

Desarrollo de una herramienta software para la gestión de datos estadísticos meteorológicos y de hidrología de las principales plantas de energía y afluentes de Colombia

Leonardo Hernando Dallos Martínez

Trabajo de grado presentado para optar por el título de Ingeniero de Sistemas

Director

Fernando Antonio Rojas Morales

Universidad Industrial De Santander

Facultad De Ingenierías Fisicomecánicas

Escuela de Ingeniería de Sistemas e Informática

Bucaramanga

2023

Agradecimientos

Quiero expresar mi mas profundo agradecimiento a todas las personas que contribuyeron de manera significativa en la culminación de este proyecto de grado. Fue un camino desafiante pero increíblemente enriquecedor, el cual no habría sido posible sin el apoyo constante y aliento de muchos.

En primer lugar quiero expresar mi gratitud a la Universidad Industrial de Santander, por brindarme la oportunidad de adquirir conocimientos, explorar nuevas ideas y desarrollar habilidades a lo largo de mi carrera académica. El respaldo de los profesores con los que tuve la oportunidad de compartir, quienes me compartieron su experiencia y orientación, que en sumatoria fue esencial para el éxito de este proyecto.

Agradecer tambien a mi familia, pues su confianza y apoyo incondicional, tanto económico como moral fue mi mayor fuente de fortaleza.

A mis amigos y compañeros de carrera, gracias por el intercambio de ideas y el sentido de comunidad que logramos apoyarnos y salir adelante juntos.

La culminación de este proyecto de grado marca el fin de una etapa significativa para mi, pero tambien el comienzo de nuevos desafíos y oportunidades. Llevo conmigo el aprendizaje y el apoyo recibido en este trayecto.

Gracias a todos de corazón.

Tabla De Contenido

Introducción	12
1. Objetivos	14
1.1. Objetivo General	14
1.2. Objetivos Específicos	14
2. Marco de referencia	15
2.1. Antecedentes y descripción del problema.....	15
2.2. Fundamentos teóricos.....	15
2.2.1. Hidrología.....	15
2.2.2. Principales Plantas de energía Hidroeléctrica de Colombia.....	16
2.2.3. Meteorología	19
2.2.4. Herramienta software	20
2.2.5. Bases de datos relacionales	21
2.2.6. Consultas a APIs	25
2.3. Fuentes de información de acceso público para datos hidrológicos	27
2.3.1. XM	27
2.3.2. UPME.....	28
2.4. Tecnologías involucradas en el desarrollo del proyecto	29
2.4.1. Streamlit	29
2.4.2. Herramientas de automatización	30
3. Metodología de desarrollo del proyecto	30

4.	Definición de requerimientos.....	32
4.1.	Requerimientos funcionales	32
4.2.	Requerimientos no funcionales	33
5.	Diseño	33
5.1.	Definición de las métricas a tratar.....	33
5.1.1.	Descripción de las métricas a tratar.....	34
5.2.	Selección de las fuentes de información	35
5.3.	Diseño de la base de datos.....	36
5.3.1.	Diagrama entidad relación.	39
5.3.2.	Entidades	40
5.3.3.	Carga de datos	45
5.3.3.1.	Descarga a partir del llamado a la API de XM:.....	48
5.3.3.2.	Descarga a partir del enlace directo ofrecido en la página web de XM:	54
5.3.3.3.	Diagrama de secuencia de extracción de datos.	55
5.3.4.	Automatización de descarga de información a la base de datos desde las fuentes de información seleccionadas.	57
5.4.	Diseño de la herramienta software	58
5.4.1.	Casos de Uso	59
5.4.2.	Diagrama de actividades	64
5.5.	Diagrama de arquitectura	66
6.	Implementación.....	66
6.1.	Diseño de Interfaces	68

