



PRESTAR SERVICIOS PROFESIONALES PARA ATENDER TEMAS
TECNICOS
RELACIONADOS PARA
LA ELABORACION DE DOCUMENTOS DE LINEAMIENTOS
TÉCNICOS, DE LOS PROGRAMAS, PROYECTOS, CONTRATOS Y/O
CONVENIOS A
CARGO DEL INVIA

INVIA
INSTITUTO NACIONAL DE VÍAS

Tercer informe de avance

INSTITUTO NACIONAL DE VÍAS TERRITORIAL ANTIOQUIA PROYECTO: MAPA DE VULNERABILIDAD FAUNÍSTICA

Avance del tercer mes de trabajo configuración y desarrollo del sistema de comunicación y transferencia de datos entre los elementos del sistema de automatización del mapa de vulnerabilidad faunística

Ing. Jairo I. Coy
Equipo de DevOps

PROLOGO

Este documento toma como base la GTC 185, la M80 Guía para la presentación de informes técnicos de avance de programas y proyectos de CTel y en la ISO 10013. El formato se realiza como una herramienta para la gestión de documentación y generación de reportes mensuales del equipo de trabajo de desarrollo del Mapa de vulnerabilidad faunística del segundo periodo de 2025.



Contenido

RESUMEN	4
1. INTRODUCCIÓN	5
2. OBJETIVOS	6
2.1 Objetivo general	6
2.2 Objetivos específicos.....	6
3. ESTRUCTURA DE TRABAJO	7
3.1 Definición del alcance y objetivos	7
3.2 Análisis de requisitos	7
3.3 Diseño de arquitectura o sistema	7
4. METODOLOGÍA.....	8
4.1 Modelos matemáticos.....	9
5. DESARROLLO Y RESULTADOS.....	11
6. CONCLUSIONES.....	15
7. RECOMENDACIONES.....	16
BIBLIOGRAFÍA	17
ANEXOS	18

Listas de figuras

Figura 1. Esquema de la metodología de la fase III.....	9
Figura 2. Estructura de sistema de colas.	10
Figura 3. Arquitectura y flujo de la cadena de subprocessos.	12
Figura 4. Arquitectura y flujo de la cadena de proceso y postproceso.	12
Figura 5. Variables de configuración y acceso al bucket.....	13
Figura 6. Resultado de función de recorrido de bucket objetivo.	13
Figura 7. Resultados del proceso de extracción de información del bucket.....	14
Figura 8. Resultados del modelo MAXENT.....	14

Listas de tablas

Tabla 1. Lenguajes y motores del sistema adicionales.	8
Tabla 2. Requisitos técnicos del sistema de procesamiento.....	8

Listas de ecuación

Ecuación 1. Ley de Amdahl	10
Ecuación 2. Ecuación original de Amdahl (tiempo de ejecución mejorado).	10
Ecuación 3. Notación generalizada para indicar el tipo de sistema.	10
Ecuación 4. Función de probabilidad de distribución de Poisson.	11

Listas de anexos

Anexo 1. Estructura del bucket	18
Anexo 2. Estructura del sistema de almacenamiento de fuentes paralelas.	18
Anexo 3. Estructura de los insumos del modelo MAXENT.	18
Anexo 4. Tiempo de vida útil de las capas y tipos de descarga.....	19

RESUMEN

Durante el tercer periodo de trabajo, comprendido entre el 1 al 31 de noviembre, los avances logrados en el proyecto del sistema de automatización del Mapa de Vulnerabilidad Faunística se centraron en la configuración del sistema de preprocesamiento y procesamiento, también se inicia la configuración del sistema que enviará la información lista y procesada al sistema de posprocesamiento (visualización de resultados) ubicado en el sistema de gestión del INVIAS gobernado por ArcGIS Enterprise.

Durante esta fase, se cumple con el objetivo de realizar la comunicación entre el sistema de preprocesamiento y el sistema de procesamiento, logrando generar una cadena automatizada aplicando el flujo de datos diseñado en la segunda fase. Durante este periodo, se realiza mejoras al diseño de la estructura del contenedor alojado en la nube de Google, y se crea la función que realizará el proceso de extracción y almacenamiento en el sistema de procesamiento.



1. INTRODUCCIÓN

El sistema de automatización del mapa de vulnerabilidad faunística se encuentra formado por tres frentes de acción; preprocesamiento, procesamiento y posprocesamiento. Los tres actúan como una cadena de procesadores, que en conjunto brindan las capas que serán publicadas en el servidor central del INVIAS donde se encuentra alojado el geo portal HERMES, destino donde se publicará la información generada. Por medio del uso de API's, se generan los canales para la transferencia de información entre cada sistema.

