modes, and more

This book presents a range of cloud computing platforms for data-intensive scientific applications. It covers systems that deliver infrastructure as a service, including: HPC as a service; virtual networks as a service; scalable and reliable storage; algorithms that manage vast cloud resources and applications runtime; and programming models that enable pragmatic programming and implementation toolkits for eScience applications. Many scientific applications in clouds are also introduced, such as bioinformatics, biology, weather forecasting and social networks. Most chapters include case studies. Cloud Computing for Data-Intensive Applications targets advanced-level students and researchers studying computer science and electrical engineering. Professionals working in cloud computing, networks, databases and more will also find this book useful as a reference.

Today, software engineers need to know not only how to program effectively but also how to develop proper engineering practices to make their codebase sustainable and healthy. This book emphasizes this difference between programming and software engineering. How can software engineers manage a living codebase that evolves and responds to changing requirements and demands over the length of its life? Based on their experience at Google, software engineers Titus Winters and Hyrum Wright, along with technical writer Tom Manshreck, present a candid and insightful look at how some of the world's leading practitioners construct and maintain software. This book covers Google's unique engineering culture, processes, and tools and how these aspects contribute to the effectiveness of an engineering organization. You'll explore three fundamental principles that software organizations should keep in mind when designing, architecting, writing, and maintaining code: How time affects the sustainability of software and how to make your code resilient over time How scale affects the viability of software practices within an engineering organization What trade-offs a typical engineer needs to make when evaluating design and development decisions

A single dramatic software failure can cost a company millions of dollars - but can be avoided with simple changes to design and architecture. This new edition of the best-selling industry standard shows you how to create systems that run longer, with fewer failures, and recover better when bad things happen. New coverage includes DevOps, microservices, and cloud-native architecture. Stability antipatterns have grown to include systemic problems in large-scale systems. This is a must-have pragmatic guide to engineering for production systems. If you're a software developer, and you don't want to get alerts every night for the rest of your life, help is here. With a combination of case studies about huge losses - lost revenue, lost reputation, lost time, lost opportunity - and practical, down-to-earth advice that was all gained through painful experience, this book helps you avoid the pitfalls that cost companies millions of dollars in downtime and reputation. Eighty percent of project life-cycle cost is in production, yet few books address this topic. This updated edition deals with the production of today's systems - larger, more complex, and heavily virtualized - and includes information on chaos engineering, the discipline of applying randomness and deliberate stress to reveal systematic problems. Build systems that survive the real world, avoid downtime, implement zero-downtime upgrades and continuous delivery, and make cloud-native applications resilient. Examine ways to architect, design, and build software - particularly distributed systems - that stands up to the typhoon winds of a flash mob, a Slashdotting, or a link on Reddit. Take a hard look at software that failed the test and find ways to make sure your software survives. To skip the pain and get the experience...get this book.

Our world is being revolutionized by data-driven methods: access to large amounts of data has generated new insights and opened exciting new opportunities in commerce, science, and computing applications. Processing the enormous quantities of data necessary for these advances requires large clusters, making distributed computing paradigms more crucial than ever. MapReduce is a programming model for expressing distributed computations on massive datasets and an execution framework for large-scale data processing on clusters of commodity servers. The programming model provides an easy-to-understand abstraction for designing scalable algorithms, while the execution framework transparently handles many system-level details, ranging from scheduling to synchronization to fault tolerance. This book focuses on MapReduce algorithm design, with an emphasis on text processing algorithms common in natural language processing, information retrieval, and machine learning. We introduce the notion of MapReduce design patterns, which represent general reusable solutions to commonly occurring problems across a variety of problem domains. This book not only intends to help the reader "think in MapReduce", but also discusses limitations of the programming model as well. This volume is a printed version of a work that appears in the Synthesis Digital Library of Engineering and Computer Science. Synthesis Lectures provide concise, original presentations of important research and development topics, published quickly, in digital and print formats. For more information visit www.morganclaypool.com

Data-intensive systems are software applications that process and generate Big Data. Data-intensive systems support the use of large amounts of data strategically and efficiently to provide intelligence. For example, examining industrial sensor data or business process data can enhance production, guide proactive improvements of development processes, or optimize supply chain systems. Designing data-intensive software systems is difficult because distribution of knowledge across stakeholders creates a symmetry of ignorance, because a shared vision of the future requires the development of new knowledge that extends and synthesizes existing knowledge. Knowledge Management in the Development of Data-Intensive Systems addresses new challenges arising from knowledge management in the development of data-intensive software systems. These challenges concern requirements, architectural design, detailed design, implementation and maintenance. The book covers the current state

and future directions of knowledge management in development of data-intensive software systems. The book features both academic and industrial contributions which discuss the role software engineering can play for addressing challenges that confront developing, maintaining and evolving systems;data-intensive software systems of cloud and mobile services; and the scalability requirements they imply. The book features software engineering approaches that can efficiently deal with data-intensive systems as well as applications and use cases benefiting from data-intensive systems. Providing a comprehensive reference on the notion of data-intensive systems from a technical and non-technical perspective, the book focuses uniquely on software engineering and knowledge management in the design and maintenance of data-intensive systems. The book covers constructing, deploying, and maintaining high quality software products and software engineering in and for dynamic and flexible environments. This book provides a holistic guide for those who need to understand the impact of variability on all aspects of the software life cycle. It leverages practical experience and evidence to look ahead at the challenges faced by organizations in a fast-moving world with increasingly fast-changing customer requirements and expectations.

