

1. Record Nr.	UNINA9910830367503321
Autore	Huang Thomas (Technologist)
Titolo	Big data analytics in earth, atmospheric, and ocean sciences // Thomas Huang, Tiffany C. Vance, and Christopher Lynnes
Pubbl/distr/stampa	Hoboken, NJ : , : American Geophysical Union, , [2022] ©2022
ISBN	1-119-46755-1 1-119-46756-X
Descrizione fisica	1 online resource (355 pages) : illustrations (some color), color maps
Collana	Special Publications
Disciplina	550.0285/57
Soggetti	Earth sciences - Data processing Atmospheric science - Data processing
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Cover -- Title Page -- Copyright -- Contents -- List of Contributors -- Preface -- Chapter 1 An Introduction to Big Data Analytics -- 1.1 Overview -- 1.1.1 What Differentiates Spatial Big Data -- 1.2 Definitions -- 1.3 Example Problems -- 1.3.1 Agriculture -- 1.3.2 Commerce -- 1.3.3 Connected Cars -- 1.3.4 Environment -- 1.3.5 Financial Services -- 1.3.6 Government Agencies -- 1.3.7 Health Care -- 1.3.8 Marketing -- 1.3.9 Mining -- 1.3.10 Petroleum -- 1.3.11 Retail -- 1.3.12 Telecommunications -- 1.3.13 Transportation -- 1.3.14 Utilities -- 1.4 Big Data Analysis Concepts -- 1.4.1 Summarizing Data -- 1.4.2 Identify Locations -- 1.4.3 Pattern Analysis -- 1.4.4 Cluster Analysis -- 1.4.5 Proximity Analysis -- 1.4.6 Predictive Modeling -- 1.5 Technology and Tools -- 1.5.1 Available Tools -- 1.6 Challenges -- 1.7 Summary -- References -- Part I Big Data Analytics Architecture -- Chapter 2 Introduction to Big Data Analytics Architecture -- References -- Chapter 3 Scaling Big Earth Science Data Systems Via Cloud Computing -- 3.1 Introduction -- 3.2 Key Concepts of Science Data Systems (SDSes) -- 3.3 Increasing Data Processing, Volumes, and Rates -- 3.3.1 Historical Example -- 3.3.2 Example of OnPremises SDS -- 3.3.3 Exceeding OnPremise Capacities -- 3.4 Cloud Concepts for SDSes -- 3.4.1 IaaS (Infrastructure as a Service) -- 3.4.2 PaaS (Platform as a Service) -- 3.4.3 SaaS (Software as a Service)