El sistema de preprocesamiento se divide entre dos elementos. Sus procesos parten de la nube de Google y cierra en el Rack de procesamiento. Al interior del Rack se halla el sistema de extracción, el sistema de alineamiento, distribución, procesamiento y unos componentes del sistema de posprocesamiento.

Cada sistema se forma de una función de segmentación, distribución y almacenamiento, como también de funciones que transformen la información de forma que el sistema a donde se valla a enviar lo acepte sin problemas o restricciones. Debido a que se trabaja con la nube de Google, se aplican una de procesos que permitan el tráfico libre de información entre los sistemas de forma segura.

El documento muestra los avances que se generaron durante la tercera fase del proyecto del sistema de automatización del mapa de vulnerabilidad faunística.

2. OBJETIVOS

2.1 Objetivo general

Prestar servicios profesionales especializados para abordar aspectos técnicos relacionados con la elaboración y desarrollo del sistema digital, generación de documentación y capacitación del equipo de trabajo para el procesamiento y aplicación de modelos de análisis de niveles de riesgo en las carreteras nacionales de Colombia.

2.2 Objetivos específicos

- Crear, desarrollar y configurar los elementos de preprocesamiento.
- Crear, desarrollar y configurar los elementos de procesamiento.
- Iniciar los procesos de testing en servidor local del sistema de preprocesamiento-procesamiento.
- Capacitar al equipo de trabajo en el uso y activación del sistema de preprocesamiento y el sistema de procesamiento.
- Generar un avance sobre el manual de configuración del sistema de preprocesamiento.

INVIAS
INSTITUTO NACIONAL DE VIAS

3. ESTRUCTURA DE TRABAJO

Para generar la cadena de procesos entre sistemas, se emplean e integran una serie de herramientas como también se desarrollan funciones que permiten el libre tráfico de información de forma segura. Es necesario recalcar que la transferencia de información solo es posible entre los sistemas matriculados, brindando capas de seguridad a los sistemas, reduciendo la posibilidad de intrusos o datos cruzados no deseados.

3.1 Definición del alcance y objetivos

Durante esta fase, se define una ruta de trabajo dividida en tres etapas. La primera etapa se centra en estructurar el contenedor en la nube (bucket) y configurar el proyecto y/o entorno de trabajo, la segunda etapa se enfoca en desarrollar el sistema de extracción y distribución alojado en el RACK, la tercera etapa cierra con la activación del sistema verificando que la información se almacene de y distribuya de forma apropiada. Esta información actuará como insumo para continuar con el proceso de alineación de las capas.

3.2 Análisis de requisitos

Para cumplir con este propósito se hace uso de los servicios que brinda la nube de Google y su respectiva configuración. Para los procesos de extracción y segmentación se identifican los requisitos técnicos y los componentes para la ejecución de las actividades. Para las actividades de procesamiento se diseñan y desarrollan las funciones que cumplen este propósito y se inicia con las rutinas de ajuste del modelo central.

3.3 Diseño de arquitectura o sistema

La cadena de procesos se realiza de una forma anidada. Durante la etapa anterior, se propuso una estructura de procesos en cada elemento (procesos independientes), no obstante, bajo una serie observaciones, se realiza hacia una la modificación, donde las responsabilidades son compartidas entre los sistemas. De esta forma se tiene preprocesamiento en la nube y preprocesamiento en el servidor de procesamiento (RACK). Este diseño de arquitectura favorece en cuanto a recursos y tiempo de respuesta. Bajo esta estructura se integran los siguientes lenguajes y motores de trabajo, adicionales a los mencionados en la segunda fase.

Tabla 1. Lenguajes y motores del sistema adicionales.

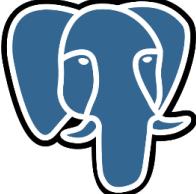
Lenguaje	Logo	Descripción
Bash		Lenguaje de programación usado para configurar los permisos de acceso de forma remota y remover restricciones de tráfico de información entre los sistemas.
PostgreSQL		Motor de base de datos usado para la gestión de datos e información geográfica.

Tabla 2. Requisitos técnicos del sistema de procesamiento.

Sistema	Logo	Función
R+		Entorno que se encarga de realizar los cálculos dentro del modelo MAXENT.
Java		Entorno de trabajo para el desarrollo de soluciones con Java.
Django		Entorno de trabajo para el desarrollo de soluciones con Python.

