Building OptiPath: A Cloud-Based system for Optimal Route Calculation using ArcGIS Enterprise

Haoyu Wang[1, 3], Devika Jain[1], Jacob Greenspon[2,4] Jeff Blossom[1]

[1] Center for Geographic Analysis, Harvard University, USA - (haoyuwang, kakkar, jblossom)@fas.harvard.edu
[2] Harvard Kennedy School – jgreenspon@hks.harvard.edu
[3] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, China - hywang@mail.bnu.edu.cn
[4]University of Oxford, UK - jacob.greenspon@economics.ox.ac.uk

Keywords: Cloud Computing, Raster Big Data, Geospatial, ArcGIS Enterprise, Routing.

Abstract

This research presents OptiPath, a system tailored for optimizing route planning within big geospatial data projects. By harnessing extensive raster data and leveraging cloud computing resources, OptiPath excels in spatial analysis and path optimization for com- plex geospatial applications. In a project involving a historical U.S. pipeline simulation, OptiPath generated 8 optimal paths, each processing 2.50 GB of raster data in an average time of 10 minutes per path. Consequently, the system efficiently processed 20GB of raster data within 1.5 hours. OptiPath accurately computes optimized paths between specified origins and destinations based on user input, aligning closely with actual routes in a time and cost-effective manner. This system's outstanding performance makes it an asset for diverse and complex route planning projects. OptiPath operates on the New England Research Cloud (NERC), running on Esri's ArcGIS Enterprise software. Beyond offering an effective, efficient routing solution, OptiPath underscores the immense potential of integrating cloud computing and ESRI's big data tools for GIS applications.

1. Introduction

Advances in Geographic Information System (GIS) technology enable processing massive amounts of data at extremely high spatial and temporal resolutions. The integration of cloud computing with GIS marks a significant advance in the capabilities for processing and analyzing GIS data. Cloud computing offers a flexible, scalable, and efficient platform for storing, managing, and analyzing large volumes of geospatial data, overcoming the limitations of traditional GIS infrastructure (Noraziah et al., 2017). Hashem et al. (2015) emphasized the transformative potential of cloud computing in handling big data, highlighting its role in providing access to powerful computing resources without the need for substantial upfront investments in hardware and software.

Optimal path or "least cost path" analysis is a commonly used type of GIS analysis, with a goal of finding the most cost-optimizing route from an origin to a destination (Meisingset et al., 2004). Costs to build a pipeline may include a variety of factors such as economic costs and environmental impacts (Abudu and Williams, 2015). This requires considering multiple factors during planning, such as the slope or gradient of the area, available land use, and soil types. The diverse considerations and interests involved make the planning process reliant on extensive computations with raster big data. In such cases, establishing a GIS system for optimal route calculation based on cloud computing platforms becomes essential.

We developed OptiPath, a cloud-based optimal route calculation system designed to provide a method for calculating the optimal routes in the field of urban planning and development. Its core goal is to harness the power of cloud computing to of- fer a robust, efficient, and scalable solution for route optimization challenges. By integrating with ArcGIS Enterprise on the New England Research Cloud (NERC), OptiPath utilizes the latest Geographic Information System (GIS) technologies and cloud resources to analyze and determine the least cost paths across urban and rural landscapes. We used the OptiPath system to calculate the optimal pipeline route in the United States in 1942. The case study not only demonstrates the system's efficiency and effectiveness in route optimization but also serves as a model for leveraging technology in achieving sustainable urban development.

