1.	Record Nr.	UNINA9910584482303321
	Autore	L'Esteve Ron
	Titolo	The Azure Data Lakehouse Toolkit : Building and Scaling Data Lakehouses on Azure with Delta Lake, Apache Spark, Databricks, Synapse Analytics, and Snowflake / / by Ron L'Esteve
	Pubbl/distr/stampa	Berkeley, CA : , : Apress : , : Imprint : Apress, , 2022
	ISBN	9781484282335 1484282337
	Edizione	[1st ed. 2022.]
	Descrizione fisica	1 online resource (467 pages)
	Disciplina	004.6782
	Soggetti	Microsoft Azure (Computing platform) Cloud computing Electronic data processing Databases
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Note generali	Includes index.
	Nota di bibliografia	Includes index.
	Nota di contenuto	Part I: Getting Started -- Chapter 1: The Data Lakehouse Paradigm -- Part II: Data Platforms -- Chapter 2: Snowflake -- Chapter 3: Databricks -- Chapter 4: Synapse Analytics -- Part III: Apache Spark ELT -- Chapter 5: Pipelines and Jobs -- Chapter 6: Notebook Code -- Part IV: Delta Lake.-Chapter 7: Schema Evolution -- Chapter 8: Change Feed -- Chapter 9: Clones -- Chapter 10: Live Tables -- Chapter 11: Sharing -- Part V: Optimizing Performance -- Chapter 12: Dynamic Partition Pruning for Querying Star Schemas -- Chapter 13: Z-Ordering & Data Skipping -- Chapter 14: Adaptive Query Execution -- Chapter 15: Bloom Filter Index -- Chapter 16: Hyperspace -- Part VI: Advanced Capabilities -- Chapter 17: Auto Loader -- Chapter 18: Python Wheels -- Chapter 19: Security & Controls.
	Sommario/riassunto	Design and implement a modern data lakehouse on the Azure Data Platform using Delta Lake, Apache Spark, Azure Databricks, Azure Synapse Analytics, and Snowflake. This book teaches you the intricate details of the Data Lakehouse Paradigm and how to efficiently design a cloud-based data lakehouse using highly performant and cutting-edge Apache Spark capabilities using Azure Databricks, Azure Synapse

Analytics, and Snowflake. You will learn to write efficient PySpark code for batch and streaming ELT jobs on Azure. And you will follow along with practical, scenario-based examples showing how to apply the capabilities of Delta Lake and Apache Spark to optimize performance, and secure, share, and manage a high volume, high velocity, and high variety of data in your lakehouse with ease. The patterns of success that you acquire from reading this book will help you hone your skills to build high-performing and scalable ACID-compliant lakehouses using flexible and cost-efficient decoupled storage and compute capabilities. Extensive coverage of Delta Lake ensures that you are aware of and can benefit from all that this new, open source storage layer can offer. In addition to the deep examples on Databricks in the book, there is coverage of alternative platforms such as Synapse Analytics and Snowflake so that you can make the right platform choice for your needs. After reading this book, you will be able to implement Delta Lake capabilities, including Schema Evolution, Change Feed, Live Tables, Sharing, and Clones to enable better business intelligence and advanced analytics on your data within the Azure Data Platform. What You Will Learn Implement the Data Lakehouse Paradigm on Microsoft's Azure cloud platform Benefit from the new Delta Lake open-source storage layer for data lakehouses Take advantage of schema evolution, change feeds, live tables, and more Write functional PySpark code for data lakehouse ELT jobs Optimize Apache Spark performance through partitioning, indexing, and other tuning options Choose between alternatives such as Databricks, Synapse Analytics, and Snowflake.