4. METODOLOGÍA

La metodología aplicada en esta fase se forma de nueve etapas. La primera etapa se enfoca en la creación del proyecto, como la creación del contenedor de almacenamiento (bucket), el contenedor tiene una configuración y distribución ideado con el objetivo de establecer una cadena de procesos automatizados, por esto se plantea un estándar al interior (ver Anexo 1). La segunda etapa consolida todo el proceso de comunicación entre el bucket y el servidor de procesamiento. La tercera etapa corresponde a todo el proceso de configuración y desarrollo de los elementos requeridos para ejecutar el tráfico de información entre los sistemas. En

esta etapa se hace el desarrollo de funciones y estructuras (scripts) que permitan integrar los permisos, levantar restricciones, como también la activación de la llave de acceso para ejecución de procesos remotos. También en esta etapa se realiza la instalación de requisitos técnicos (librerías), no obstante, primero se verifica que no se cuente con ellos. Este proceso se ejecuta tanto en el contenedor en la nube como en el servidor de procesamiento. La cuarta y quinta etapa se centra en realizar todo el proceso de extracción y almacenamiento de información en el entorno de trabajo, también se verifica que la distribución durante el almacenamiento sea al diseñada. En la sexta y séptima etapa se realiza el proceso de alineación de las capas y la concatenación de estas formando una sola, generando uno de los insumos del modelo de procesamiento. En la octava etapa, se realiza el cálculo de los índices y se separan las bandas en cada capa concatenada. Finalmente, en novena etapa se lleva al modelo para la realización de pruebas. Todo este proceso inicialmente se hace forma manual, pero el sistema terminado, ejecuta toda la cadena de forma automatizada.

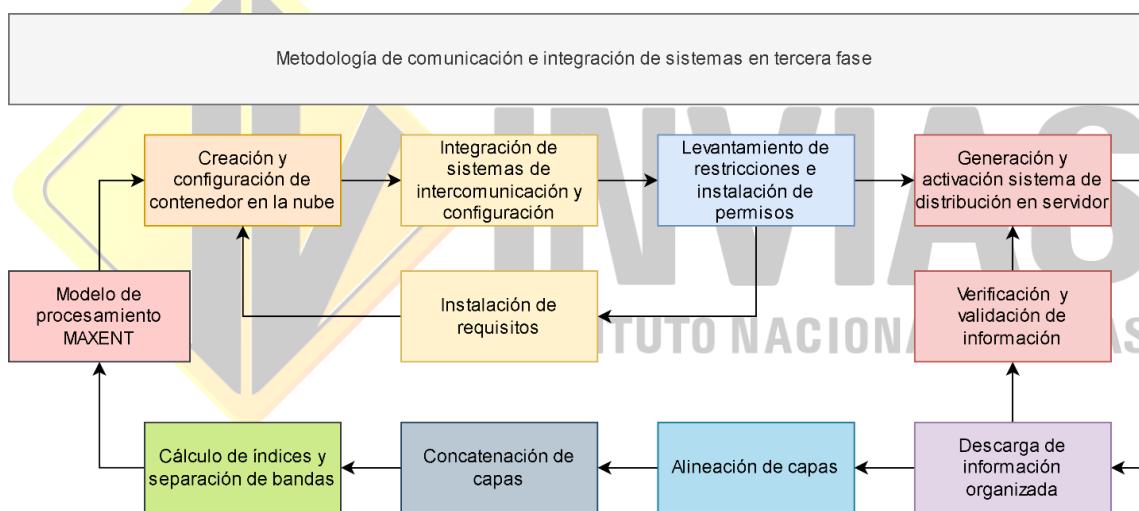


Figura 1. Esquema de la metodología de la fase III.

4.1 Modelos matemáticos

Debido al volumen de datos que se requiere descargar para cada ejecutar cada proceso con MAXENT, se aplica la ley de Amdahl, esto con el objetivo de favorecer en cuanto a tiempos de descarga y consumo de recursos (sistema ágil). La ley como tal estima el valor de la máxima aceleración de un programa al igual que el número de hilos ideal para un proceso.

$$A_{ccp} = \frac{1}{(1 - P) + \frac{P}{N}}$$

Ecuación 1. Ley de Amdahl

Donde P, es la proporción del programa que puede ser paralelizada, N es el número de procesadores usados y (1-P) es la parte secuencial del programa que debe ejecutarse de forma serial. Para el cálculo del tiempo optimizado (ganancia temporal), se hace uso de la ecuación original de la ley de Amdahl.

$$T_m = T_a \cdot \left((1 - F_m) + \frac{F_m}{A_m} \right)$$

Ecuación 2. Ecuación original de Amdahl (tiempo de ejecución mejorado).

