Project Group 3

Johnathan Salamanca, Mario Cerón,
Carol Martinez, Javier Cocunubo, Jairo Nino, Alvaro Munoz

December 9, 2019

1 Introduction

In 2012 the National System of Risk Management (NSRM) was created in Colombia. The system includes public, private, and community entities that will work closely with the government to coordinate the different risk management procedures. The NSRM is comprised of 6 instances:

- The National Risk Management Council (Consejo Nacional para la Gestion de Riesgo): coordinates the national system. At the head is the President and his govement.

- The National Risk of Disaster Management Unit UNGRD (Unidad Nacional para la Gestin del Riesgo de Desastres): it coordinates the nacional system and manage the risk management system.

- National Comittee for Risk Awareness (Comité Nacional para el Conocimiento del Riesgo): advises and plans the constant implementation process of risk awareness

- National Comittee for Risk Reduction (Comité Nacional para la Reduccin del Riesgo): it advises and plans the implementation of the process to reduce the risk of disasters.

- National Comittee for Risk Management (Comité Nacional para el Manejo de Desastres): it advises and plans the implementation of the process of disaster management

- City and Departmental Risk Management Council (Consejos departamentales distritales y municipales para la Gestión del Riesgo): they coordinate, advise, plan and control the processes of risk management in each territorial subdivision.

All six instances are responsible of preventing and managing possible disasters that occur in the country.

In April 2018, the National Planning Department (DNP) presented a report [?] that shows the national situation of the Risk Management in Colombia. The report presents a general overview of Disaster Risk in the world and the situation of Colombia in that matter.

Some of the information from that report is summarized as follows:

International Situation

- From 1980's the disasters have triplicate worldwide. 90% of disasters are hydrometeorological and generate 74% financial losses (e.g. Japan Tsunami, Katrina Hurricane, Japan Earthquake).

- The number of deaths due to disasters is higher in developing countries that in develop countries.

- Countries with high incomes are the ones that have more policy frameworks on risk management.

National Situation

- 88% of the disasters in Colombia are hydrometeorological (Inundaciones, movimientos de masa, flujo torrenciales, sequias e incendios, geologicos, otros).

- Infraestructure looses increase by Nina and Nino natural phenomena.

- Colombian departments with less incomes are the ones that have more people affected during the disasters.

Additionally, the report introduces the Risk Management Index of Colombia adjusted on the basis of capacities. The index measures the risk of a territorial subdivision under hydrometeorologic events and the capacity of that subdivision to manage the risk. The index takes into account two indexes: the risk index and the capacity index. The risk component analyzes the thread, exposition, and vulnerability to a risk. Additionally, the capacity to manage the risk is analyzed based on the economic point of view, socio-economic, and risk management.

The index was created based on the following information:

- 15% of deaths are due to slow flooding (generated by constant and heavy rain that increases the rivers levels) and 85% of the homes affected during a disaster are due to this phenomena.

- Landslide: it causes 19% of death and 1% of affected homes.

- Torrential flow: it causes 66% of death and 14% of affected homes.

- 29% of the national territory has conditions of critical thread of hydrometereological phenomena.

- 13% of the population are socially vulnerable and are highly exposed to the most critical hydrometereological threads.

- Colombia territorial subdivisions have heterogeneous capacity of risk management.

Figure 1 describes the country situation on the basis of the 3 indexes: the capacity index (image on the left), the disaster risk index (image in the center), and the risk management index that combines both (image on the left).

