



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
HAROKOPIO UNIVERSITY

School of Digital Technology Department of Informatics
and Telematics

Postgraduate Program “Informatics and Telematics”

End-to-End Computational Infrastructure for Geospatial Analysis

Master Thesis

Angelos Naoum



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
HAROKOPIO UNIVERSITY

Σχολή Ψηφιακής Τεχνολογίας Τμήμα Πληροφορικής και
Τηλεματικής

Πρόγραμμα Μεταπτυχιακών Σπουδών “Πληροφορική και
Τηλεματική”

Υπολογιστική υποδομή από άκρο σε άκρο για γεωχωρική ανάλυση

Διπλωματική Εργασία

Άγγελος Ναούμ



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
HAROKOPIO UNIVERSITY

School of Digital Technology Department of Informatics
and Telematics

Postgraduate Program “Informatics and Telematics”

Examining Committee Tsadimas Anargyros Laboratory
Teaching Staff, Department of Informatics and Telematics,
Harokopio University of Athens

Kousiouris Georgios Assistant Professor, Department of
Informatics and Telematics, Harokopio University of
Athens

Dalakas Vasileios Laboratory Teaching Staff, Department
of Informatics and Telematics, Harokopio University of
Athens

Ethics and Copyright Statement

I, Angelos Naoum, hereby declare that:

1) I am the owner of the intellectual rights of this original work and to the best of my knowledge, my work does not insult persons, nor does it offend the intellectual rights of third parties.

2) I accept that the Library and Information Centre of Harokopio University may, without changing the content of my work, make it available in electronic form through its Digital Library, copy it in any medium and / or any format and hold more than one copy for maintenance and safety purposes.

3) I have obtained, where necessary, permission from the copyright owners to use any third-party copyright material reproduced in the master thesis while the corresponding material is visible in the submitted work.

Acknowledgements

I would like to thank all the people who contributed through their own way in the work presented in this Master Thesis. Firstly, I would like to thank my academic supervisor, Dr. Anargyros Tsadimas for his guidance and assistance. Additionally, I am deeply grateful to Lydia for her continuous support, patience and understanding. I would like to thank my friends and family too, for their encouragement. Last but not least, I would like to thank my colleagues for the beautiful moments we shared both on a social and academic level.

Table of Contents

| | |
|--|-----------|
| Abstract..... | 6 |
| Introduction..... | 7 |
| Geospatial Analysis core concepts..... | 8 |
| GIS software tools..... | 9 |
| QGIS..... | 9 |
| ArcGIS..... | 10 |
| Jupyter Ecosystem..... | 11 |
| Problems on geospatial analysis..... | 15 |
| Thesis objectives and purpose..... | 18 |
| Technologies and Tools Used..... | 18 |
| Apache Sedona: Capabilities and usage for geospatial analysis..... | 18 |
| Apache Spark: Distributed processing of large-scale data..... | 18 |
| JupyterHub: Provision of a collaborative and interactive environment.. | 18 |
| Kubernetes: Management and organization of infrastructure..... | 18 |
| Architecture of the Proposed Infrastructure..... | 19 |
| End-to-end design of the infrastructure..... | 19 |
| Integration of Apache Sedona, Spark, and JupyterHub into | |
| Kubernetes..... | 19 |
| Implementation of the Infrastructure..... | 19 |
| Experimental Results and Evaluation..... | 19 |
| Testing and evaluation of the infrastructure's performance..... | 19 |
| Comparison with existing solutions..... | 19 |
| Performance analysis in large-scale geospatial data analysis..... | 19 |
| Conclusions and Future Work..... | 19 |
| Recap of the main findings and contributions..... | 19 |
| Suggestions for future improvements and extensions..... | 19 |
| Bibliography..... | 20 |
| Appendices..... | 23 |
| Technical details, code, and additional results..... | 23 |

Abstract

Many organizations, such as urban planning agencies, environmental research institutes, and transportation networks, require efficient solutions for managing and analyzing vast amounts of geospatial data to support decision-making and resource optimization.

This thesis introduces a robust computational infrastructure for geospatial analysis, combining Apache Sedona, Apache Spark, and JupyterHub, all deployed on Kubernetes.

Apache Sedona's specialized capabilities for large-scale geospatial data processing are integrated with the distributed computing framework of Apache Spark to enable efficient data handling, analysis, and visualization.