Donde, Fm es la fracción de tiempo que el sistema utiliza el subsistema mejorado, Am el factor de mejora que se ha introducido en el subsistema mejorado, Ta tiempo de ejecución antigua y Tm tiempo de ejecución mejorado. En situaciones donde se genere un apilamiento de tareas, se aplica la teoría de colas [1].

$A/B/c$

Ecuación 3. Notación generalizada para indicar el tipo de sistema.

Siendo, A la distribución de probabilidad que siguen las llegadas al sistema. B la distribución de probabilidad que sigue el tiempo de servicio y c la cantidad de servidores con los que cuenta el sistema [2].

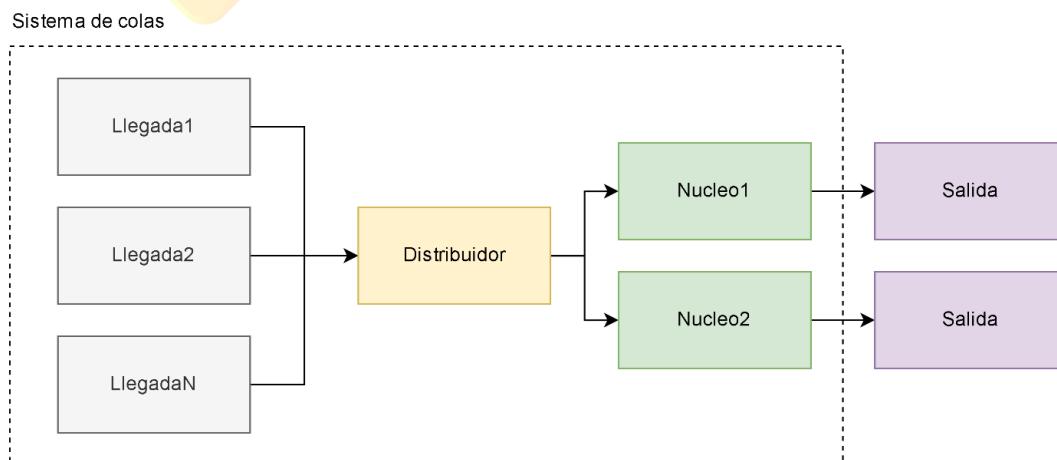


Figura 2. Estructura de sistema de colas.

El elemento de distribución de llegadas se activa al momento en que el servidor se encuentra ocupado y su distribución se fundamenta a partir de un proceso aleatorio de la llegada de tareas al igual que de las variaciones en el tiempo en registro de servicios, para su estimación se calcula la distribución de probabilidad exponencial.

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Ecuación 4. Función de probabilidad de distribución de Poisson.

Donde, Pn es la probabilidad de que n procesos estén en el sistema en el tiempo t, λ el número medio de eventos en el intervalo [3] [4].

5. DESARROLLO Y RESULTADOS

Los resultados logrados durante la fase III corresponden al desarrollo de funciones soportado bajo los modelos de distribución y procesamiento. Así mismo, se desarrollan los insumos suficientes para el descargue de información y el tráfico libre y seguro de datos entre la nube de Google y el servidor de procesamiento donde se aloja el modelo MAXENT. Con el objetivo de optimizar los procesos y lograr dar un mejor uso a los recursos físicos con los que se cuentan, se fusiona la cadena preproceso entre la nube y servidor de procesamiento. La distribución de tareas inicia en Google Earth Engine (GEE), donde sea adquiere y aadecua la parte de las capas que se requieren, esta información se almacena en un bucket del módulo de Google Cloud Storage (GCS), desde el este, se extrae la información para ser almacenada en el servidor de procesamiento empleando la API de Google. De forma paralela se realiza el descargue de las otras capas a partir de la información recolectada de otras fuentes paralelas (IGAC, IDEAM, SIAC, NASA, etc) y de igual forma se almacenan. Con a la información almacenada, el sistema activa la función de alineación y continua con la concatenación de las capas preprocesadas. Con estos procesados completos, se calculan los índices y se realiza la separación de bandas.