7. Pruebas funcionales de validación	74
8. Conclusiones	81
9. Recomendaciones	82
Referencias Bibliográficas	83

Lista de figuras

Figura 1 <i>Represa, vertedero de borde</i>	17
Figura 2 <i>Principales embalses de Colombia</i>	19
Figura 3 <i>Estructura tabla base de datos relacional</i>	21
Figura 4 <i>Ejemplo diagrama E-R</i>	23
Figura 5 <i>Ejemplo del modelo relacional</i>	24
Figura 6 <i>Diagrama Entidad-relación del modelo relacional de la base de datos</i>	40
Figura 7 <i>Funcionamiento API XM</i>	50
Figura 8 <i>Ejemplo de una métrica con múltiple indicador</i>	52
Figura 9 <i>Ejemplo configuración de una consulta API en “Postman”</i>	53
Figura 10 <i>Ejemplo Formato de salida respuesta API</i>	54
Figura 11 <i>Diagrama extracción de datos por parte de la implementación que descarga las métricas automáticamente</i>	56
Figura 12. <i>Diagrama de extracción de datos por parte del usuario</i>	57
Figura 13 <i>Casos de Uso</i>	59
Figura 14. <i>Diagrama de actividades para la generación de reportes estadístico de acuerdo con una métrica seleccionada</i>	65
Figura 15 <i>Diagrama de arquitectura - extracción de datos, generación y visualización de reportes</i>	66

Figura 16 <i>Interfaz 1 - inicio (Pantalla de bienvenida) de la herramienta software</i>	69
Figura 17 <i>Interfaz 2 - selección de reporte estadístico a visualizar.</i>	70
Figura 18 <i>Interfaz 3 - selección de fecha a generar reporte.</i>	71
Figura 19 <i>Interfaz 4 – Visualización de reporte</i>	72
Figura 20 <i>Interfaz 5 - Visualización de tabla de datos del grafico del reporte y enlace a descarga de estos</i>	73
Figura 21 <i>Interfaz 6 – Visualización de la selección de filtros para aplicar.</i>	74

Lista de tablas

Tabla 1	<i>Centrales hidroeléctricas en Colombia clasificadas por regiones</i>	17
Tabla 2	<i>Métricas seleccionadas a descargar</i>	34
Tabla 3	<i>Fuentes de información de las métricas seleccionadas</i>	36
Tabla 4	<i>Entidad base de datos - "dim_recurso"</i>	41
Tabla 5	<i>Entidad base de datos - "dim_rio"</i>	41
Tabla 6	<i>Entidad base de datos - "dim_embalse "</i>	42
Tabla 7	<i>Entidad base de datos - "dim_region_hidro"</i>	42
Tabla 8	<i>Entidad base de datos - "dim_sistema"</i>	43
Tabla 9	<i>Entidad base de datos - "met_rio"</i>	43
Tabla 10	<i>Entidad base de datos - "met_sistema"</i>	43
Tabla 11	<i>Entidad base de datos - "met_embalse"</i>	44
Tabla 12	<i>Entidad base de datos - "met_recurso"</i>	44
Tabla 13	<i>Entidad base de datos - "dim_parametro"</i>	46
Tabla 14	<i>Caso de uso 1 – Ingreso al sistema</i>	59
Tabla 15	<i>Caso de uso 2 – Selección de reportes a visualizar</i>	60
Tabla 16	<i>Caso de uso 3 - Selección Fecha de reporte</i>	61
Tabla 17	<i>Caso de uso 4 - Selección de conjunto de datos para el reporte</i>	62
Tabla 18	<i>Caso de uso 5 – Descargar tablas de los datos de los reportes</i>	63

Tabla 19 <i>Especificación secciones interfaz 1</i>	69
---	----

Tabla 20 <i>Especificación secciones interfaz 2</i>	70
---	----

Resumen

Título: Desarrollo de una herramienta software para la gestión de datos estadísticos meteorológicos y de hidrología de las principales plantas de energía y afluentes de Colombia.*

Autor: Leonardo Hernando Dallos Martínez

Palabras Clave: Bases de datos, herramienta software, API, hidrología, reportes, automatización.**

Descripción: Actualmente en Colombia el acceso a la información concerniente a las variables involucradas en la generación de Energía es de acceso público. La medición y la distribución de esta información es realizada por entidades especializadas a tal fin, las cuales comparten los datos por medio de plataformas web y por servicio de APIs. Estas entidades proporcionan información histórica y actualizada de distintas métricas relacionadas con los recursos, embalses y ríos que son partícipes en la generación de energía en el país. El acceso a esa información es algo ambiguo ya que para conseguir una base de datos completa, desde la cual se pueda realizar análisis estadístico de los patrones y comportamientos de dichas variables, es necesario consultar diversas fuentes de información con sus respectivos diversos métodos de interacción.