The deployment of JupyterHub on Kubernetes fosters a collaborative, interactive environment, allowing users to seamlessly work with geospatial data in a secure manner by integrating centralized authentication mechanisms.

This solution offers a scalable and flexible platform, addressing the growing demands of geospatial analysis while enhancing both performance and usability within the Jupyter ecosystem.

Introduction

In the modern world, analysis of geospatial data in an efficient and a robust manner is very significant for several applications areas such as:

- **Urban planning and development** where it plays a critical role in land use planning, smart city initiatives, and environmental impact assessments. [1]
- **Environmental monitoring and conservation** efforts leverage geospatial tools to track climate change, manage natural resources, and preserve biodiversity. [2]
- **Disaster management** uses geospatial data for risk assessment, emergency response, and recovery planning, while transportation and logistics benefit from route optimization and traffic analysis. [3]
- **Public health and epidemiology** employ geospatial analysis to map disease spread and assess healthcare access. [4]
- **Agriculture** utilizes it for precision farming and crop monitoring. [5]
- In **geospatial intelligence and defense**, it supports military operations, border security, and humanitarian aid. [6]
- In the **energy industry** spatial analysis use it for infrastructure monitoring and renewable energy site selection. [7]
- Geospatial tools are essential in **crime mapping and predictive policing** for understanding crime patterns and improving law enforcement strategies. [8]
- **Hydrology and water resource management** apply it for flood modeling and watershed management. [9]

- Finally, in **meteorology** for climate modeling and weather forecasting. [10]

Geospatial Analysis core concepts

All the concepts about geography and information technology come into **Geographic Information System (GIS)**. It allows users to visualize, interpret, and understand patterns and relationships in data through maps, graphs, and other analytical tools. GIS integrates various types of data (e.g., satellite imagery, maps, environmental data) and ties them to specific locations on Earth,

Geospatial analysis is the process of gathering, displaying, and manipulating geographic data to understand spatial relationships, patterns, and trends. This type of analysis combines geographic information systems (GIS) technology with statistical methods to analyze spatial data effectively. [11]

Geospatial Data (Geodata) are data formats where location information is stored, often represented through coordinates such as latitude and longitude [12]. There are several types of such data:

1. **Vector data** (points, lines, polygons) for discrete features like buildings, roads, or boundaries.
2. **Raster data** (grids or pixels) for continuous data like satellite imagery or elevation.
3. **Geographic databases** are able to house vectors and rasters. Databases store geographic data as a structured set of data/information.
4. **Web Files** are data formats specialized for web applications. As the internet becomes the largest library in the world, geodata has adapted with its own types of storage and access. For example, GeoJSON, GeoRSS, and web mapping services (WMS) were built

specifically to serve and display geographic features over the internet.

GIS software tools

There are several GIS software tools specialized for small and large organizations to perform analysis, create detailed maps, analyze spatial relationships, and manage geographic information. Two of the most popular GIS software tools are **QGIS** and **ArcGIS**, each offering unique features and capabilities for various use cases.

QGIS

QGIS (Quantum GIS) is a free and open-source GIS platform. It is widely used by individuals and organizations that prefer or require open-source software for budgetary reasons or to benefit from community-driven development [13].

QGIS offers extensive support for vector and raster data formats, including common geospatial data formats such as shapefiles, GeoTIFFs, and KML. The platform is highly customizable and can be extended through plugins developed by the global QGIS community, allowing users to tailor the software to their specific needs. Some of the key features of QGIS include:

- **Cost and Accessibility:** QGIS is entirely free, which significantly reduces the financial barriers to using powerful GIS software.
- **Cross-Platform Compatibility:** The software is available on multiple operating systems, including Windows, macOS, and Linux, ensuring that it is accessible to a wide range of users.
- **Extensive Plugin Library:** QGIS has an extensive library of plugins that enable users to perform specialized tasks such as remote sensing analysis, hydrological modeling, and georeferencing.

- **Cartographic Capabilities:** QGIS provides a range of tools for high-quality map creation, including symbolization, labeling, and map layout design, making it suitable for professional cartography.
- **Open-Source Community Support:** QGIS is maintained by an active community of developers and users who continuously contribute to its development and provide support to other users.

