

O'Reilly Data Engineering

Excel Scientific and Engineering Cookbook
 Data Teams
 Database Internals
 The Pragmatic Programmer
 Data Science on AWS
 Pragmatic AI
 Official Google Cloud Certified Professional Data Engineer Study Guide
 Azure Data Engineering Cookbook
 The Evolving Role of the Data Engineer
 Data Science with Java
 Database Reliability Engineering
 Data Engineering with Python
 Engineering Dynamics
 97 Things Every Cloud Engineer Should Know
 Python Data Science Handbook
 97 Things Every Data Engineer Should Know
 Programming C# 8.0
 Introducing MLOps
 Building Machine Learning Powered Applications
 Scala for Data Science
 The Self-Service Data Roadmap
 Architectural Patterns
 Data Science on the Google Cloud Platform
 Software Architecture: The Hard Parts
 Trino: The Definitive Guide
 Fundamentals of Data Engineering
 Site Reliability Engineering
 Data Management at Scale
 Data Mesh
 Data Engineering with Apache Spark, Delta Lake, and Lakehouse
 Feature Engineering for Machine Learning
 Jumpstart Snowflake
 The Rails Way
 Software Engineering at Google
 Doing Data Science
 Engineering MLOps
 Designing Data-Intensive Applications
 Data Pipelines Pocket Reference
 Python for Excel

O'Reilly Data Engineering

Downloaded from dev.mabts.edu by guest

EVIE JIMENA

Excel Scientific and Engineering Cookbook Addison-Wesley Professional

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

[Data Teams](#) Packt Publishing Ltd

Master Powerful Off-the-Shelf Business Solutions for AI and Machine Learning Pragmatic AI will help you solve real-world problems with contemporary machine learning, artificial intelligence, and cloud computing tools. Noah Gift demystifies all the concepts and tools you need to get results—even if you don't have a strong background in math or data science. Gift illuminates powerful off-the-shelf cloud offerings from Amazon, Google, and Microsoft, and demonstrates proven techniques using the Python data science ecosystem. His workflows and examples help you streamline and simplify every step, from deployment to production, and build exceptionally scalable solutions. As you learn how machine language (ML) solutions work, you'll gain a more intuitive understanding of what you can achieve with them and how to maximize their value. Building on these fundamentals, you'll walk step-by-step through building cloud-based AI/ML applications to address realistic issues in sports marketing, project management, product pricing, real estate, and beyond. Whether you're a business professional, decision-maker, student, or programmer, Gift's expert guidance and wide-ranging case studies will prepare you to solve data science problems in virtually any environment. Get and configure all the tools you'll

need Quickly review all the Python you need to start building machine learning applications Master the AI and ML toolchain and project lifecycle Work with Python data science tools such as IPython, Pandas, Numpy, Jupyter Notebook, and Sklearn Incorporate a pragmatic feedback loop that continually improves the efficiency of your workflows and systems Develop cloud AI solutions with Google Cloud Platform, including TPU, Colaboratory, and Datalab services Define Amazon Web Services cloud AI workflows, including spot instances, code pipelines, boto, and more Work with Microsoft Azure AI APIs Walk through building six real-world AI applications, from start to finish Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

[Database Internals](#) Addison-Wesley Professional Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by Facebook, open source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn about tools that will help you connect to Trino and query data Go deeper: Learn Trino's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more Put Trino in production: Secure Trino, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino

[The Pragmatic Programmer](#) "O'Reilly Media, Inc."

The infrastructure-as-code revolution in IT is also affecting database administration. With this practical book, developers, system administrators, and junior to mid-level DBAs will learn how the modern practice of site reliability engineering applies to the craft of database architecture and operations. Authors Laine Campbell and Charity Majors provide a framework for professionals looking to join the ranks of today's database reliability engineers (DBRE). You'll begin by exploring core operational concepts that DBREs need to master. Then you'll examine a wide range of database persistence options, including how to implement key technologies to provide resilient, scalable, and performant data storage and retrieval. With a firm foundation in database reliability engineering, you'll be ready to dive into the architecture and operations of any modern database. This book covers: Service-level requirements and risk management Building

and evolving an architecture for operational visibility Infrastructure engineering and infrastructure management How to facilitate the release management process Data storage, indexing, and replication Identifying datastore characteristics and best use cases Datastore architectural components and data-driven architectures

[Data Science on AWS](#) O'Reilly Media

Companies working to become data driven often view data scientists as heroes, but that overlooks the vital role that data engineers play in the process. While data scientists focus on finding new insights from datasets, data engineers deal with preparation-obtaining, cleaning, and creating enhanced versions of the data an organization needs. In this report, Andy Oram examines how the role of data engineer has quickly evolved. DBAs, software engineers, developers, and students will explore the responsibilities of modern data engineers and the skills and tools necessary to do the job. You'll learn how to deal with software engineering concepts such as rapid and continuous development, automation and orchestration, modularity, and traceability. Decision makers considering a move to the cloud will also benefit from the in-depth discussion this report provides. This report covers: Major tasks of data engineers today The different levels of structure in data and ways to maximize its value Capabilities of third-party cloud options Tools for ingestion, transfer, and enrichment Using containers and VMs to run the tools Software engineering development Automation and orchestration of data engineering.