En este documento se presenta la descripción de una herramienta software que permite reunir periódica y automáticamente la información más relevante de las métricas que se consideran importantes para la generación de energía en Colombia, relacionadas a los principales recursos de generación, para posteriormente ser presentadas en una aplicación que permite generar reportes con gráficos estadísticos interactivos, los cuales sirven para realizar un análisis que permita en base a un intervalo de fecha definido establecer patrones de comportamiento en los datos tratados.

* Proyecto de Grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fernando Antonio Rojas Morales

Abstract

Title: Development of a software tool for the management of meteorological and hydrological statistical data of the main power plants and troughs of Colombia*

AUTHOR: Leonardo Hernando Dallos Martinez**

KEY WORDS: Databases, software tool, API, hydrology, reports, automation.

DESCRIPTION: Currently in Colombia, access to information concerning the variables involved in energy generation is publicly accessible. The measurement and distribution of this information is carried out by specialized entities for this purpose, which share the data through web platforms and API services. These entities provide historical and updated information on different metrics related to the resources, reservoirs, and rivers that participate in the generation of energy in the country. Access to this information is somewhat ambiguous since in order to obtain a complete database, from which statistical analysis of the patterns and behaviors of said variables can be carried out, it is necessary to consult various sources of information with their respective various methods of interaction.

This document presents the description of a software tool that allows to periodically and automatically gather the most relevant information of the metrics that are considered important for the generation of energy in Colombia, related to the main generation resources, to later be presented in a application that allows generating reports with interactive statistical graphs, which are used to carry out an analysis that allows, based on a defined date interval, to establish behavior patterns in the processed data.

* Degree work

** Faculty of Physical-Mechanical Engineering. School of Systems Engineering. Director: Fernando Antonio Rojas Morales

Introducción

En Colombia el acceso a la información de energía tanto del mercado de energía como del comportamiento de las variables meteorológicas e hidrológicas que se presentan en las principales plantas de energía y embalses en su mayoría es de acceso público; así pues, cualquier ente o persona interesados en estos datos pueden obtener acceso. El acceso y análisis a esta información, tanto histórica como en tiempo real es de importancia para las empresas o personas interesadas en realizar análisis y/o trabajos de comportamientos y pronósticos respecto al valor de las variables tratadas, con el fin de tomar decisiones y emprender proyectos relacionados.

Actualmente, la mayoría de esa información se obtiene mediante la consulta en algunos portales web a fines para esto (especializados en la administración de servicios de información de energía), los cuales brindan la opción de acceder a la información mediante consultas en su portal interactivo o mediante el uso de una API web. El acceso a esta información mediante esas opciones es algo ambiguo y muchas veces difícil de entender, pues se encuentra en diferentes formatos y alguna de esta no se encuentra categorizada, por lo que su localización y entendimiento es difícil, adicionalmente esos portales tampoco brindan un apartado para mostrar la información mediante graficas interactivas que puedan facilitar una comprensión rápida del comportamiento de esta.

Se quiere crear una herramienta software que pueda serle útil a cualquier entidad o persona de interés, en el que se tenga a disposición toda la información real e histórica de variables correspondientes a las estadísticas correspondientes a hidrología y a la generación de energía hidroeléctrica de las principales plantas de energía y fuentes hídricas del país. La herramienta software a crear debe facilitar la búsqueda y descarga de la información en tiempo real e histórico (por lo que se debe alimentar automáticamente de las fuentes de información y guardar esa

información en una base de datos) distribuida en categorías, además de tener a disposición diferentes cruces entre esa información, mediante la implementación de graficas interactivas para la realización de diversos análisis.

1. Objetivos

1.1. Objetivo General

Desarrollar una herramienta software automatizada para la recolección, visualización, interacción y descarga de información relacionada con los datos estadísticos meteorológicos y de hidrología de las principales plantas de energía y fuentes hídricas de Colombia.

1.2. Objetivos Específicos

Definir los reportes estadísticos que generará y mostrará la herramienta software y a partir de estos seleccionar las fuentes de información a usar.

Identificar los requerimientos para el diseño y desarrollo de la herramienta software.

Diseñar un modelo de datos relacional a usar en la base de datos.

Diseñar e implementar a partir de los requisitos previstos una herramienta software, la cual ofrecerá al usuario final la posibilidad de generar y descargar reportes actuales e históricos de los datos que seleccione de su interés.