2. Case Study: U.S. Pipeline Simulation Project

This case study examines the application of the OptiPath system in a historical context, focusing on modeling the most efficient routes for constructing oil and gas pipelines in the United States during the 1940s. This study aims to investigate economic effects on local communities that happened to gain early access to natural gas due to pipelines built nearby. In World War II, the United States Government built "emergency pipelines" to transport crude oil from Texas in the Southeast to multiple locations in the Northeast for refinery and shipment. This project seeks to answer the research question "Did increased energy access due to being near a major gas pipeline route have long-run effects on location of manufacturing?" (Greenspon and Hanson, 2023). The U.S. Government was trying to build the pipelines as fast and as efficiently as possible given wartime construction constraints and therefore prioritized the most direct and inexpensive route. For this part of the project, we want to test if the wartime pipelines built follow the optimal path to build a pipeline when considering topography, land cover, and materials costs. This would mean that communities were situated near the pipeline route post-WW2 "quasi-randomly" because of its least-cost construction routing, rather than because of external factors. For example, for economic reasons the pipeline route may have been directed toward large potential markets for the gas it carried. This would allow for a comparison of long-term manufacturing production between communities near vs. distant to the pipeline.

In this project, three major pipelines were modeled: The Big Inch, Little Inch, and TN Gas Pipelines. These pipelines will follow the paths defined below:

- Big Inch Pipeline (TETCO)
  - Start: Longview, TX
  - Midpoints: Little Rock, AR; Norris City, IL; Phoenixville, PA (splitting into two afterwards)
  - Endpoints: Chester Junction, PA and Linden, NJ
- Little Inch Pipeline (TETCO)
  - Start: Baytown, TX
  - Midpoint: Little Rock, AR
  - Follows the same route as the Big Inch Pipeline afterward
- TN Gas Pipeline
  - Start: Banquette, TX
  - Initial Endpoint: Cornwell Station, WV
  - Extension (built later): From West Bend, KY to Buffalo, NY

This creates a total of 8 segments of pipeline that need to be modelled. The objective is to calculate the least cost path to build these pipelines based on physical factors alone. Specifically the costs are, the topographic slope of the land, the land cover itself, and materials cost per mile. Land cover categories include wetlands, developed land, agricultural areas, forests, and water bodies. Topographic slope was calculated from the 90m spatial resolution elevation data source Shuttle Radar Topography Mission (SRTM - (`https://bigdata.cgiar.org/srtm-90m-digital-elevation-database`). For land cover, we used the historic land cover modelling by the U.S. Geological Survey (USGS). The USGS produced these models yearly, between 1938 and 1992 at a resolution of 250m. (`https://www.sciencebase.gov/catalog/item/59d3c73de4b05fe04cc3d1d1`). We downloaded and utilized the land cover raster for the year 1942 for this project. Relative costs ranging from 1 (easy travel) to 30 (extremely difficult) were assigned to each land cover type. These are listed in Table 1 below.

Pixel Characteristic Group	Land Cover (1942)	Cost
Wetlands	15 - Herbaceous Wetland; 16 - Wetland	15
Developed	2 - Urban/Developed	23
Agricultural	13 - Cultivated Cropland	7
Forest	8 - Deciduous Forest; 9 - Evergreen Forest; 10 - Mixed Forest	15
Water (major rivers/lakes)	1 - Open Water	30
Rocky Land	6 - Mining; 7 - Barren	12
Grassland/Brush/Pasture	11 - Grassland; 12 - Shrubland; 14 - Hay/Pasture	2

Table 1. Pixel characteristic group and land cover classification.

3. Literature Review

The determination of optimal routes represents a complex and multi- faceted challenge, as evidenced by various studies across different domains. Lebedeva et al. (2020) outlines the inherent complexities in calculating optimal routes, underscoring the critical role of GIS tools in navigating these challenges. Macharia and Mundia (2022) and Hamid-Mosaku et al. (2020) both demonstrate the application of GIS analysis in identifying optimal routes for pipeline projects, with a particular emphasis on the importance of environmental and social considerations. Additionally, the selection of optimal paths for rail lines and power transmission has been effectively addressed using GIS technology by De Luca et al. (2012) and Eroglu and Aydin (2015) respectively. These studies collectively underscore the indispensable value of GIS in route optimization, especially within the pipeline sector, highlighting the necessity to account for a myriad of factors in the decision-making process.