2 Multiple Problem Versions

Version 1

The Municipal Capacities-Adjusted Disaster Risk Index is an innovative indicator for policymakers to make informed decisions about how to better preserve citizensẃell-being in the presence of real and potential threats. However, to be actionable, information needs not only to be available
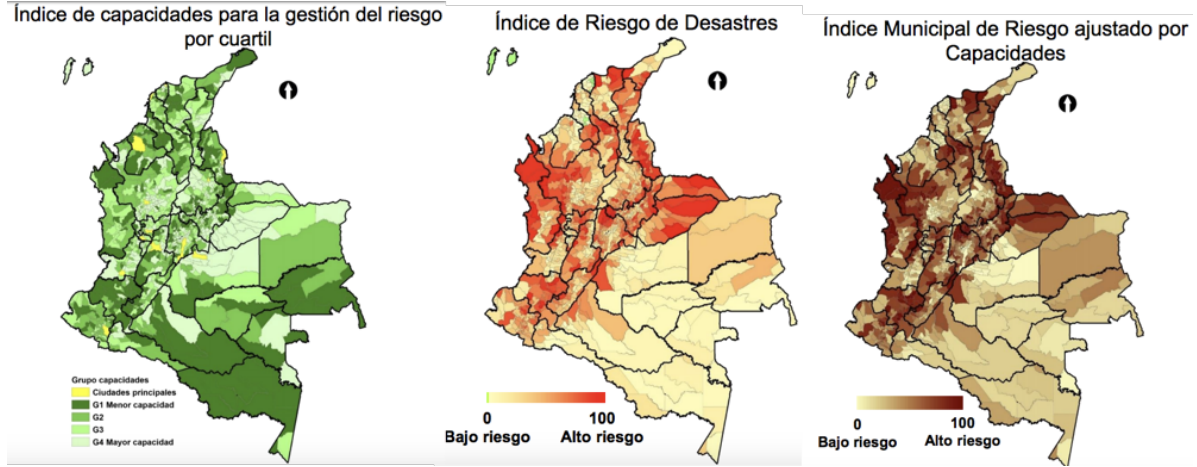
Figure 1: Risk Management Index of Colombia adjusted on the basis of capacities. The index which is illustrated on the right image combines the capacity of territorial subdivisions to manage the risk (image on the left), and their risk of a disaster (image in the center). Image taken from [?]

but efficiently delivered to communities, as a mean of protecting citizenś rights, foster economic growth, and make government officials accountable.

As per its current state, official risk management information lacks a delivery system that enables local communities to improve their risk awareness and disaster coping capabilities in different scenarios marked for global phenomena such as extreme temperatures and changing weather patterns. For instance, it is not apparent how similar events have impacted communities with different risk and vulnerability profiles, and there is no relevant information to assess the performance of risk management activities.

2.1 Question

Based on the previously presented information different questions have arisen.

1. In 2012, the National System of Risk Management was created in Colombia. Based on the available datasets is it possible to analyze and find patterns that show (**V3**),

   - how does the risk map of Colombia changed after the creation of this system?

2. There is a disproportionate impact of similar events among Colombia's municipalities, given by disparities in available infrastructure and first response resources (**V2**).

3. Is it possible to analyze a specific event (disaster) and show how does the same event affects different zones of the country? Based on that, we can analyze (**V1**):

   - Are there factors that make some zones more vulnerable than others?
   - How does the specific infrastructure affects the impact of the specific event?

3 Datasets sourced

The main dataset used in the project is from the Colombia Risk of Disaster Management Unit (Unidad de Gestión de Riesgos y Desastres) UNGRD [?]. The dataset contains information about

the risk management associated with natural phenomena, socio-natural, technologic, and human-based non-intentional incidents reported in Colombia in the last 10 years (38626 records). Some of the fields found in the dataset are: Date, Department, Municipality, Event Name, Code, Dead, Wounded, Disappeared, Affected People, Affected Families, Affected Houses, among others.

The team will also use a dataset from the National Administrative Department for Statistics DANE. It is a time series between 1985 to 2020 and contains information, per department code about [**?**] .

Both datasets contain "DIVIPOLA" codes, which is the codification of the Politica-Administrative Division of Colombia (Codification of the departments,). Figure 2 describes the meaning of the code. The first two numbers correspond to the department, followed by the Municipality Code and the Populated Center [[**?**]].
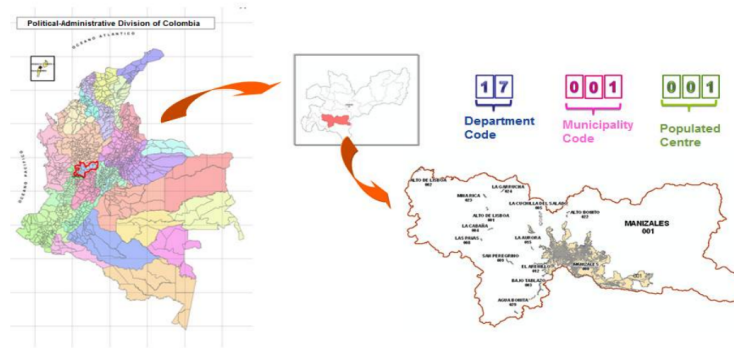
Figure 2: Explanation of "DIVIPOLA" code. The codes provide information of the Politica-Administrative Division of Colombia. Image taken from [**?**]

4 Application Overview

4.1 Users

A non-technical user should be able to use this to get meaningful outputs and visualizations. The target audience is a government official or someone in the private sector - it should be clear to them how to get value out of this and what that value is.