Pragmatic AI "O'Reilly Media, Inc."

Understand the complexities of modern-day data engineering platforms and explore strategies to deal with them with the help of use case scenarios led by an industry expert in big data Key Features Become well-versed with the core concepts of Apache Spark and Delta Lake for building data platforms Learn how to ingest, process, and analyze data that can be later used for training machine learning models Understand how to operationalize data models in production using curated data Book Description In the world of ever-changing data and schemas, it is important to build data pipelines that can auto-adjust to changes. This book will help you build scalable data platforms that managers, data scientists, and data analysts can rely on. Starting with an introduction to data engineering, along with its key concepts and architectures, this book will show you how to use Microsoft Azure Cloud services effectively for data engineering. You'll cover data lake design patterns and the different stages through which the data needs to flow in a typical data lake. Once you've explored the main features of Delta Lake to build data lakes with fast performance and governance in mind, you'll advance to implementing the lambda architecture using Delta Lake. Packed with practical examples and code snippets, this

book takes you through real-world examples based on production scenarios faced by the author in his 10 years of experience working with big data. Finally, you'll cover data lake deployment strategies that play an important role in provisioning the cloud resources and deploying the data pipelines in a repeatable and continuous way. By the end of this data engineering book, you'll know how to effectively deal with ever-changing data and create scalable data pipelines to streamline data science, ML, and artificial intelligence (AI) tasks. What you will learn Discover the challenges you may face in the data engineering world Add ACID transactions to Apache Spark using Delta Lake Understand effective design strategies to build enterprise-grade data lakes Explore architectural and design patterns for building efficient data ingestion pipelines Orchestrate a data pipeline for preprocessing data using Apache Spark and Delta Lake APIs Automate deployment and monitoring of data pipelines in production Get to grips with securing, monitoring, and managing data pipelines models efficiently Who this book is for This book is for aspiring data engineers and data analysts who are new to the world of data engineering and are looking for a practical guide to building scalable data platforms. If you already work with PySpark and want to use Delta Lake for data engineering, you'll find this book useful. Basic knowledge of Python, Spark, and SQL is expected.

Official Google Cloud Certified Professional Data Engineer Study Guide "O'Reilly Media, Inc."

The proven Study Guide that prepares you for this new Google Cloud exam The Google Cloud Certified Professional Data Engineer Study Guide, provides everything you need to prepare for this important exam and master the skills necessary to land that coveted Google Cloud Professional Data Engineer certification. Beginning with a pre-book assessment quiz to evaluate what you know before you begin, each chapter features exam objectives and review questions, plus the online learning environment includes additional complete practice tests. Written by Dan Sullivan, a popular and experienced online course author for machine learning, big data, and Cloud topics, Google Cloud Certified Professional Data Engineer Study Guide is your ace in the hole for deploying and managing analytics and machine learning applications.

- Build and operationalize storage systems, pipelines, and compute infrastructure
- Understand machine learning models and learn how to select pre-built models
- Monitor and troubleshoot machine learning models
- Design analytics and machine learning applications that are secure, scalable, and highly available.

This exam guide is designed to help you develop an in depth understanding of data engineering and machine learning on Google Cloud Platform.

Azure Data Engineering Cookbook Pearson Education
If you create, manage, operate, or configure systems running in the cloud, you're a cloud engineer—even if you work as a system administrator, software developer, data scientist, or site reliability engineer. With this book, professionals from around the world provide valuable insight into today's cloud engineering role. These concise articles explore the entire cloud computing experience, including fundamentals, architecture, and migration. You'll delve into security and compliance, operations and reliability, and software development. And examine networking, organizational culture, and more. You're sure to find 1, 2, or 97 things that inspire you to dig deeper and expand your own career. "Three Keys to Making the Right Multicloud Decisions," Brendan O'Leary "Serverless Bad Practices," Manases Jesus Galindo Bello "Failing a Cloud Migration," Lee Atchison "Treat Your Cloud Environment as If It Were On Premises," Iyana Garry "What Is Toil, and Why Are SREs Obsessed with It?," Zachary Nickens "Lean QA: The QA Evolving in the DevOps World," Theresa Neate "How Economies of Scale Work in the Cloud," Jon Moore "The Cloud Is Not About the Cloud," Ken Corless "Data Gravity: The Importance of Data Management in the Cloud," Geoff Hughes "Even in the Cloud, the Network Is the Foundation," David Murray "Cloud Engineering Is About Culture, Not Containers," Holly Cummins **The Evolving Role of the Data Engineer** "O'Reilly Media, Inc."