ArcGIS, as a leading GIS platform, offers an extensive array of tools for path optimization and spatial analysis. Among these, the network analysis tool is pivotal for selecting the optimal path by considering impedance fac- tors such as road length, as detailed by Tan (2021). The platform also encompasses various spatial statistical tools for analyzing feature patterns, as discussed by Scott and Janikas (2009). The incorporation of the ArcGIS Geodatabase Network Model and the Dijkstra algorithm significantly enhances the platform's capabilities for conducting thorough path analysis, as detailed by Cao et al. (2010). In this context, ArcGIS Enterprise emerges as a corner-stone for developing efficient, cloud-based GIS applications, enabling intricate route calculations and advanced spatial data analysis. This enterprise solution leverages the power of cloud computing to facilitate scalable and collaborative GIS applications, underscoring its critical role in supporting complex geo- spatial analyses and optimizations.

The integration of cloud computing and big data technologies with GIS has significantly propelled the field of path optimization. Junior et al. (2017) and Mahmoud et al. (2013) both underscore the cost-effectiveness and enhanced performance of processing geospatial data through public cloud providers. Yang et al. (2017) delves into the utilization of cloud computing to navigate the challenges posed by big geospatial data, highlighting its potential to transform big data's inherent characteristics into actionable insights for smart cities and geospatial science. Further, Yang et al. (2011) explores the symbiotic relationship between geospatial sciences and cloud computing, revealing the vital role of cloud computing in overcoming the substantial in- formation technology challenges faced by the geospatial sciences, particularly in managing and analyzing spatial data at scale.

4. System Design

The New England Research Cloud (NERC) is a professionally operated regional resource that provides on-premises cloud service. The varying level of services include self-service Software-as-a-Service (SaaS) allowing easy access to users, automated Platform-as-a-Service (PaaS) for custom workflow creation, and Infrastructure-as-a-Service (IaaS) which catalyzes hardware acceleration. NERC is commonly used by researchers in the Boston area to facilitate quality research using more-efficient computational environments (NERC, 2024; MOC, 2024).

NERC OpenStack (2024) is an open-source cloud computing platform that allows organizations to build and manage both public and private clouds. It provides a set of software tools for building and managing cloud infrastructure as a service (IaaS). OpenStack is designed to be scalable and flexible, enabling users to create and manage virtualized resources such as compute instances, storage, and networking. It provides a versatile and open alternative to proprietary cloud solutions, offering flexibility and control over cloud infrastructure. Figure 1 below shows the user-friendly interface of NERC OpenStack which enables easy install and deployment of applications.

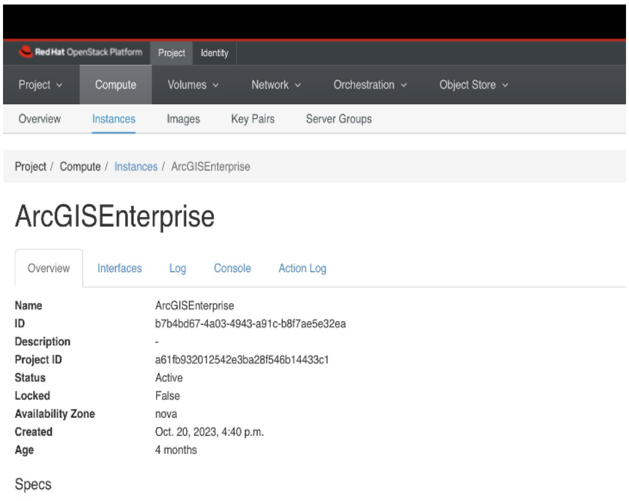

Figure 1. ArcGIS Enterprise deployed on NERC OpenStack Platform.

ArcGIS Enterprise is a comprehensive geospatial platform developed by Esri (Environmental Systems Research Institute). ArcGIS Enterprise allows organizations to create, manage, share, and analyze spatial and location- based data in a collaborative and secure environment (Esri, 2024). It provides a set of integrated components and services for building and deploying GIS applications within an enterprise setting. Key com- ponents of ArcGIS Enterprise include ArcGIS Server, Portal for ArcGIS, ArcGIS Data Store, ArcGIS Web Adaptor, ArcGIS GeoEvent Server and ArcGIS Image Server. ArcGIS Enterprise allows organizations to create and deploy a wide range of GIS applications.