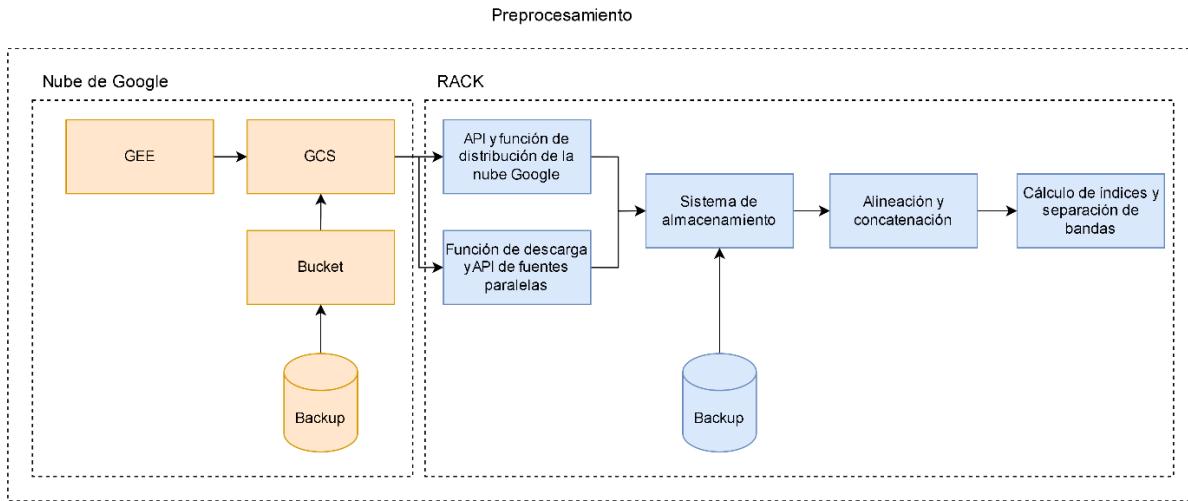


Figura 3. Arquitectura y flujo de la cadena de subprocessos.

Con la información ya preprocesada y lista para el modelo MAXENT. Este proceso se realiza de forma paralela, aplicando la Ecuación 1, para estimar el número de hilos y lanzar el proceso. Para estimar el tiempo de ganancia se hace uso de Ecuación 2, con el propósito de aumentarlo en referencia a la configuración usada.

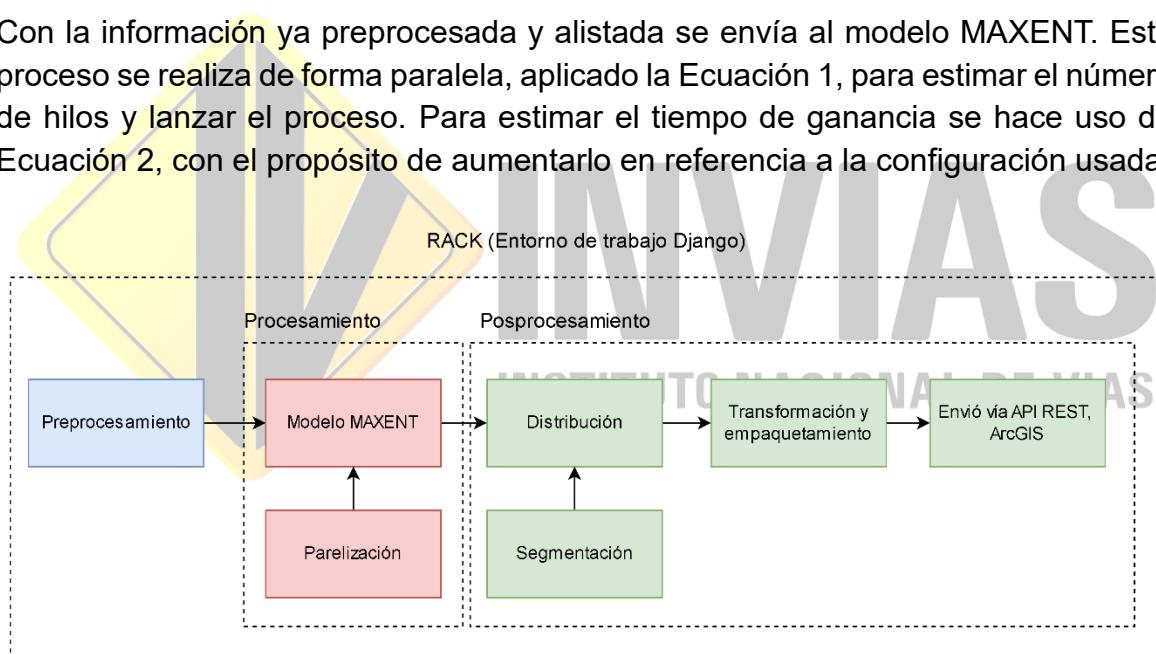


Figura 4. Arquitectura y flujo de la cadena de proceso y postproceso.