ArcGIS

ArcGIS, developed by **Esri**, is one of the most widely used commercial GIS platforms in the world. It offers a comprehensive suite of tools for spatial data management, analysis, and visualization, making it the preferred choice for professional GIS users in industries such as government, utilities, transportation, and environmental consulting. As a proprietary software, ArcGIS comes with a cost, but its wide array of features and robust support services make it a leading choice for organizations that require high-end geospatial analysis capabilities [14].

ArcGIS provides an extensive range of functionalities that support advanced geospatial analysis and decision-making processes. Some of the notable features of ArcGIS include:

- **Comprehensive Geospatial Tools:** ArcGIS includes powerful tools for spatial analysis, such as geostatistics, network analysis, and 3D modeling. These tools enable users to conduct in-depth analysis of geographic data and derive insights that support decision-making processes.
- **Integration with Esri Ecosystem:** ArcGIS seamlessly integrates with other Esri products, such as ArcGIS Online and ArcGIS Pro, which allows users to work across desktop, web, and cloud environments. This integration enhances collaboration and provides access to real-time data and services.
- **Advanced Cartography:** ArcGIS is known for its professional-grade cartographic capabilities, enabling users to produce highly detailed and aesthetically pleasing maps for

presentations and reports.

- **Enterprise Solutions:** ArcGIS supports enterprise-level geospatial data management through server-based applications, which can handle vast datasets and multiple users simultaneously.
- **Support and Training Resources:** As a commercial product, ArcGIS comes with extensive documentation, technical support, and training resources, which helps users maximize the software's capabilities.

While both QGIS and ArcGIS provide powerful tools for spatial data analysis, they cater to different types of users and organizations. **QGIS** is a more accessible option due to its open-source nature, making it ideal for users with limited budgets, such as academic researchers or small non-profit organizations. In contrast, **ArcGIS** offers a more feature-rich and enterprise-ready solution, but at a higher cost. ArcGIS is better suited for large organizations and industries that require advanced functionalities, professional support, and scalability [15].

Also both are primarily GUI-driven where operation involves menu-driven workflows with the users being able to interact with spatial data without writing code. While they support scripting their purpose is not made for such scenarios where users want to interact directly with the geospatial data in a high flexibility and customization manner, especially when dealing with different types of datasets (numeric, textual, spatial, etc.).

Jupyter Ecosystem

The **Jupyter ecosystem** refers to a collection of open-source tools and technologies centered around **Jupyter notebooks**, which provide an interactive computing environment for data science, analysis, education, and research.

The ecosystem has grown beyond the original Jupyter notebook to include a wide variety of interfaces, extensions, and libraries that support interactive workflows in multiple programming languages. It enables users to write code, visualize data, and document their thought process in an integrated, reproducible environment. Among others, some of its core components are the below [16]:

Jupyter Notebooks:

- **What it is:** A web-based interactive environment where you can write and execute code in a cell-based format. It allows combining code, text, equations, and visualizations in a single document.
- **Features:**
 - Supports multiple programming languages (e.g., Python, R, Julia, Scala).
 - Code execution and visualization happen in real-time.
 - Markdown support for writing rich text, including LaTeX for equations.
 - Integrated with interactive widgets and libraries for visualizations.
- **Use Cases:** Data exploration, machine learning experiments, educational material, presentations, and scientific research.

JupyterLab:

- **What it is:** The next-generation interface for Jupyter notebooks. It's a more flexible, modular, and extensible environment that supports working with notebooks, text editors, terminals, and data file viewers all in one interface [17].
- **Features:**
 - Drag-and-drop functionality for rearranging panes.
 - Multiple document types (notebooks, code scripts, markdown files) in one workspace.
 - Real-time collaboration and extension support.