ArcGIS Enterprise was installed on NERC OpenStack which allows for easy scalability, enabling us to expand or shrink the infrastructure based on demand. This flexibility ensures optimal performance and resource utilization, accommodating varying workloads of the project.

5. Methodology

In OptiPath, calculating the optimal path can be divided into three steps: data processing, cost surface calculation, and optimal path calculation, as shown in Figure 2. Below, these three steps are introduced sequentially.

5.1 Data Processing

When searching for the optimal path, various factors need to be considered, such as elevation, land cover, and topography. In OptiPath, these factors must be input in the form of raster

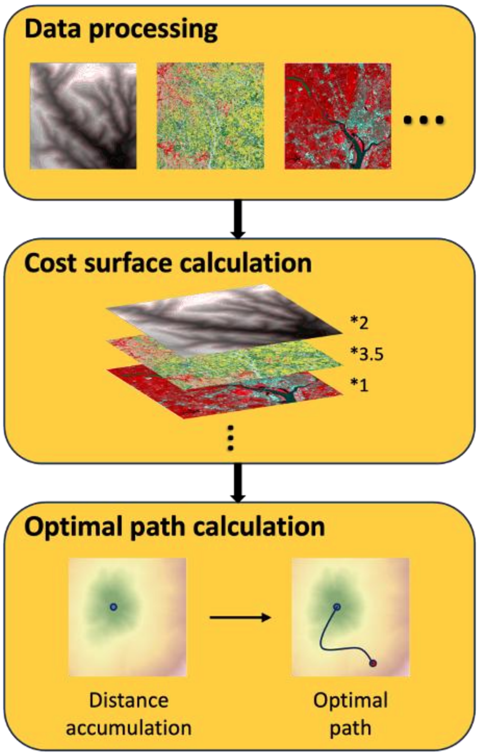

Figure 2. Optimal path calculation flowchart.

data. Based on the decisionmaker's judgment, different travel costs are assigned to different grid cells, in other words the cost incurred when the path passes through that grid cell. The higher the value, the higher the cost. This process was accomplished using the ArcGIS Pro Reclassification tool. It's important to note that to ensure all raster data can be smoothly aggregated in the next step, all the rasters need to be resampled to the same resolution and aligned before reclassification. This was achieved using the Resample tool. This allowed us to achieve the goal of aligning the cells of all the rasters and converting them to the same resolution of 250m.

5.2 Cost Surface Calculation

This step involves merging all the processed raster data from the first step into a total cost surface raster. The specific method is to assign weights to each influencing factor according to the decisionmaker's judgment, and then to sum the weighted the rasters to obtain the final cost surface. This cost surface reflects the comprehensive cost of constructing paths through each grid cell. This process was achieved using the Raster Calculator tool. The determination of weights has an almost decisive im- pact on the generation of the final path, so weights should be chosen carefully. Multiple simulations with different weights may need to be run to determine weights that produce the most realistic results.

5.3 Optimal Path Calculation

This is the most important and core part of the formal optimal path calculation process. This step is divided into two stages, each implemented with two tools. First, on the cost surface generated in the second step, the distance accumulation raster and back direction raster from the path start point are calculated. Both rasters were created using the Distance Accumulation tool. Then, based on the distance accumulation raster and back direction raster, the lowest cost path from the start points to the endpoint is calculated. This was achieved using the Optimal Path as Line tool.