-- 3.4.4 XaaS (Everything as a Service) -- 3.5 Architecture Components of CloudBased SDS -- 3.5.1 Algorithm Development Environment -- 3.5.2 Processing Algorithm Catalog -- 3.5.3 Resource Management -- 3.5.4 Processing Orchestration/Workflow Management -- 3.5.5 Compute Services -- 3.5.6 Data Catalog -- 3.5.7 Data Storage Services -- 3.5.8 Common Services for Logging, Metrics, Events, and Analytics. 3.5.9 Integrating Science Data Processing and Algorithm Development, Software Catalog, and Analysis -- 3.6 Considerations for Multicloud and Hybrid SDS -- 3.6.1 Collocation of GDS, SDS, and DAAC -- 3.6.2 AllIn and LockIn -- 3.7 Cloud Economics -- 3.8 LargeScaling Considerations -- 3.8.1 Metrics for Anomalies -- 3.8.2 Thundering Herd -- 3.8.3 Higher SLA -- 3.8.4 Watchdogs -- 3.9 Example of Cloud SDSes -- 3.9.1 SMAP in the Cloud -- 3.9.2 NISAR SDS -- 3.10 Conclusion -- 3.10 References -- Chapter 4 NOAA Open Data Dissemination (Formerly NOAA Big Data Project/Program) -- 4.1 Obstacles to the Public's Use of NOAA Environmental Data -- 4.2 Public Access of NOAA Data Creates Challenges for the Agency -- 4.3 The Vision for NOAA's "Oddball" Approach to Big Data -- 4.4 A NOAA Cooperative Institute Data Broker provides Research and Operational Agility -- 4.5 PublicPrivate Partnerships Provide the Pipeline -- 4.6 BDP Exceeds Expectations and Evolves into Enterprise Operations -- 4.7 Engaging Users in the Cloud -- 4.7.1 Early Insights From User Engagement -- 4.7.2 Data Analytics and Metrics Informing User Engagement -- 4.7.3 NODD Supports Industry Challenges in Sustainability -- 4.7.4 Advancing User Engagement for NODD -- 4.8 Challenges and Opportunities -- 4.8.1 Format Conversions and Cloud Based Tools -- 4.8.2 Attention to Data Quality and Provenance -- 4.9 Vision for the Future -- 4.9 Acknowledgments -- 4.9 References -- Chapter 5 A Data Cube Architecture for CloudBased Earth Observation Analytics -- 5.1 Introduction -- 5.1.1 The Open Data Cube (ODC) Architecture -- 5.2 Open Data Cube for the Cloud Design -- 5.2.1 Storage Model -- 5.2.2 The ODC S3 Native Storage Driver -- 5.2.3 S3 Array Structure -- 5.2.4 Implementation of the S3 Array I/O Module -- 5.2.5 Execution Model -- 5.2.6 Execution Engine -- 5.3 S3 Array I/O Performance -- 5.3.1 Experiment Setup. 5.3.2 Raw S3 Read/Write Performance -- 5.3.3 Data Cube Ingest Scaling (Write) -- 5.3.4 Data Cube Load Scaling (Read) -- 5.4 Discussion and Conclusion -- 5.4.1 S3AIO Advantages -- 5.4.2 S3AIO Limitations -- 5.4.3 Next Steps for This Research -- 5.4 References -- Chapter 6 Open Source Exploratory Analysis of Big Earth Data With NEXUS -- 6.1 Introduction -- 6.1.1 Cloud Computing -- 6.1.2 MapReduce Programming Model -- 6.2 Architecture -- 6.3 Deployment Architecture -- 6.4 Benchmarking and Studies -- 6.4.1 Hurricane Katrina Case Study -- 6.5 Analytics Collaborative Framework -- 6.6 Federated Analytics Collaborative Systems -- 6.7 Conclusion -- 6.7 References -- Chapter 7 Benchmark Comparison of Cloud Analytics Methods Applied to Earth Observations -- 7.1 Introduction -- 7.2 Experimental Setup -- 7.3 AODS Candidates -- 7.4 Experimental Results -- 7.5 Conclusions -- 7.5 References -- Part II Analysis Methods for Big Earth Data -- Chapter 8 Introduction to Analysis Methods for Big Earth Data -- References -- Chapter 9 Spatial Statistics for Big Data Analytics in the Ocean and Atmosphere: Perspectives, Challenges, and Opportunities -- 9.1 Spatial Data and Spatial Statistics -- 9.2 What Constitutes Big Spatial Data? -- 9.3 Statistical Implications of the Four Vs of Big Spatial Data -- 9.3.1 Volume -- 9.3.2 Variety -- 9.3.3 Velocity -- 9.3.4 Veracity -- 9.4 Challenges to the Statistical Analysis of Big Spatial Data -- 9.4.1 Randomness and Sampling -- 9.4.2 High Dimensionality -- 9.4.3 Independence of Samples and