- Provides a rich, IDE-like experience.
- **Use cases:** Exploratory Data Analysis (EDA), Real-time Collaboration

JupyterHub:

- **What it is:** A multi-user version of Jupyter, which allows multiple users to run Jupyter notebooks on shared infrastructure, such as in schools, research labs, or businesses.
- **Features:**
 - Provides access control and user management.
 - Scalable, with integrations for cloud services (AWS, Google Cloud, etc.) and Kubernetes for larger deployments.
 - Ideal for educational settings, where students can work on notebooks without needing to install software locally.
- **Use cases:** Centralized Notebook Hosting, Multi-user Access to Notebooks, Resources management

Binder:

- **What it is:** An online platform that allows users to share Jupyter notebooks that run directly in the browser without the need to install anything locally.
- **Features:**
 - Converts a GitHub repository containing Jupyter notebooks into an interactive, shareable environment.
 - Enables others to run code in a temporary cloud environment with no setup required.
- **Use Cases:** Sharing reproducible research, educational tutorials, demo notebooks.

nbconvert:

- **What it is:** A tool that allows you to convert Jupyter notebooks into different formats such as HTML, PDF, LaTeX, and slideshows.
- **Features:**
 - Converts notebooks into static files for sharing or publication.
 - Supports exporting notebooks for presentations or reports.
- **Use Cases:** Publishing research results, creating static reports, or generating presentations from notebooks.

Regarding geospatial analysis, It has become a powerful platform for this type of data analysis by combining interactive computing with the flexibility of Python and other languages. Several libraries and tools in this ecosystem are specifically designed for handling geospatial data, mapping, visualization, and advanced spatial analytics.

For instance, ISPRS (International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences), which is a key platform for disseminating scientific and technical information related to photogrammetry, remote sensing, and spatial information systems, published a paper where case studies of Jupyter notebooks usage for mapping projects were presented [18].

The study focused on using geospatial analysis tools to map educational and demographic data, highlighting applications in creating thematic maps for policy development and decision-making. Python libraries like *Folium*, *Geopandas*, *Matplotlib*, and *Ipywidgets* were employed for data visualization and analysis and the data that were gathered were from public repositories, including national statistics agencies and open-access platforms like OpenStreetMap with the goal of mapping the distribution of schools and the spatial relationships of educational variables (e.g., teacher qualifications and school accessibility) to demographic factors.

The study emphasizes Jupyter notebooks as a collaborative and visual tool that simplifies communication among multidisciplinary teams, including developers, educators, and policymakers.

Problems on geospatial analysis

As already stated earlier, geospatial analysis involves dealing with geographic data to address questions about spatial relationships, patterns, and trends leading to understanding complex environmental, social, and economic phenomena. Despite its growing importance, researchers in this domain face numerous challenges, many of which stem from technical, computational, and theoretical issues tied to data collection, integration, and management. These challenges highlight the necessity of innovative and interdisciplinary approaches to overcome barriers and harness the full potential of geospatial analysis.

One of the primary difficulties in geospatial analysis arises from the diversity and heterogeneity of data sources. Geospatial data can originate from satellite imagery, GPS devices, sensors, surveys, and historical maps, among others. This variety introduces inconsistencies in data resolution, accuracy, and format. For instance, satellite imagery may offer high-resolution spatial details, but its temporal resolution might not align with data collected from ground sensors. Such disparities necessitate preprocessing techniques, such as data resampling or transformation, to align data sets, which can significantly increase computational demands. Moreover, preprocessing steps, such as enhancing spatial resolution or correcting errors in sensor readings, often lead to larger data sizes, further complicating storage and processing requirements [19][20].

The integration of geospatial data with non-spatial data—such as demographic, economic, or health-related statistics—introduces additional complexities. Differences in granularity, measurement units, and data formats make it challenging to establish meaningful connections between spatial and non-spatial datasets. For example, administrative boundaries used in demographic data might not align perfectly with ecological zones defined in geospatial datasets, leading to

mismatched scales or overlapping features. Addressing these issues often requires advanced data harmonization techniques, which are computationally intensive and may require domain-specific knowledge [21][22].

Furthermore, the sheer volume of geospatial data poses significant computational challenges. Modern data sources, such as high-frequency satellite imagery or continuous sensor networks, produce vast amounts of information at an unprecedented scale. Managing, storing, and processing this data demands robust computational infrastructure, including high-performance computing (HPC) systems, cloud storage, and specialized software for spatial analysis. Additionally, advanced machine learning and artificial intelligence (AI) algorithms are increasingly employed to extract patterns and insights from geospatial data, but these methods often require extensive training and fine-tuning to address spatial heterogeneity and noise in the data [23][24].

