1. Record Nr. UNINA9910899895203321

Autore	Demchenko Yuri
Titolo	Big Data Infrastructure Technologies for Data Analytics : Scaling Data Science Applications for Continuous Growth / / by Yuri Demchenko, Juan J. Cuadrado-Gallego, Oleg Chertov, Marharyta Aleksandrova
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2024
ISBN	3-031-69366-3
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (553 pages)
Altri autori (Persone)	Cuadrado-GallegoJuan J ChertovOleg AleksandrovaMarharyta
Disciplina	005.7
Soggetti	Artificial intelligence - Data processing Quantitative research Software engineering Artificial intelligence Application software Data Science Data Analysis and Big Data Software Engineering Artificial Intelligence Computer and Information Systems Applications
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Chapter 1 Introduction. - Chapter 2 Big Data Technologies Foundation: Definition, Reference Architecture, use cases. - Chapter 3 Cloud Computing Foundation: Definition, Reference Architecture, Foundational Technologies, Use cases. - Chapter 4 Cloud and Big Data Service Providers and Platforms. - Chapter 5 Big Data Algorithms, MapReduce and Hadoop ecosystem -- Chapter 6 Streaming Analytics and Spark -- Chapter 7 Data Structures for Big Data, Modern Big Data SQL and NoSQL Databases.-Chapter 8 Enterprise Data Governance and Management -- Chapter 9 Research Data Management -- Chapter 10 Big Data Security and Compliance, Data Privacy Protection -- Chapter

11 Finding Data on the Web, Data sets, Web Scraping, Web API -- Chapter 12 Data Science Projects Management,DataOps, MLOPs -- Chapter13 Data Science Projects Development with Amazon SageMaker -- Chapter 14 Data Validation for Data Science Projects.

Sommario/riassunto

This book provides a comprehensive overview and introduction to Big Data Infrastructure technologies, existing cloud-based platforms, and tools for Big Data processing and data analytics, combining both a conceptual approach in architecture design and a practical approach in technology selection and project implementation. Readers will learn the core functionality of major Big Data Infrastructure components and how they integrate to form a coherent solution with business benefits. Specific attention will be given to understanding and using the major Big Data platform Apache Hadoop ecosystem, its main functional components MapReduce, HBase, Hive, Pig, Spark and streaming analytics. The book includes topics related to enterprise and research data management and governance and explains modern approaches to cloud and Big Data security and compliance. The book covers two knowledge areas defined in the EDISON Data Science Framework (EDSF): Data Science Engineering and Data Management and Governance and can be used as a textbook for university courses or provide a basis for practitioners for further self-study and practical use of Big Data technologies and competent evaluation and implementation of practical projects in their organizations.