Spatial Autocorrelation -- 9.4.4 Effect Size -- 9.5 Opportunities in Spatial Analysis of Big Data -- 9.5.1 Ecological Marine Units -- 9.5.2 Spatiotemporal Analysis of California Maximum Temperature -- 9.6 Conclusion -- 9.6 References -- Chapter 10 Giving Scientists Back Their Flow: Analyzing Big Geoscience Data Sets in the Cloud -- 10.1 Introduction. 10.2 Where's the Opportunity? -- 10.2.1 Progression of Typical Approaches Over the Past 5 Years -- 10.2.2 On Premises Approaches -- 10.2.3 OffPremises Solutions: Cloud Computing -- 10.3 The Future -- 10.3 Reference -- Chapter 11 The Distributed Oceanographic MatchUp Service -- 11.1 Introduction -- 11.2 DOMS Capabilities -- 11.3 System Architecture -- 11.3.1 Data Sets -- 11.3.2 In Situ Data -- 11.3.3 Satellite Data -- 11.3.4 User Interface -- 11.4 Workflow -- 11.4.1 MatchUp Algorithms -- 11.4.2 DOMS API -- 11.4.3 DOMS File Output -- 11.4.4 User Interface -- 11.5 Future Development -- 11.5 Acknowledgments -- 11.5 Availability Statement -- 11.5 References -- Part III Big Earth Data Applications -- Chapter 12 Introduction to Big Earth Data Applications -- References -- Chapter 13 Topological Methods for Pattern Detection in Climate Data -- 13.1 Introduction -- 13.2 Topological Methods for Pattern Detection -- 13.2.1 Step 1: Topological Feature Descriptors of Weather Patterns -- 13.2.2 Step 2: Machine Learning for Classifying Weather Patterns -- 13.3 Case Study: Atmospheric Rivers Detection -- 13.3.1 Atmospheric Rivers -- 13.3.2 Data -- 13.3.3 Results -- 13.4 Conclusions and Recommendations -- 13.4 Acknowledgments -- 13.4 References -- Chapter 14 Exploring Large Scale Data Analysis and Visualization for Atmospheric Radiation Measurement Data Discovery Using NoSQL Technologies -- 14.1 Introduction -- 14.2 Software and Workflow -- 14.2.1 Software -- 14.2.2 Workflow -- 14.3 Hardware Architecture -- 14.4 Applications -- 14.4.1 LASSO Bundle Browser -- 14.4.2 ARMBE Visualizations -- 14.4.3 Data Analytics -- 14.5 Conclusions -- 14.5 Acknowledgments -- 14.5 References -- Chapter 15 Demonstrating Condensed Massive Satellite Data Sets for Rapid Data Exploration: The MODIS Land Surface Temperatures of Antarctica -- 15.1 Introduction -- 15.2 Data -- 15.3 Methods. 15.3.1 Data Set Cleaning -- 15.3.2 Baseline Statistics Generation -- 15.3.3 Anomaly Determination -- 15.3.4 Database Storage -- 15.4 Results -- 15.4.1 Efficiency and Performance -- 15.4.2 Cloud Masking Quality -- 15.4.3 Coldest Temperature -- 15.4.4 Spurious Warm Observations -- 15.5 Conclusions -- 15.5 Acknowledgments -- 15.5 Availability Statement -- 15.5 References -- Chapter 16 Developing Big Data Infrastructure for Analyzing AIS Vessel Tracking Data on a Global Scale -- 16.1 Introduction -- 16.2 Background -- 16.3 Use Case: Producing Heat Maps of Vessel Traffic using AIS Data -- 16.3.1 Overview -- 16.3.2 Data Preparation -- 16.4 Data Processing Overview -- 16.4.1 Data Processing Steps -- 16.4.2 Data Processing: Results -- 16.4.3 Data Curation and Open Access -- 16.5 Future Work -- 16.6 Conclusions -- 16.6 References -- Chapter 17 Future of Big Earth Data Analytics -- 17.1 Introduction -- 17.2 How Data Get Bigger -- 17.3 The Evolution of Analytics Algorithms -- 17.4 Analytics Architectures -- 17.5 Conclusions -- 17.5 References -- Index -- EULA.

Sommario/riassunto

"Big Data Analytics in Earth, Atmospheric and Ocean Sciences introduces the different aspects of big earth data analytics which include a brief introduction to big data about the Earth, an introduction to big earth data analytics, methodologies for big earth data analytics, and tools for analysis and applications in different earth science domains. - Contributions from leading researchers in the field compiled into an advanced volume available to a new audience of

students, researchers in allied disciplines in earth and environmental sciences, geography and geoinformatics - Significant results reported through funded projects from NASA, NOAA, NSF, and other resources both nationally and internationally, for sharing with the community about the source, management, analytics, visualization, and utilization of big earth data in a variety of applications and decision supporting contexts - Identification and discussion of the critical topics in the subject, and leading experts in the field, including non-scientists and business analysts seeking an introduction to the topic, e.g. companies moving into big earth data analysis Big Data Analytics in Earth, Atmospheric and Ocean Sciences is a valuable resource for geoinformatics and geoscience professionals who are working on providing big earth data processing, and scientists or engineers who need big earth data processing to support their research and development."