6. Results

The input DEM has a resolution of 90 meters, whereas the land cover data has a resolution of 250 meters. To align the DEM data with the land cover data in terms of projection, range, and resolution, the Clip, Project Raster, and Resample tool were sequentially applied. Alignment of the two raster datasets was ensured by assigning the historic land cover raster as the 'snap raster.' Subsequently, the percentage slope of the DEM data was calculated using the Slope tool, identifying slope and land cover as the cost factors for the pipeline path. Corresponding cost values were then assigned to the resulting raster. Using the Reclassify tool, grids with a slope greater than 30 percent slope were set to 'NoData', indicating areas unsuitable for pipeline crossing due to steep slopes. The cost values for other grids were determined based on their raster values, with higher slopes indicating higher costs. Similarly, for the land cover data, cost values for each land cover type were assigned based on the decisionmaker's analysis, listed above in Table 1.

The slope cost and landcover cost rasters were merged into a single cost surface. Based on the analysis by decision makers and subsequent trials, the formula for calculating the cost sur- face was determined as:

$$\text{cost}_i = \text{slope}_i + \text{LC}_i + 10 \qquad (1)$$

where cost_i represents the total cost value for raster cell_i, slope_i indicates the slope cost value, and LC_i denotes the land cover cost value for the same cell. Moreover, to limit the pipeline length, a constant term is introduced (+ 10), which effectively prioritizes the cost implications of pipeline materials. By ap- plying equation (1) through the Raster Calculator tool, the final cost surface is derived, as illustrated in Figure 3.

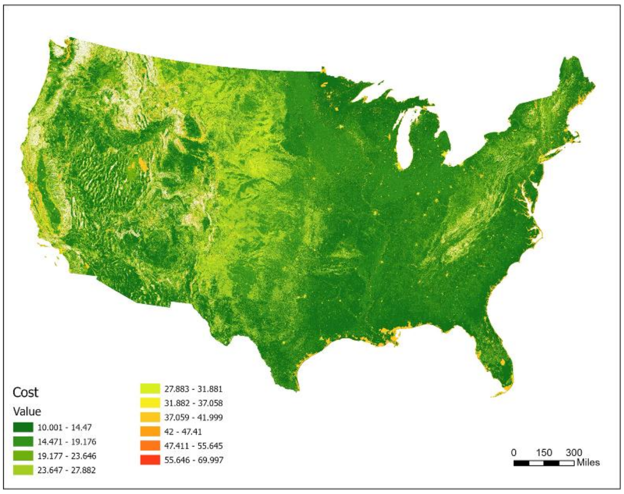

Figure 3. Cost surface.

After obtaining the cost surface, the next step involved generating the segments of the path to be calculated. To accomplish this, three main routes were calculated and divided into eight specific segments for detailed analysis. These segments are:

- Longview, TX to Little Rock, AR
- Little Rock, AR to Norris City, IL
- Norris City, IL to Phoenixville, PA
- Phoenixville, PA to Chester Junction, PA
- Phoenixville, PA to Linden, NJ
- Baytown, TX to Little Rock AR
- Banquette, TX to Cornwell Station, WV
- West Bend, KY to Buffalo, NY.

To facilitate the operation, a model was constructed in ArcGIS Pro's Model Builder software module. This involves chaining the analytical tools together to produce the desired output. See the model presented in Figure 4 below.

Figure 4. Model for optimal calculation of the 8 paths.

Once built, the model was published to ArcGIS Enterprise for execution. This model was applied to all 8 segments, to simultaneously generate all 8 path segments.

Running the model, the calculation of all 8 optimal path segments was completed in 1 hour and 8 minutes. For each segment, a total of 2.50 GB of raster data (cost surface + distance accumulation raster + back direction raster) needed to be processed, with an average time of 8.5 minutes per segment. Our cloud computing system efficiently and stably accomplished the optimal path calculation.

To validate the accuracy of the computational results, they were mapped alongside the actual pipeline paths from 1949 (see Figure 5, below). These results demonstrate that the calculated lowest cost paths for the Big Inch Pipeline, Little Inch Pipeline, and the initial TN Gas Pipeline closely align with the historical routes. This alignment highlights the model's precision, reflecting the historical context of resource optimization during war-time. The route from West Bend, KY to Buffalo, NY pertains to a post-war extension, where factors beyond cost (e.g. market potential) influenced the path, explaining the slight deviations from the calculated least cost path. This comparison emphasizes the reliability of the optimal paths determined by OptiPath in replicating real-world scenarios.