Finalizado el tiempo de proceso, el modelo genera las capas **resultado en bruto**. En esta etapa, la información resultante no se encuentra en formato para que el sistema de visualización (geo portal), acepte la información y logre renderizarla en el sistema HERMES. Para este propósito, se envía la información a una función de transformación y empaquetamiento. Con la información en condiciones para ser publicada se activa el sistema de envío a la nube del cliente.

El sistema requiere de unas variables de configuración y permisos de acceso. Parte de estas variables se ejecutan desde el entorno de trabajo en Django.

```
# Default primary key field type
# https://docs.djangoproject.com/en/5.2/ref/settings/#default-auto-field

DEFAULT_AUTO_FIELD = 'django.db.models.BigAutoField' # Permiso para almacenamiento de datos de gran carga
GS_CREDENTIALS = os.environ.get('GOOGLE_APPLICATION_CREDENTIALS') # Accesos a la llave privada

DEFAULT_FILE_STORAGE = 'storages.backends.gcloud.GoogleCloudStorage' # Acceso al Google Cloud
GS_BUCKET_NAME = 'invias' # Nombre del bucket objetivo
GS_PROJECT_ID = 'complete-energy-448804-i2' # Identificador del bucket objetivo
```

Figura 5. Variables de configuración y acceso al bucket.

Como estrategia para probar el sistema de comunicación entre el bucket y el entorno de trabajo con Django, se trabaja con un bucket demo. Se desarrolla y ejecuta una función que inicialmente recorra todo el bucket ejecutándose cada periodo deseado, con el objetivo de identificar cambios realizados a la información almacenada.

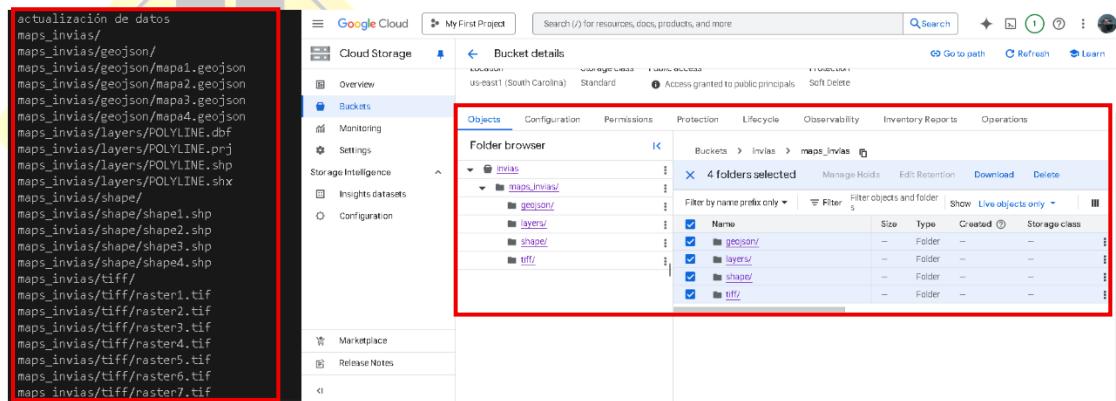


Figura 6. Resultado de función de recorrido de bucket objetivo.

Tras realizar la debida verificación de información, se activa la descarga de toda la información del bucket en el sistema de trabajo en Django, posterior se activa nuevamente el entorno, favoreciendo en cuanto a consumo de recursos y eficiencia en los procesos. Para el descargue de información se trabajan con varios hilos y estos se estiman a partir de la Ecuación 1, haciendo despeje de la variable N. Para lograrlo, se estima una serie de tiempos deseados posibles calculados usando la Ecuación 2.

Figura 7. Resultados del proceso de extracción de información del bucket.