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

Data Science with Java "O'Reilly Media, Inc."

Take advantage of today's sky-high demand for data engineers. With this in-depth book, current and aspiring engineers will learn powerful real-world best practices for managing data big and small. Contributors from notable companies including Twitter, Google, Stitch Fix, Microsoft, Capital One, and LinkedIn share their experiences and lessons learned for overcoming a variety of specific and often nagging challenges. Edited by Tobias Macey, host of the popular Data Engineering Podcast, this book presents 97 concise and useful tips for cleaning, prepping, wrangling, storing, processing, and ingesting data. Data engineers, data architects, data team managers, data scientists, machine learning engineers, and software engineers will greatly benefit from the wisdom and experience of their peers. Topics include: The Importance of Data Lineage - Julien Le Dem Data Security for Data Engineers - Katharine Jarmul The Two Types of Data Engineering and Data Engineers - Jesse Anderson Six Dimensions for Picking an Analytical Data Warehouse - Gleb Mezhanskiy The End of ETL as We Know It - Paul Singman Building a Career as a Data Engineer - Vijay Kiran Modern Metadata for the Modern Data Stack - Prukalpa Sankar Your Data Tests Failed! Now What? - Sam Bail

Database Reliability Engineering Packt Publishing Ltd

With this practical book, AI and machine learning practitioners will learn how to successfully build and deploy data science projects on Amazon Web Services. The Amazon AI and machine learning stack unifies data science, data engineering, and application development to help level up your skills. This guide shows you how to build and run pipelines in the cloud, then integrate the results into applications in minutes instead of days. Throughout the book, authors Chris Fregly and Antje Barth demonstrate how to reduce cost and improve performance. Apply the Amazon AI and ML stack to real-world use cases for natural language processing, computer vision, fraud detection, conversational devices, and more Use automated machine learning to implement a specific subset of use cases with SageMaker Autopilot Dive deep into the complete model development lifecycle for a BERT-based NLP use case including data ingestion, analysis, model training, and deployment Tie everything together into a repeatable machine learning operations pipeline Explore real-time ML, anomaly detection, and streaming analytics on data streams with Amazon Kinesis and Managed Streaming for Apache Kafka Learn security best practices for data science projects and workflows including identity and access management, authentication, authorization, and more

Data Engineering with Python "O'Reilly Media, Inc."

Feature engineering is a crucial step in the machine-learning pipeline, yet this topic is rarely examined on its own. With this practical book, you'll learn techniques for extracting and transforming features—the numeric representations of raw data—into formats for machine-learning models. Each chapter guides you through a single data problem, such as how to represent text or image data. Together, these examples illustrate the main principles of feature engineering. Rather than simply teach these principles, authors Alice Zheng and Amanda Casari focus on practical application with exercises throughout the book. The closing chapter brings everything together by tackling a real-world, structured dataset with several feature-engineering techniques. Python packages including numpy, Pandas, Scikit-learn, and Matplotlib are used in code examples. You'll examine: Feature engineering for numeric data: filtering, binning, scaling, log transforms, and power transforms Natural text techniques: bag-of-words, n-grams, and phrase detection Frequency-based filtering and feature scaling for eliminating uninformative features Encoding techniques of categorical variables, including feature hashing and bin-counting Model-based feature engineering with principal component analysis The concept of model stacking, using k-means as a featurization technique Image feature extraction with manual and deep-learning techniques **Engineering Dynamics** John Wiley & Sons

This Primer is intended to provide the theoretical background for the standard undergraduate, mechanical engineering course in dynamics. The book contains several worked examples and summaries and exercises at the end of each chapter to aid readers in their understanding of the material. Teachers who wish to have a source of more detailed theory for the course, as well as graduate students who need a refresher course on undergraduate dynamics when preparing for certain first year graduate school examinations, and students taking the course will find the work very helpful.

97 Things Every Cloud Engineer Should Know Springer

Now that people are aware that data can make the difference in an election or a business model, data science as an occupation is gaining ground. But how can you get started working in a wide-ranging, interdisciplinary field that's so clouded in hype? This insightful book, based on Columbia University's Introduction to Data Science class, tells you what you need to know. In many of these chapter-long lectures, data scientists from companies such as Google, Microsoft, and eBay share new algorithms, methods, and models by presenting case studies and the code they use. If

you're familiar with linear algebra, probability, and statistics, and have programming experience, this book is an ideal introduction to data science. Topics include: Statistical inference, exploratory data analysis, and the data science process Algorithms Spam filters, Naive Bayes, and data wrangling Logistic regression Financial modeling Recommendation engines and causality Data visualization Social networks and data journalism Data engineering, MapReduce, Pregel, and Hadoop Doing Data Science is collaboration between course instructor Rachel Schutt, Senior VP of Data Science at News Corp, and data science consultant Cathy O'Neil, a senior data scientist at Johnson Research Labs, who attended and blogged about the course.

