

1. Record Nr.	UNINA9911046616603321
Autore	Kumar Manoj
Titolo	Mastering Data Engineering and Analytics with Databricks : A Hands-On Guide to Build Scalable Pipelines Using Databricks, Delta Lake, and MLflow (English Edition)
Pubbl/distr/stampa	Delhi : , : Orange Education PVT Ltd, , 2024 ©2024
ISBN	9788196862046 8196862040
Edizione	[1st ed.]
Descrizione fisica	1 online resource (331 pages)
Soggetti	Big data Data mining
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Cover Page -- Title Page -- Copyright Page -- Dedication Page -- About the Author -- About the Technical Reviewers -- Acknowledgements -- Preface -- Errata -- Table of Contents -- SECTION 1 Getting Started with Data Engineering and Databricks -- 1. Introducing Data Engineering with Databricks -- Introduction -- Structure -- The Basics of Data Engineering -- Data -- Data Layers -- Raw Data -- Enriched Data -- Curated Data -- Big Data -- Data Quality -- Master Data/Dimensions -- Transactions/Facts -- Times Series Data -- Data Serialization -- Parquet -- JavaScript Object Notation (JSON) -- Comma Separated Values (CSV) -- Schema
Sommario/riassunto	In today's data-driven world, mastering data engineering is crucial for driving innovation and delivering real business impact. Databricks is one of the most powerful platforms which unifies data, analytics and AI requirements of numerous organizations worldwide. Mastering Data Engineering and Analytics with Databricks goes beyond the basics, offering a hands-on, practical approach tailored for professionals eager to excel in the evolving landscape of data engineering and analytics. This book uniquely blends foundational knowledge with advanced

applications, equipping readers with the expertise to build, optimize, and scale data pipelines that meet real-world business needs. With a focus on actionable learning, it delves into complex workflows, including real-time data processing, advanced optimization with Delta Lake, and seamless ML integration with MLflow--skills critical for today's data professionals. Drawing from real-world case studies in FMCG and CPG industries, this book not only teaches you how to implement Databricks solutions but also provides strategic insights into tackling industry-specific challenges. From setting up your environment to deploying CI/CD pipelines, you'll gain a competitive edge by mastering techniques that are directly applicable to your organization's data strategy. By the end, you'll not just understand Databricks--you'll command it, positioning yourself as a leader in the data engineering space.