Another critical challenge in geospatial analysis lies in the theoretical understanding and interpretation of spatial relationships. Unlike traditional statistical analysis, where independence between observations is often assumed, spatial data inherently exhibit spatial autocorrelation—the principle that nearby locations are more likely to exhibit similar characteristics than distant ones. Ignoring this spatial dependency can lead to incorrect inferences and model biases. Consequently, researchers must employ spatial statistical methods, such as geographically weighted regression (GWR) or spatial econometrics, which require specialized knowledge and tools [25][26].

Given these challenges, interdisciplinary approaches and advanced computational tools have become indispensable. Collaborations between geographers, computer scientists, statisticians, and domain-specific experts enable the development of tailored solutions that address the unique challenges of geospatial analysis. For instance, the integration of Geographic Information Systems (GIS) with AI and machine learning has opened new avenues for automating tasks such as land cover classification or anomaly detection in spatial data [27]. Cloud-based platforms, such as Google Earth Engine, provide scalable

solutions for processing and analyzing large geospatial datasets, democratizing access to advanced geospatial tools [28].

Consider flood risk assessment as an example. This involves integrating hydrological models, topographic maps, and weather forecast data to predict areas susceptible to flooding. These datasets often vary in resolution and format, requiring significant preprocessing to achieve compatibility. AI algorithms may then be applied to identify high-risk zones, but the results must be validated with historical flood data to ensure reliability [26].

Another practical example of spatial autocorrelation can be found in epidemiology. Researchers mapping the spread of infectious diseases often observe clusters of cases in specific regions, driven by factors like population density or environmental conditions. Identifying these clusters using spatial statistical techniques helps in designing targeted interventions, such as vaccination campaigns or resource allocation, to mitigate disease spread [29].

Given these challenges, interdisciplinary approaches and advanced computational tools have become indispensable. Collaborations between geographers, computer scientists, statisticians, and domain-specific experts enable the development of tailored solutions that address the unique challenges of geospatial analysis. For instance, the integration of Geographic Information Systems (GIS) with AI and machine learning has opened new avenues for automating tasks such as land cover classification or anomaly detection in spatial data [30]. Cloud-based platforms, such as Google Earth Engine, provide scalable solutions for processing and analyzing large geospatial datasets, democratizing access to advanced geospatial tools [31].

In conclusion, geospatial analysis is a rapidly evolving field that demands innovative solutions to overcome technical, computational, and theoretical challenges. As the volume and diversity of geospatial data continue to grow, the need for interdisciplinary collaboration and the development of sophisticated computational methods becomes increasingly critical. By addressing these challenges, researchers can unlock the full potential of geospatial analysis to inform decision-making

and contribute to a deeper understanding of complex spatial phenomena.