Python Data Science Handbook "O'Reilly Media, Inc."

Given the improved analytical capabilities of Excel, scientists and engineers everywhere are using it—instead of FORTRAN—to solve problems. And why not? Excel is installed on millions of computers, features a rich set of built-in analyses tools, and includes an integrated Visual Basic for Applications (VBA) programming language. No wonder it's today's computing tool of choice. Chances are you already use Excel to perform some fairly routine calculations. Now the Excel Scientific and Engineering Cookbook shows you how to leverage Excel to perform more complex calculations, too, calculations that once fell in the domain of specialized tools. It does so by putting a smorgasbord of data analysis techniques right at your fingertips. The book shows how to perform these useful tasks and others: Use Excel and VBA in general Import data from a variety of sources Analyze data Perform calculations Visualize the results for interpretation and presentation Use Excel to solve specific science and engineering problems Wherever possible, the Excel Scientific and Engineering Cookbook draws on real-world examples from a range of scientific disciplines such as biology, chemistry, and physics. This way, you'll be better prepared to solve the problems you face in your everyday scientific or engineering tasks. High on practicality and low on theory, this quick, look-up reference provides instant solutions, or "recipes," to problems both basic and advanced. And like other books in O'Reilly's popular Cookbook format, each recipe also includes a discussion on how and why it works. As a result, you can take comfort in knowing that complete, practical answers are a mere page-flip away.

97 Things Every Data Engineer Should Know "O'Reilly Media, Inc."

We're at an inflection point in data, where our data management solutions no longer match the complexity of organizations, the proliferation of data sources, and the scope of our aspirations to get value from data with AI and analytics. In this practical book, author Zhamak Dehghani introduces data mesh, a decentralized sociotechnical paradigm drawn from modern distributed architecture that provides a new approach to sourcing, sharing, accessing, and managing analytical data at scale. Dehghani guides practitioners, architects, technical leaders, and decision makers on their journey from traditional big data architecture to a distributed and multidimensional approach to analytical data management. Data mesh treats data as a product, considers domains as a primary concern, applies platform thinking to create self-serve data infrastructure, and introduces a federated computational model of data governance. Get a complete introduction to data mesh principles and its constituents Design a data mesh architecture Guide a data mesh strategy and execution Navigate organizational design to a decentralized data ownership model Move beyond traditional data warehouses and lakes to a distributed data mesh

Programming C# 8.0 Apress

For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use: IPython and Jupyter: provide computational environments for data scientists using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib: includes capabilities for a flexible range of data visualizations in Python Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms **Introducing MLOps** Packt Publishing Ltd

Data Science is booming thanks to R and Python, but Java brings the robustness, convenience, and ability to scale critical to today's data science applications. With this practical book, Java software engineers looking to add data science skills will take a logical journey through the data science pipeline. Author Michael Brzustowicz explains the basic math theory behind each step of the data science process, as well as how to apply these concepts with Java. You'll learn the critical roles that data IO, linear algebra,

statistics, data operations, learning and prediction, and Hadoop MapReduce play in the process. Throughout this book, you'll find code examples you can use in your applications. Examine methods for obtaining, cleaning, and arranging data into its purest form Understand the matrix structure that your data should take Learn basic concepts for testing the origin and validity of data Transform your data into stable and usable numerical values Understand supervised and unsupervised learning algorithms, and methods for evaluating their success Get up and running with MapReduce, using customized components suitable for data science algorithms

Building Machine Learning Powered Applications O'Reilly Media Data pipelines are the foundation for success in data analytics. Moving data from numerous diverse sources and transforming it to provide context is the difference between having data and actually gaining value from it. This pocket reference defines data pipelines and explains how they work in today's modern data

stack. You'll learn common considerations and key decision points when implementing pipelines, such as batch versus streaming data ingestion and build versus buy. This book addresses the most common decisions made by data professionals and discusses foundational concepts that apply to open source frameworks, commercial products, and homegrown solutions.

You'll learn: What a data pipeline is and how it works How data is moved and processed on modern data infrastructure, including cloud platforms Common tools and products used by data engineers to build pipelines How pipelines support analytics and reporting needs Considerations for pipeline maintenance, testing, and alerting

"O'Reilly Media, Inc."

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

Related with Oreilly Data Engineering:

© [Oreilly Data Engineering Materials Today Chemistry Impact Factor](#)

© [Oreilly Data Engineering Math 55 Harvard Syllabus](#)

© [Oreilly Data Engineering Math 110 Penn State](#)