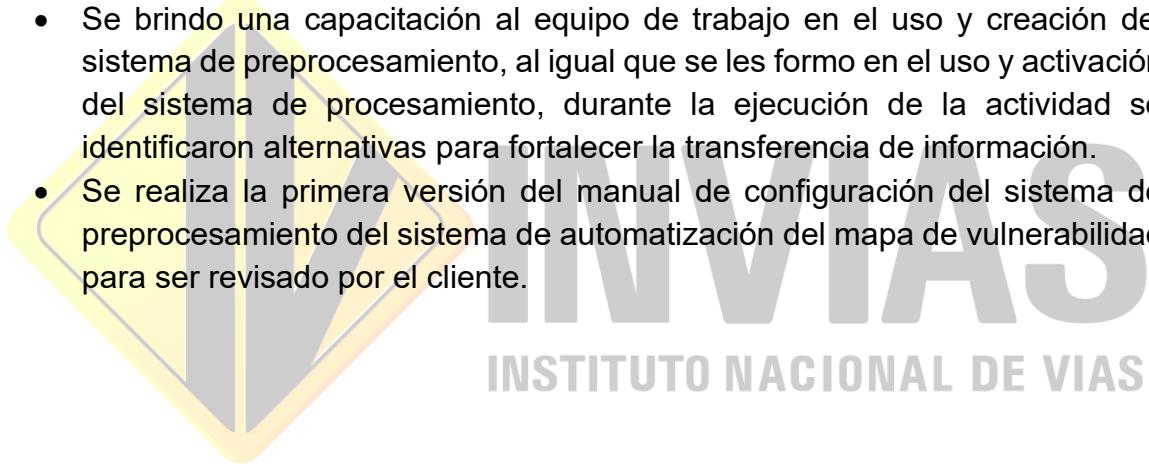
Para ejecutar el modelo MAXENT se crean los directorios con la información correspondiente a cada región geográfica y se almacena la información preprocesada en cada directorio. Los directorios tienen una estructura definida como un estándar en el sistema y esta se evidencia en el Anexo 3. El sistema presenta tres distintos períodos de actualización, estos se pueden identificar en el Anexo 4.

Dentro de todos los resultados entregados por el modelo MAXENT, el archivo “resultado_maxent.tif”, es el que se procesara para ser finalmente visualizado por en el geo portal del INVIA. El tiempo empleado por el MAXENT para realizar el procesamiento de la región de prueba tarda un aproximado de 48 minutos.

Figura 8. Resultados del modelo MAXENT.

6. CONCLUSIONES

- Se realiza de forma exitosa la actividad de creación de un contenedor de prueba en la nube Google, almacenando una serie de información de diferente tipo y extensión, no obstante, se identifican una serie de limitaciones que se pueden mostrar como oportunidades de mejora en una versión actualizada del sistema de almacenamiento.
- Se realiza de forma exitosa la activación y pruebas de funcionalidad del sistema de procesamiento, se identifican elementos adicionales inicialmente no identificados que permitieron el buen funcionamiento del sistema, por lo que se realizaron modificaciones al diseño y arquitectura inicialmente planteada, así mismo se identifican opciones de mejora para una versión futura.
- Se brindo una capacitación al equipo de trabajo en el uso y creación del sistema de preprocesamiento, al igual que se les formó en el uso y activación del sistema de procesamiento, durante la ejecución de la actividad se identificaron alternativas para fortalecer la transferencia de información.
- Se realiza la primera versión del manual de configuración del sistema de preprocesamiento del sistema de automatización del mapa de vulnerabilidad para ser revisado por el cliente.



INVIAS
INSTITUTO NACIONAL DE VIAS

7. RECOMENDACIONES

- Se recomienda iniciar con las pruebas en el servidor de producción en aras de ir reconociendo posibles requisitos carentes, o problemáticas técnicas no contempladas.
- Se recomienda hacer un refuerzo en la capacitación de la activación del sistema de procesamiento, así mismo se recomienda que el equipo por parte del cliente participe en los procesos de capacitación técnica en el uso de los sistemas en desarrollo.
- Se recomienda aplicar el método de Amdahl a los demás elementos de la cadena de procesamiento, con el propósito de optimizar los tiempos de respuesta y favorecer en cuanto al uso de recursos tecnológicos.