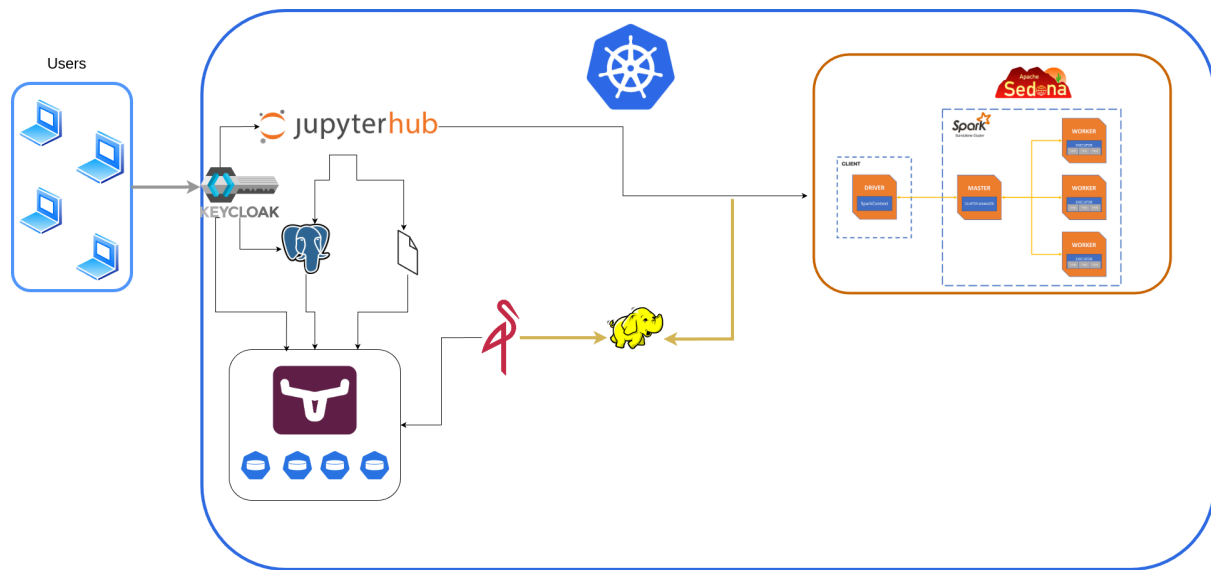
Thesis objectives and purpose

The purpose of this thesis is the design and the development of a user collaborative platform that addresses the multifaceted challenges of geospatial analysis by combining state-of-the-art tools and technologies to create a robust, scalable, and user-friendly analytical environment ready for users to use without having to manage infrastructure and dependencies focusing on what matters most, research. Its objectives and purpose are centered on enabling efficient processing, analysis, and visualization of geospatial data for researchers and practitioners, addressing both technical and computational barriers.

The platform is designed to support efficient geospatial analysis by enabling users to process, manage, and analyze large and complex spatial datasets seamlessly. It facilitates secure access, collaborative workflows, and scalable data processing, addressing challenges like data heterogeneity, computational demands, and storage requirements.

By providing an adaptable and user-friendly environment, the platform empowers researchers and analysts across disciplines to extract meaningful insights from geospatial data, drive innovation, and support decision-making in various applications such as urban planning, environmental monitoring, and disaster management.

Technologies and Tools Used



The above diagram illustrates the architecture of the platform. Below is a breakdown of the components and their roles based on the diagram:

- **Kubernetes:** Serving as the foundation of the platform, Kubernetes orchestrates the deployment, scaling, and management of containerized applications. It ensures high availability, fault tolerance, and efficient resource utilization.
- **JupyterHub:** As a multi-user server for Jupyter notebooks, JupyterHub provides a collaborative workspace for data scientists and analysts. It enables users to share and work on data analysis tasks within a centralized environment.
- **Specialized Jupyter Container Images:** These pre-configured container images offer a streamlined development experience with Python & R environments by bundling essential geospatial software and libraries for data analysis tasks.
- **Keycloak:** This identity and access management (IAM) solution safeguards the platform by providing user authentication, authorization, and single sign-on (SSO) capabilities. Keycloak ensures secure access to resources and protects sensitive data.

- **Apache Spark:** A distributed data processing engine, Spark is optimized for large-scale data analysis tasks. It excels at data cleaning, transformation, and machine learning, delivering exceptional performance and scalability.
- **Apache Sedona:** Apache Sedona is a powerful open-source framework designed to efficiently process and analyze large-scale spatial data and operates as an extension of Spark. It extends these systems with a set of out-of-the-box Spatial Resilient Distributed Datasets (SRDDs) and Spatial SQL, enabling developers to easily perform complex spatial operations on massive datasets.
- **Longhorn:** Longhorn is a distributed block storage system that offers persistent storage for the users applications. It ensures high availability and performance, guaranteeing data durability and accessibility by offering distributed cloud-native persistent block storage on Kubernetes. Efficient backup mechanisms take place too on which critical data can be recovered from a disk failure by re-replicating the entire data store at minimal downtime.
- **MinIO** is an open-source, high-performance object storage system designed to store unstructured data. It is used as the **HDFS** and the overlay storage between JupyterHub and Spark for storing data.
- **PostgreSQL** database is integrated for managing structured data, likely serving as the backend for metadata or intermediate results storage.

Apache Sedona: Capabilities and usage for geospatial analysis.

Apache Spark: Distributed processing of large-scale data.

JupyterHub: Provision of a collaborative and interactive environment.

Kubernetes: Management and organization of infrastructure.

Architecture of the Proposed Infrastructure

End-to-end design of the infrastructure.

Integration of Apache Sedona, Spark, and JupyterHub into Kubernetes.

Implementation of the Infrastructure

Experimental Results and Evaluation

Testing and evaluation of the infrastructure's performance.

Comparison with existing solutions.

Performance analysis in large-scale geospatial data analysis.

Conclusions and Future Work

Recap of the main findings and contributions.

Suggestions for future improvements and extensions.

