Database Administration
Final Project

Andrés Oswaldo Calderón Romero, Ph.D.

October 20, 2025

1 Introduction

The rapid transformation of energy infrastructure in Colombia calls for innovative approaches to identify suitable locations for the deployment of renewable energy solutions. As part of a broader national strategy to promote sustainable development and improve energy access in underserved regions, the Research and Information Office of the Unidad de Planeación Minero Energética (UPME) has launched a strategic initiative focused on evaluating the feasibility of solar energy in selected territories.

This project supports UPME's objective by designing and implementing a reproducible geospatial analysis workflow to estimate the solar energy potential of building rooftops across prioritized municipalities. In particular, the project emphasizes the analysis of PDET (Programas de Desarrollo con Enfoque Territorial) territories, which represent a key focus area for post-conflict development and infrastructure enhancement in Colombia.

Leveraging openly available geospatial datasets the project aims to estimate the number and area of building rooftops suitable for solar panel installation. These datasets, encompassing billions of building outlines derived from high-resolution satellite imagery, provide a valuable resource for quantifying potential energy-harvesting surfaces in urban and rural contexts.

To align with the UPME's technological infrastructure and modernization goals, the methodology is designed using NoSQL data solutions, enabling scalable storage, efficient querying, and flexible spatial operations. The outcome of this work will be a technical report outlining the methodology, results, and recommendations for identifying optimal locations for proof-of-concept solar farms in the selected regions.

By integrating modern data science tools with real-world energy policy needs, this project bridges technical innovation and strategic planning in support of Colombia's energy transition and territorial equity.

2 Problem Statement and Scope

The Research and Information Office of UPME (Unidad de Planeación Minero Energética) in Colombia is seeking a strategic partner to support the initial phases of a major initiative aimed at identifying potential locations for proof-of-concept projects involving alternative energy solutions. Among these, UPME plans to assess the feasibility of solar farms in various regions across the country.

In order to select potential locations, stakeholders aim to prioritize local governments (municipalities) that have been designated as PDET[1] territories. The primary goal of the project is to determine the number of buildings within each municipality and to estimate the total rooftop area suitable for solar panel installation. The greater the rooftop area in a municipality, the higher the potential for collecting clean energy through solar panels installed on those roofs. UPME requires a **detailed report** outlining a reproducible methodology for counting the number of rooftops and aggregating their total area for each PDET municipality evaluating different datasets.

Given the new technologies and infrastructure acquired by the Research and Information Office, a key requirement is the application of NoSQL solutions in the proposed methodology. Additionally, the Office mandates the evaluation of open datasets to query and identify buildings within the target territories, with the goal of comparing the outputs from each source. Two open datasets have been selected as the primary sources for this study:

- **Microsoft Building Footprints:** Bing Maps has released open building footprint data covering various parts of the world. The dataset includes over 999 million building detections derived from Bing Maps imagery collected between 2014 and 2021, incorporating sources such as Maxar and Airbus. The data is freely available under the Open Data Commons Open Database License (ODbL). More information is available at: https://planetarycomputer.microsoft.com/dataset/ms-buildings.

- **Google Open Buildings:** This dataset contains 1.8 billion building detections spanning an inference area of 58 million km^2, covering Africa, South Asia, Southeast Asia, Latin America, and the Caribbean. Currently in its third version, the dataset is distributed under the Creative Commons Attribution (CC BY-4.0) and Open Data Commons Open Database License (ODbL) v1.0. More information is available at: https://sites.research.google/gr/open-buildings/.

In addition to the aforementioned datasets, the Colombian administrative agency DANE (Departamento Administrativo Nacional de Estadística) provides the *Marco Geoestadístico Nacional* (MGN)[2] at various administrative levels, including the boundaries of all Colombian municipalities[3]. However, as previously mentioned, only the municipalities designated as PDET territories will be considered in this analysis.

3   Proposed Deliveries

Each deliverable is expected to be submitted each week and defended on sessions at October 27, November 10 and November 24. The submission will be done via commits in the GitHub project repository of each group.

1. **NoSQL Database Schema Design and Implementation Plan**: This deliverable focuses on meeting the key requirement of using NoSQL solutions for scalable storage and

---

[1]Programas de Desarrollo con Enfoque Territorial.

[2]MGN User Guide v. 2.0 (in Spanish): https://geoportal.dane.gov.co/descargas/descarga_mgn/Manual_MGN.pdf

[3]https://geoportal.dane.gov.co/servicios/descarga-y-metadatos/datos-geoestadisticos/?cod=111

efficient spatial operations. Expected delivery on October 27 before 2:00pm. It will lay the foundational data structure. This delivery should include:

- Implementation Plan.
- Data Modeling.
- Schema Design & Appropriateness.

2. **PDET Municipality Boundaries Dataset Integration**: This deliverable ensures the project focuses only on the mandated PDET territories, using the required DANE/MGN administrative boundaries. Expected delivery on November 3 before 2:00pm. This will be checked by:

- Data Acquisition & Verification.
- Data Integrity & Format.
- NoSQL Spatial Integration.
- Documentation of Process.

3. **Building Footprint Data Loading and Integration Report (Microsoft & Google Datasets)**: This task addresses the mandate to evaluate and compare outputs from the Microsoft and Google open datasets. Expected delivery on November 10 before 2:00pm. We expect for this delivery:

- Same than previous delivery (including spatial indexing).
- Data Loading Efficiency.
- Initial Data Audit (EDA).

4. **Reproducible Geospatial Analysis Workflow (Rooftop Count and Area Estimation)**: This is the core task, fulfilling the primary goal of the project to estimate the number and total area of building rooftops in each PDET municipality. Expected delivery on November 17 before 2:00pm. The delivery should include:

- Reproducibility & Methodology.
- Accuracy of Spatial Operations.
- Output Data Structure (tables and maps).

5. **Final Technical Report and Recommendations**: This comprehensive report is the final deliverable, summarizing the entire project and providing the required recommendations for UPME. Expected delivery on November 24 before 2:00pm. Here we expect the final report and evaluate:

- Documentation of the whole process.
- Results and data visualizations.
- Content & Completeness.
- Clarity of Recommendations.
- Alignment with UPME Objectives.

4   Conclusions

This project presents the framework for an analysis aimed at addressing a real-world challenge: the identification and quantification of solar energy potential in developing regions of Colombia. By enabling the loading, processing, and analysis of geospatial data, the proposed solution facilitates a comprehensive evaluation of potential locations through a NoSQL-based approach. The inclusion of both theoretical and empirical analyses ensures a robust assessment of the available data and contextual conditions. The final deliverable, **a detailed technical report**, will serve as a solid foundation for selecting the most appropriate methodologies and locations in future implementation phases.