Data-intensive systems are a technological building block supporting Big Data and Data Science applications.This book familiarizes readers with core concepts that they should be aware of before continuing with independent work and the more advanced technical reference literature that dominates the current landscape. The material in the book is structured following a problem-based approach. This means that the content in the chapters is focused on developing solutions to simplified, but still realistic problems using data-intensive technologies and approaches. The reader follows one reference scenario through the whole book, that uses an open Apache dataset. The origins of this volume are in lectures from a master's course in Data-intensive Systems, given at the University of Stavanger. Some chapters were also a base for guest lectures at Purdue University and Lodz University of Technology.

Summary Cloud Native Patternsis your guide to developing strong applications that thrive in the dynamic, distributed, virtual world of the cloud. This book presents a mental model for cloud-native applications, along with the patterns, practices, and tooling that set them apart. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Cloud platforms promise the holy grail: near-zero downtime, infinite scalability, short feedback cycles, fault-tolerance, and cost control. But how do you get there? By applying cloudnative designs, developers can build resilient, easily adaptable, web-scale distributed applications that handle massive user traffic and data loads. Learn these fundamental patterns and practices, and you'll be ready to thrive in the dynamic, distributed, virtual world of the cloud. About the Book With 25 years of experience under her belt, Cornelia Davis teaches you the practices and patterns that set cloud-native applications apart. With realistic examples and expert advice for working with apps, data, services, routing, and more, she shows you how to design and build software that functions beautifully on modern cloud platforms. As you read, you will start to appreciate that cloud-native computing is more about the how and why rather than the where. What's inside The lifecycle of cloud-native apps Cloud-scale configuration management Zero downtime upgrades, versioned services, and parallel deploys Service discovery and dynamic routing Managing interactions between services, including retries and circuit breakers About the Reader Requires basic software design skills and an ability to read Java or a similar language. About the Author Cornelia Davis is Vice President of Technology at Pivotal Software. A teacher at heart, she's spent the last 25 years making good software and great software developers. Table of Contents PART 1 - THE CLOUD-NATIVE CONTEXT You keep using that word: Defining "cloud-native" Running cloud-native applications in production The platform for cloud-native software PART 2 - CLOUD-NATIVE PATTERNS Event-driven microservices: It's not just request/response App redundancy: Scale-out and statelessness Application configuration: Not just environment variables The application lifecycle: Accounting for constant change Accessing apps: Services, routing, and service discovery Interaction redundancy: Retries and other control loops Fronting services: Circuit breakers and API gateways Troubleshooting: Finding the needle in the haystack Cloud-native data: Breaking the data monolith

Annotation Over the past 10 years, distributed systems have become more fine-grained. From the large multi-million line long monolithic applications, we are now seeing the benefits of smaller self-contained services. Rather than heavy-weight, hard to change Service Oriented Architectures, we are now seeing systems consisting of collaborating microservices. Easier to change, deploy, and if required retire, organizations which are in the right position to take advantage of them are yielding significant benefits. This book takes an holistic view of the things you need to be cognizant of in order to pull this off. It covers just enough understanding of technology, architecture, operations and organization to show you how to move towards finer-grained systems.

This text represents a breakthrough in the process underlying the design of the increasingly common and important data-driven Web applications.

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

This goal of this book is to provide a reliable and easy to understand strategy to approach system design questions. The process and justification of your ideas are the most important things in system design interviews. Thus the combination of right strategy and knowledge is vital to the success of your interview. Some candidates fail because lack of knowledge while some fail because they do not find the right way to approach the problem. This book provides valuable ways to fix both problems. By the time you finish the book, you are exceptionally well-equipped to tackle any system design questions.About the authorAlex is an experienced software engineer and entrepreneur. He enjoys

hand-on engineering and the thrill of working on a variety of software products including business applications, web apps and mobile apps. He has worked at Apple and Twitter among other internet companies. While not doing software development, Alex enjoys hiking and gaming. During the job interviews, he learned many things about system design interviews and achieved many successes. But, it is very time consuming to find the effective materials to prepare the interview, so Alex wrote this book offering the best knowledge to ace the design interviews. Alex hopes this book will save you a lot of time, energy to master the system design questions. TABLE OF CONTENTSCHAPTER ONE: SCALE FROM ZERO TO TEN MILLION USERSCHAPTER TWO: DESIGN CONSISTENT HASHINGCHAPTER THREE: DESIGN A KEY-VALUE STORECHAPTER FOUR: DESIGN A URL SHORTENER