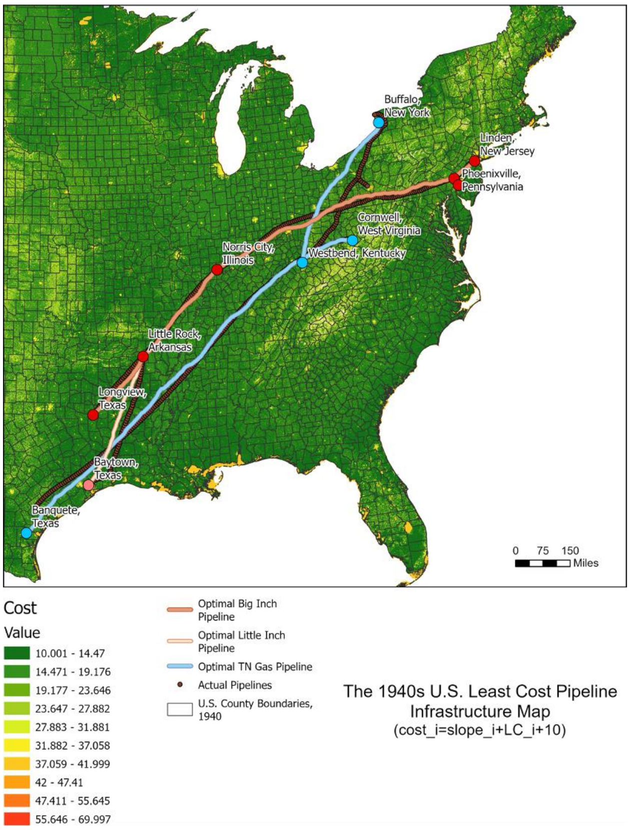

Figure 5. Calculated optimal paths and actual paths map.

7. Conclusion and Next Steps

This paper has introduced OptiPath, an optimal path calculation system based on a cloud platform utilizing ArcGIS Enterprise. Hosted on the cloud computing platform NERC, OptiPath leverages the sharing capabilities of ArcGIS Enterprise along with GIS tools in ArcGIS Pro to compute optimal paths. To validate OptiPath's efficacy, a case study calculating the lowest cost path for a 1942 U.S. natural gas pipeline was conducted. The results indicate:

- OptiPath efficiently and stably processed 20GB of raster data, generating 8 optimal paths in just 1 hour and 8 minutes.
- The generated optimal paths closely match the actual war-time paths, demonstrating the accuracy of the results.
- Utilizing OptiPath for optimal path calculations can save local resources for users, enhancing computational efficiency and stability.

While OptiPath is adept at calculating least cost paths for pipeline construction, it can also be applied to calculate any other raster based optimal path calculations. Additionally, by leveraging the cloud platform OptiPath has the potential for expanding computational resources to further enhance computational efficiency; utilizing ArcGIS Enterprise's sharing services to enable collaborative projects; improving tool integration to lower barriers to use; and integrating more tools could enrich the system's functionality.

Acknowledgements

We would like to thank the New England Research Cloud (NERC) for providing the technical support and computing resources needed for the project. We would also like to thank ESRI for providing us the software license to their product. Additionally, the China Scholarship Council (CSC) supported Haoyu Wang during this research.

The authors' contributions are as follows: Study conception and design by Jacob Greenspon and Jeff Blossom; Software and computing design by Devika Jain; Software development by Haoyu Wang; Manuscript preparation by Haoyu Wang, Devika Jain, and Jeff Blossom; Overall manuscript editing by Jeff Blossom and Jacob Greenspon. All authors participated in reviewing the results and approved the final version of the manuscript.

References

Abudu, D., Williams, M., 2015. GIS-based optimal route selection for oil and gas pipelines in Uganda. *Advances in Computer Science: an International Journal*, 4(4), 93–104.