4.2 Architecture

Figure 3 shows the architecture of the proposed solution including the elements of the application at component level and its connections at high level (see deployment diagram). Additionally, it shows the application elements used for the Front and Back End. The figure also shows the names of the technologies used hosted on AWS cloud, i.e.: (Python, dash and libraries).

The following is the list AWS components used in the project:

1. The machine who host the solution (Elastic Compute Cloud - EC2).

2. The Database (Relational Database Service -RDS).

3. The storage for the datasets and GeoJson files for Colombia on the service (Simple Storage Service - S3) to save these files.

4. The Security group for these services talks with each other and have access from the internet as well.

5. The remote DNS (Domain Name Service) to have a friendly URL for the application on the Apache Web server.

6. A remote code repository (hosted by github). It is used for hosting the source code and documentation.

4.3 Front End Design

The following is the link of the web-page created for the project.

- http://ds4a-colombia-group03.tk/

Figure 4 shows the muckup of the project. In general terms our information system will provide a dynamic map of Colombia delivering:

1. The impact metric (or metrics) at the municipal level for a given category of events. Ability to display complementary metrics of interest for specific locations (utilities, healthcare facilities, first-responder facilities, etc).

2. Considering the established association between extreme temperatures and the frequency of hydro-meteorological events, a projected extreme-temperature indicator for the 100 most vulnerable municipalities with 3 data points: Indicator value at Time 0 (1998), Time 1 (2018) and Time 3 (projected 2040)

3. The indicator corresponds to the extreme temperature projection made by Climate Impact Lab for the number of days a year that register temperatures above 32 degrees Celsius.

4. The timeline slider will facilitate the visualization of indexes using time intervals.

5. The map (Main View) will show the index adjusted by capacity (ndice Ajustado por Capacidades) or other indexes previously defined.

6. The selection of a political division (departamento) in the main view restarts the map and will show the political view (departamento). In the latter the user can select the small political division (municipio) for update the graphics on the right side of the dashboard screen.

7. The graphics on the right side of the dashboard shows absolute values.

8. By default, the map starts with the country view with political division (departamento)

The production version of dashboard will show the following impact variables:

1. Deaths by 1000 inhabitants

2. People affected by 1000 inhabitants

3. Houses affected by 1000 inhabitants

- Outline of who the primary users would be and how they would interact with the app

- Descriptions of app features

- The datasets your team is planning on using

- How you plan on using each of these datasets/how they contribute to your app

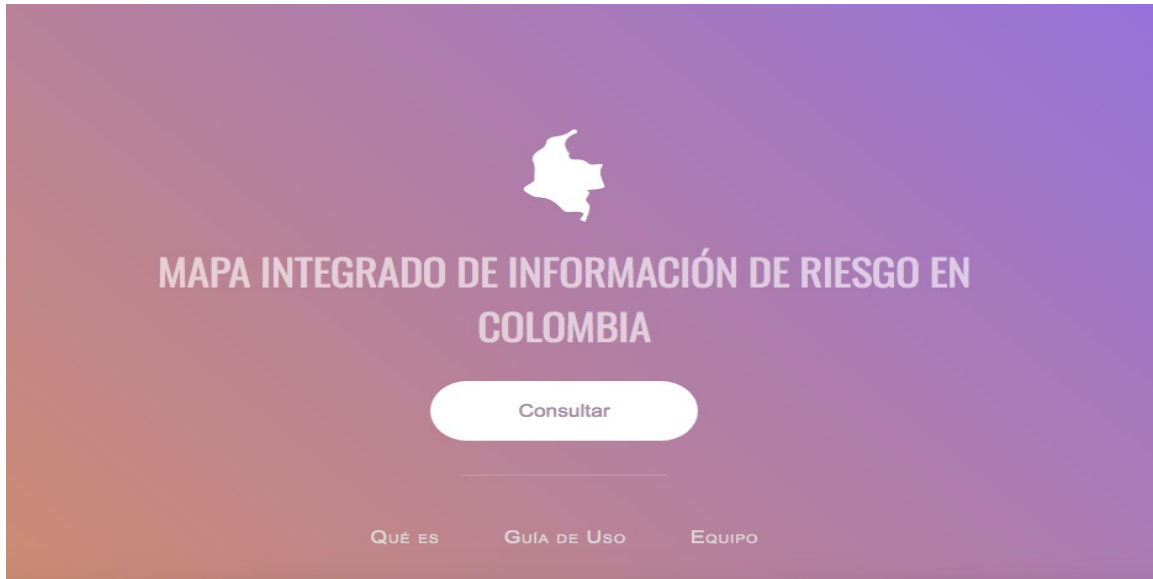**DS4 FINAL PROJECT : GROUP 03 - DEPLOYMENT DIAGRAM**
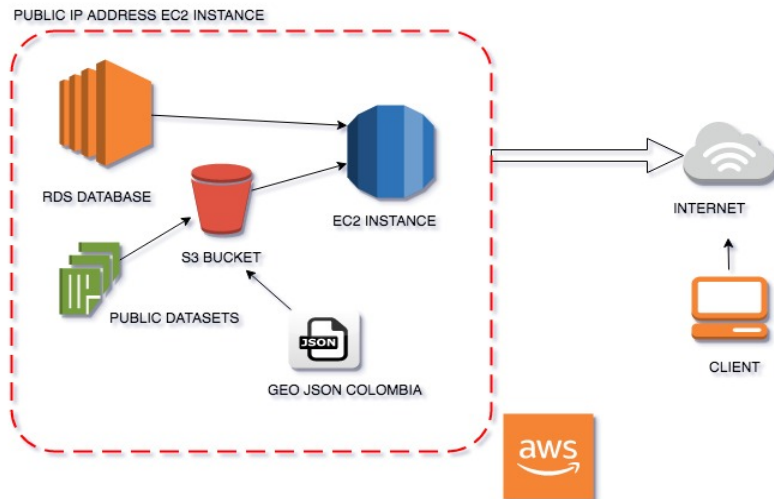
(a) Deployment diagram