Streaming data is a big deal in big data these days. As more and more businesses seek to tame the massive unbounded data sets that pervade our world, streaming systems have finally reached a level of maturity sufficient for mainstream adoption. With this practical guide, data engineers, data scientists, and developers will learn how to work with streaming data in a conceptual and platform-agnostic way. Expanded from Tyler Akidau's popular blog posts "Streaming 101" and "Streaming 102", this book takes you from an introductory level to a nuanced understanding of the what, where, when, and how of processing real-time data streams. You'll also dive deep into watermarks and exactly-once processing with co-authors Slava Chernyak and Reuven Lax. You'll explore: How streaming and batch data processing patterns compare The core principles and concepts behind robust out-of-order data processing How watermarks track progress and completeness in infinite datasets How exactly-once data processing techniques ensure correctness How the concepts of streams and tables form the foundations of both batch and streaming data processing The practical motivations behind a powerful persistent state mechanism, driven by a real-world example How time-varying relations provide a link between stream processing and the world of SQL and relational algebra

"This book focuses on the challenges of distributed systems imposed by the data intensive applications, and on the different state-of-the-art solutions proposed to overcome these challenges"--Provided by publisher.

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

Architect scalable, reliable, and maintainable applications for enterprises with Python Key Features Explore various Python design patterns used for enterprise software development Apply best practices for testing and performance optimization to build stable applications Learn about different attacking strategies used on enterprise applications and how to avoid them Book Description Dynamically typed languages like Python are continuously improving. With the addition of exciting new features and a wide selection of modern libraries and frameworks, Python has emerged as an ideal language for developing enterprise applications. Hands-On Enterprise Application Development with Python will show you how to build effective applications that are stable, secure, and easily scalable. The book is a detailed guide to building an end-to-end enterprise-grade application in Python. You will learn how to effectively implement Python features and design patterns that will positively impact your application lifecycle. The book also covers advanced concurrency techniques that will help you build a RESTful application with an optimized frontend. Given that security and stability are the foundation for an enterprise application, you'll be trained on effective testing, performance analysis, and security practices, and understand how to embed them in your codebase during the initial phase. You'll also be guided in how to move on from a monolithic architecture to one that is service oriented, leveraging microservices and serverless deployment techniques. By the end of the book, you will have become proficient at building efficient enterprise applications in Python. What you will learn Understand the purpose of design patterns and their impact on application lifecycle Build applications that can handle large amounts of data-intensive operations Uncover advanced concurrency techniques and discover how to handle a large number of requests in production Optimize frontends to improve the client-side experience of your application Effective testing and performance profiling techniques to detect issues in applications early in the development cycle Build applications with a focus on security Implement large applications as microservices to improve scalability Who this book is for If you're a developer who wants to build enterprise-grade applications, this book is for you. Basic to intermediate-level of programming experience with Python and database systems is required to understand the concepts covered in this book.

If you need to build a scalable, fault tolerant system with requirements for high availability, discover why the Erlang/OTP platform stands out for the breadth, depth, and consistency of its features. This hands-on guide demonstrates how to use the Erlang programming language and its OTP framework of reusable libraries, tools, and design principles to develop complex commercial-grade systems that simply cannot fail. In the first part of the book, you'll learn how to design and implement process behaviors and supervision trees with Erlang/OTP, and bundle them into standalone nodes. The second part addresses reliability, scalability, and high availability in your overall system design. If you're familiar with Erlang, this book will help you understand the design choices and trade-offs necessary to keep your system running. Explore OTP's building blocks: the Erlang language, tools and libraries collection, and its abstract principles and design rules Dive into the fundamentals of OTP reusable frameworks: the Erlang process structures OTP uses for behaviors Understand how OTP behaviors support client-server structures, finite state machine patterns, event handling, and runtime/code integration Write your own behaviors and special processes Use OTP's tools, techniques, and architectures to handle deployment, monitoring, and operations

Describes principles of the emerging field of data-intensive computing, along with methods for designing, managing and analyzing the big data sets of today.

A guide to developing Web sites using scalable applications.

In this practical book, four Cloudera data scientists present a set of self-contained patterns for performing large-scale data analysis with Spark. The authors bring Spark, statistical methods, and real-world data sets together to teach you how to approach analytics problems by example. You'll start with an introduction to Spark and its ecosystem, and then dive into patterns that apply common techniques—classification, collaborative filtering, and anomaly detection among others—to fields such as genomics, security, and finance. If you have an entry-level understanding of machine learning and statistics, and you program in Java, Python, or Scala, you'll find these patterns useful for working on your own data applications. Patterns include: Recommending music and the Audioscrobbler data set Predicting forest cover with decision trees Anomaly detection in network traffic with K-means clustering Understanding Wikipedia with Latent Semantic Analysis Analyzing co-occurrence networks with GraphX Geospatial and temporal data analysis on the New York City Taxi Trips data Estimating financial risk through Monte Carlo simulation Analyzing genomics data and the BDG project Analyzing neuroimaging data with PySpark and Thunder