Cao, J., Song, Z., Wang, Y., Wang, T., 2010. Analysis and application of the optimal path based on arcgis geodatabase'network model. *2010 The 2nd Conference on Environmental Science and Information Application Technology*, 1, IEEE, 434–437.

De Luca, M., Dell'Acqua, G., Lamberti, R., 2012. High-speed rail track design using GIS and multi-criteria analysis. *Procedia-Social and Behavioral Sciences*, 54, 608–617.

Erog˘lu, H., Aydin, M., 2015. Optimization of electrical power transmission lines"routing using AHP, fuzzy AHP, and GIS. *Turkish Journal of Electrical Engineering and Computer Sciences*, 23(5), 1418–1430.

Esri, 2024. What is Arcgis enterprise? `https://enterprise.arcgis.com/en/get-started/latest/windows/what-is-arcgis-enterprise-.htm`. Accessed: 2024-02-25.

Greenspon, J., Hanson, G., 2023. Local energy access and industry location: evidence from world war ii 'emergency pipelines'. *Pre-Conference Workshop on Economic Impacts of World War II*, National Bureau of Economic Research (NBER).

Hamid-Mosaku, I. A., Oguntade, O. F., Ifeanyi, V. I., Balo- gun, A.-L., Jimoh, O. A., 2020. Evolving a comprehensive geomatics multi-criteria evaluation index model for optimal pipeline route selection. *Structure and Infrastructure Engineer- ing*, 16(10), 1382–1396.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., Khan, S. U., 2015. The rise of “big data” on cloud computing: Review and open research issues. *Information systems*, 47, 98–115.

Junior, J. B., Sousa Reis, M. A., Araujo, A. P. d., Holanda, M., 2017. Cost optimization on public cloud provider for big geospatial data. *Proceedings of the 7th International Conference on Cloud Computing and Services Science*, 82–90.

Lebedeva, O., Poltavskaya, J., Gozbenko, V., Mikhailov, A., 2020. Solving a dynamic routing problem using an optimization algorithm. *Journal of Physics: Conference Series*, 1614num- ber 1, IOP Publishing, 012101.

Macharia, P. M., Mundia, C. N., 2022. Gis analysis and spatial modelling for optimal oil pipeline route location. a case study of proposed isiolo nakuru pipeline route. *Proceedings of the Sustainable Research and Innovation Conference*, 91–94.

Mahmoud, E., Hegazy, O., El-Dien, M. N., 2013. Integration of GIS and cloud computing for emergency system. *International Journal of Engineering and Computer Science*, 2(10), 2889–2893.

Meisingset, H., Hove, J., Olsen, G., 2004. Optimization of pipeline routes. *ISOPE International Ocean and Polar Engineering Conference*, ISOPE, ISOPE–I.

MOC, 2024. Massachusetts open cloud alliance. `https://massopen.cloud/`. Accessed: 2024-01-11.

NERC, 2024. Massachusetts green high-performance computing center. `https://nerc.mghpcc.org/`. Accessed: 2024-01-11.

NERC OpenStack, 2024. New england research cloud(nerc). `https://nerc-project.github.io/nerc-docs/openstack/`. Accessed: 2024-02-25.

Noraziah, A., Fakherldin, M. A. I., Adam, K., Majid, M. A., 2017. Big data processing in cloud computing environments. *Advanced Science Letters*, 23(11), 11092–11095.

Scott, L. M., Janikas, M. V., 2009. Spatial statistics in arcgis. *Handbook of applied spatial analysis: Software tools, methods and applications*, Springer, 27–41.

Tan, Y., 2021. Research on optimal path selection based on webgis. *Journal of Physics: Conference Series*, 1871number 1, IOP Publishing, 012138.

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Bambacus, M., Fay, D., 2011. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? *International Journal of Digital Earth*, 4(4), 305–329.

Yang, C., Yu, M., Hu, F., Jiang, Y., Li, Y., 2017. Utilizing cloud computing to address big geospatial data challenges. *Computers, environment and urban systems*, 61, 120–128.