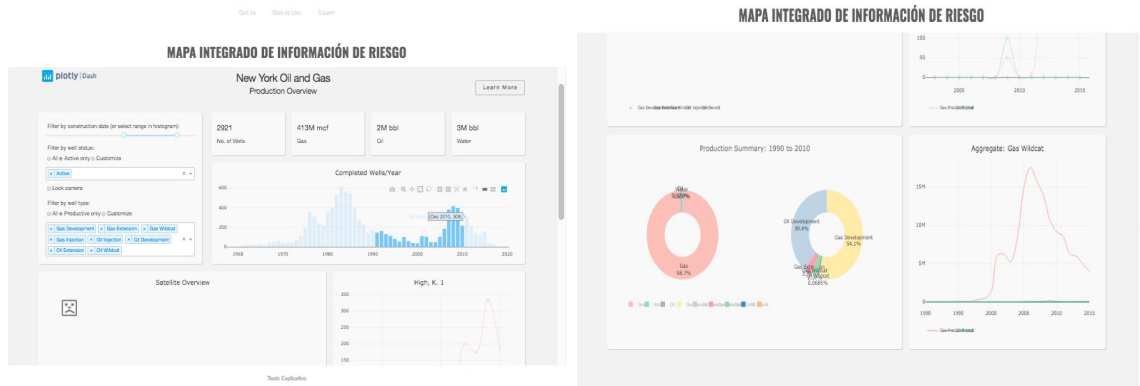**DS4 FINAL PROJECT : GROUP 03 - COMPONENT DIAGRAM**
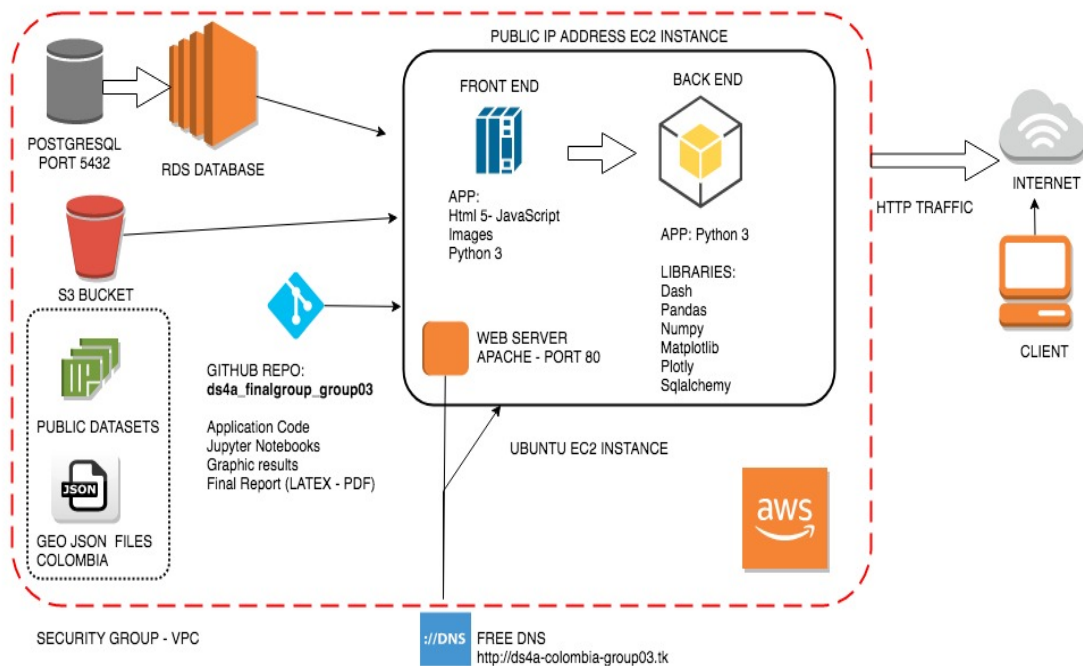
(b) Component diagram