Bibliography

- [1] Batty, M. (2013). "Big data, smart cities and city planning." *Dialogues in Human Geography*, 3(3), 274-279.
- [2] Foody, G. M. (2002). "Status of land cover classification accuracy assessment." *Remote Sensing of Environment*, 80(1), 185-201.
- [3] Cutter, S. L., Burton, C. G., & Emrich, C. T. (2010). "Disaster resilience indicators for benchmarking baseline conditions." *Journal of Homeland Security and Emergency Management*, 7(1), 1-22.
- [4] Gatrell, A. C., & Elliott, S. J. (2015). "Geographies of health: An introduction." *Wiley-Blackwell*.

- [5] Zhang, C., & Kovacs, J. M. (2012). "The application of small unmanned aerial systems for precision agriculture: A review." *Precision Agriculture*, 13(6), 693-712.
- [6] NRC (National Research Council). (2003). "Mapping the zone: Improving flood map accuracy." *The National Academies Press*.
- [7] Behrens, R. (2004). "GIS-based modeling and analysis of energy distribution systems." *Energy Policy*, 32(17), 2013-2022.
- [8] Chainey, S., & Ratcliffe, J. (2005). *GIS and Crime Mapping*. Wiley.
- [9] Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill.
- [10] Hewitson, B. C., & Crane, R. G. (2006). "Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa." *International Journal of Climatology*, 26(10), 1315-1337.
- [11] "Geospatial Analysis: A Comprehensive Guide" *Authors: Paul A. Longley, Goodchild, M.F., Maguire, D.J., & Rhind, D.W.*
- [12] What is Geodata? A Guide to Geospatial Data (<https://gisgeography.com/what-is-geodata-geospatial-data/>)
- [13] QGIS (<https://www.qgis.org/>)
- [14] ArcGIS (<https://www.arcgis.com/>)
- [15] "Comparative Assessment of the Unsupervised Land Use Classification by Using Proprietary GIS and Open Source Software" *Authors: Muhammad Azizol Ismail, Ahmad Nazri Muhamad Ludin. , Nafisa Hosni.*

- [16] Kluyver, T. et al. "Jupyter Notebooks — A publishing format for reproducible computational workflows". Positioning and Power in Academic Publishing: Players, Agents and Agendas (2016),
- [17] Granger, B. and Grout, J. "JupyterLab: Building Blocks for Interactive Computing". Slides of presentation made at SciPy'2016. 2016.
- [18] Gabriele Silveira Camara, Silvana Philippi Camboim, João Vitor Meza Bravo "USING JUPYTER NOTEBOOKS FOR VIEWING AND ANALYZING GEOSPATIAL DATA: TWO EXAMPLES FOR EMOTIONAL MAPS AND EDUCATION DATA"
- [19] Goodchild, M. F. (2018). Spatial Autocorrelation and Geographic Analysis. *GeoJournal*, 1(1), 15-22. doi:10.1007/s10708-018-9734-5
- [20] Schuurman, N. (2004). *GIS: A Short Introduction*. Wiley-Blackwell.
- [21] Brunsdon, C., & Comber, L. (2015). *An Introduction to R for Spatial Analysis and Mapping*. SAGE Publications.
- [22] Cressie, N., & Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley.
- [23] Li, S., Dragicevic, S., & Castro, F. A. (2020). Big Data Analytics in GIS: A Comprehensive Review. *ISPRS International Journal of Geo-Information*, 9(3), 124.
- [24] Lary, D. J., et al. (2016). Machine Learning in Geospatial Analysis. *Environmental Modelling & Software*, 80, 1-12.
- [25] Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93-115.
- [26] Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

- [27] Hu, T., & Lo, C. P. (2007). Modelling Urban Growth in Atlanta Using Logistic Regression. *Geographical Analysis*, 39(4), 1-25.
- [28] Gorelick, N., et al. (2017). Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone. *Remote Sensing of Environment*, 202, 18-27.
- [29] Bailey, T. C., & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Longman Scientific & Technical.
- [30] Jianya, G., et al. (2011). Object-Based Image Analysis for Urban Mapping. *Journal of Remote Sensing*, 15(3), 345-360.
- [31] Gorelick, N., et al. (2017). Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone. *Remote Sensing of Environment*, 202, 18-27.

Appendices

Technical details, code, and additional results.