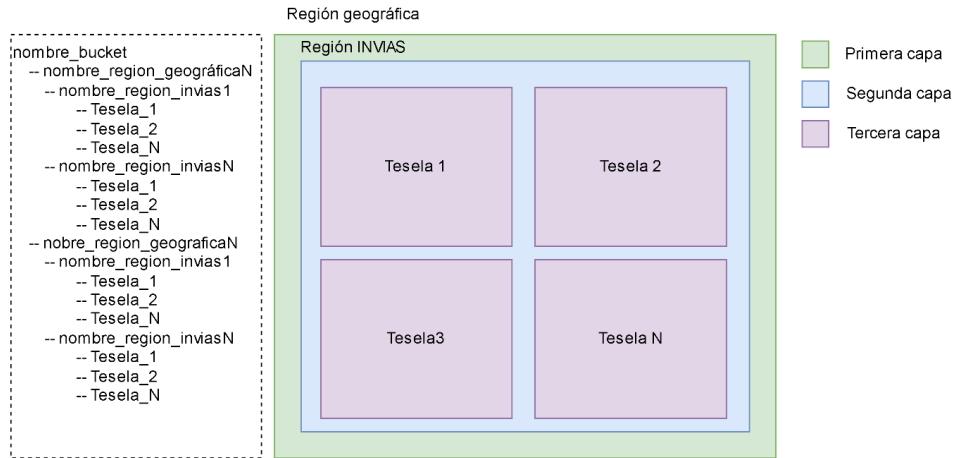


BIBLIOGRAFÍA

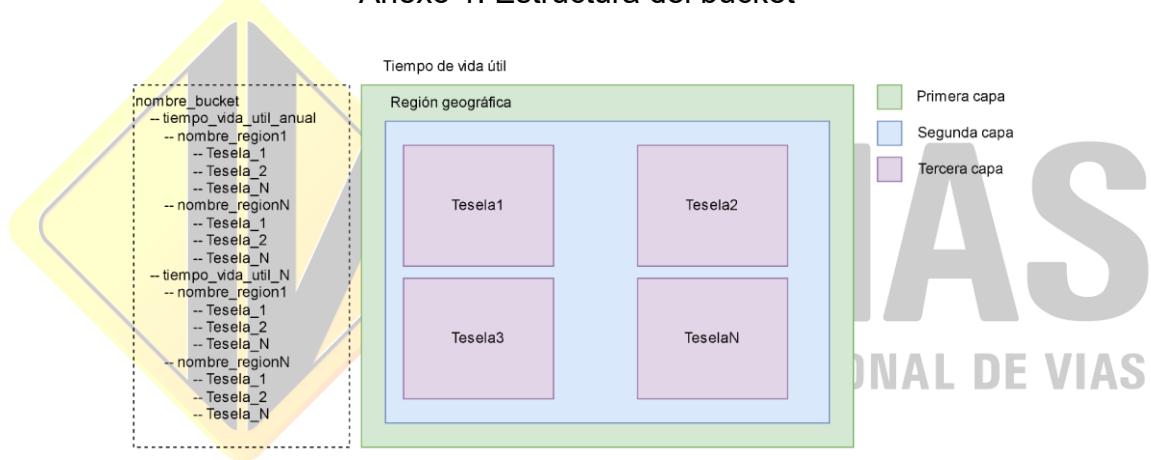
- [1] L. M. Portilla, L. Montoya Arias y S. A. Fernández Henao, «Análisis de líneas de espera a través de teoría de colas y simulación,» *Scientia et Technica*, vol. 17, nº 46, pp. 56-61, 2010.
- [2] V. M. Burbano Pantoja, M. A. Valdivieso Miranda y Á. S. Burbano Valdivieso, «Teoría de colas en la práctica investigativa: Generación de modelos probabilísticos para líneas de espera,» *Revista de investigación Desarrollo e Innovación*, vol. 14, nº 2, pp. 9-24, 2024.
- [3] Universidad Autónoma de Madrid, «Estadística,» [En línea]. Available: <https://www.estadistica.net/INVESTIGACION/TEORIA-COLAS.pdf>. [Último acceso: 27 11 2025].
- [4] Universidad de Sonora, «Introducción a la teoría de colas y su simulación,» Universidad de Sonora, Hermosillo Sonora, 2013.



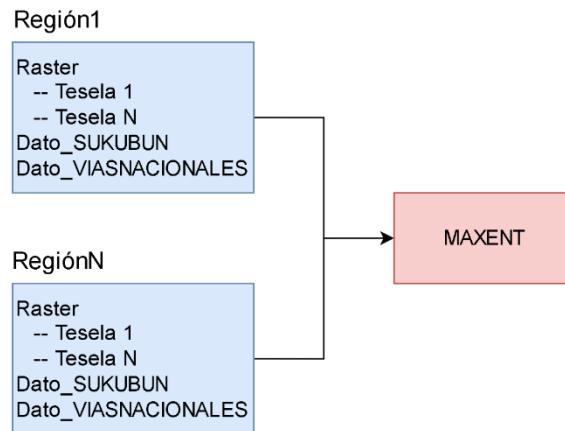
ANEXOS



Anexo 1. Estructura del bucket



Anexo 2. Estructura del sistema de almacenamiento de fuentes paralelas.



Anexo 3. Estructura de los insumos del modelo MAXENT.

Tiempo de vida útil de las capas y tipo de descarga			
Tiempo		Tipo	
Bianual	3	Automática	24
Anual	2	Semiautomática	3
Trimestral	23	Manual	3
Semestral	1		
Estático	1		

Anexo 4. Tiempo de vida útil de las capas y tipos de descarga.