Figure 3: Project architecture.

(a)

(b)

(c)

Figure 4: Muckup of the Hydro-meteorological Risks In Colombia project. The main page (Figure 4(a)) will show the map of Colombia with the Risk Index, and will offer interactive options to get detail information of selected Municipalities. The page will provide the current risk index in addition to a line plot that shows a projected temperature indicator.

5 Data Wrangling and Data Cleaning

The main dataset used in the project contains information about Disasters in Colombia from 1998 to 2017. We found 8 files to integrate from the following links:

– Datos Abiertos Colombia:
  https://www.datos.gov.co/

– Unidad Nacional para la Gestin del Riesgo de Desastres:
  http://portal.gestiondelriesgo.gov.co/http://portal.gestiondelriesgo.gov.co/.

The files are Microsoft Excel Files and contain the same information in the first 40 columns. However, the column names are not equal.

All the information has been analyzed in notebooks, the notebooks can be found in the following links:

– Disasters dataset integration: link

– Disaster exploratory analysis: link.

For Data Wrangling and cleaning the following steps were conducted:

– Load the 8 Excel Files as Pandas Data Frames. df_desastres_1998_2012, df_desastres_2013, df_desastres_2014, df_desastres_2015, df_desastres_2016, df_desastres_2017, df_desastres_2018

– Parsing the date columns as pandas.datetime.

  * Two corrections were conducted with some missing values:
    31/11/2018 to 30/11/2018, 27/11/201 to 27/11/2018

– Columns names standarization:

  * Change column names in df_desastres_2018.
  * Normalize type of disaster: we found the same kind of disaster written in different ways

– Concatenate all dataSets.

– Transform Data type of the following columns to numerical values:

  * muertos (deaths)
  * personas (Population)
  * heridos (Injured)
  * familias (Families)
  * viviendasdestru (Affected houses)
  * ptepeat
  * hectareas (hectares)
  * otros (others)

6 Technical Information

6.1 AWS-hosted database

Figure 5 shows the entity relationship model of the project. The information contained in the model is:

1. Disasters : information of disasters of Colombia from 1998 to 2017, with the following fields: date, Colombian political division, number of disasters, death, injuries, missing people, families, houses, public and education services infrastructures and some economical information. (table disasters)

2. Events: name and category of the event (table eventos)

3. Political division of Colombia (divipola): ID of division, subdivision and name of both (divipola table)

4. Population estimates: relates population by period, age groups, political division, id, and gender

5. Weather estimates: weather information from 1990 to 2018 (minimal temperatures, maximal temperatures, precipitations) (tables: historico_cond_metereologicas, load_mintemp, load_maxtemp, load_precipitaciones)

6. Load tables: temporal tables for loading disasters, political division, population (tables: load_disasters, load_divipola, load_populations)

7. Views: wv_disasters (summary table for disasters).

The SQL script for the creation of the database on AWS can be download from this link:
https://marioceron-case-51.s3.amazonaws.com/final_project/Script_Desastres_DB.sql

Full backup Database:
https://marioceron-case-51.s3.amazonaws.com/final_project/Script_Desastres_DB_full.sql

Figure 6 shows the database uploaded to AWS.

Figure 7 shows the database loaded to the AWS hosted database.

The link between the front-end and the AWS-hosted databased was also stablished (see Figure 8).

Figure 5: Entity relationship model.

Figure 6: AWS connection.

Figure 7: Database loaded to AWS.

Figure 8: Link between DASH and AWS stablished