Validar el funcionamiento de la solución implementada mediante un conjunto de pruebas funcionales.

2. Marco de referencia

2.1. Antecedentes y descripción del problema

Este proyecto surge en la socialización sobre algunas necesidades que se presentan en empresas y personas interesadas en el sector energético del país, en donde se manifiesta la necesidad de tener una herramienta en donde se puedan acceder a ciertos reportes y análisis estadísticos de interés en cuestión de la hidrología del país y concerniente a la generación de energía, esto con el fin de tener a fácil disposición reportes estadísticos y la información actual e histórica.

2.2. Fundamentos teóricos

2.2.1. Hidrología

Por su etimología de origen griego, ‘hidros’ significa agua y ‘logos’ ciencia: hidrología es la ciencia que estudia el agua. Pero una definición más apta y completa sería “Hidrología es la ciencia que trata de las aguas de la tierra, su ocurrencia, circulación y distribución, sus propiedades físicas y químicas y su influencia sobre el medio ambiente, incluyendo su relación con los seres vivos” (U.S. Federal Council for Science and technology, 1962). La hidrología también según la organización meteorológica mundial es la ciencia que trata de los procesos involucrados en el agotamiento y recuperación de los recursos hídricos disponibles en las áreas continentales de la tierra en las diversas fases del ciclo hidrológico(Aranda, 1998).

Cuando se habla del ciclo hidrológico se refiere a los distintos comportamientos que forman la hidrosfera (agua que se encuentra sobre la tierra sólida y bajo su superficie, en la corteza terrestre), en el que el agua se traslada de lugar debido a un ciclo bioquímico.

2.2.2. Principales Plantas de energía Hidroeléctrica de Colombia

Las fuentes hídricas son todas las corrientes de agua, ya sean subterráneas o sobre superficie, las cuales son aprovechadas para generar energía o para el uso personal. Colombia es uno de los países que más agua tiene en el mundo gracias a su ubicación geográfica, se tienen seis tipos de fuentes hídricas diferentes: Superficiales, subterráneas, marinas y oceánicas, termominerales, de alimentación glacial y de agua lluvia. El 97.5% del agua del mundo es oceánica y solo el 2.5% es de agua dulce, dividida en ríos, lagos, ciénagas y humedales; los cuales están divididos en centros hidrológicos.

Para el aprovechamiento de las fuentes hídricas para la producción de energía se usan los llamados embalses, los cuales son la acumulación de agua que es producida por la construcción de una represa sobre el lecho de un río o un arroyo, el cual cierra total o parcialmente su cauce. La represa es construida con piedra, tierra, hormigo o materiales sueltos la cual es levantada generalmente sobre un desfiladero (paso de montañas estrecho) con el fin de embalsar el agua. El embalse cuenta con un vertedero, el cual es una estructura hidráulica construida con el objetivo de dar paso a un flujo de agua controlado proveniente del embalse, cuando este alcanza niveles altos (generalmente en época de lluvias).

Figura 1

Represa, vertedero de borde.



Nota. Fotografía de la represa vertedero el borde. Tomado de Wikipedia.org

La mayoría de los embalses son unipropósito para generar energía eléctrica, pero existen también los multipropósitos, los cuales adicionalmente se usan para regular las crecientes y el abastecimiento de agua para acueductos (Palacios, 2013).

Las centrales hidroeléctricas están vinculadas a su respectivo embalse, siendo los principales embalses de Colombia los siguientes, clasificados por regiones (XM, 2022).

Tabla 1

Centrales hidroeléctricas en Colombia clasificadas por regiones

Región	Embalse
Antioquia	Amani
	Miraflores
	Peñol
	Playas
	Porce II
	Porce III
	Punchina
	RioGrande2

	San Lorenzo Troneras
Caribe	Urral
Centro	Agregado Bogotá Betania El Quimbo Muna Prado Topocoro Chuza Esmeralda Guavio
Oriente	Chuza Esmeralda Guavio
Valle	Altoanchicaya Calima1 Salvajina

Figura 2

Principales embalses de Colombia



Nota. Fotografías de los principales embalses de Colombia, Tomada de: oupme.gov.co

2.2.3. Meteorología

Es la ciencia que estudia los fenómenos que suceden en la baja atmosfera, en la cual se consideran las condiciones químicas, físicas y termodinámicas. Algunos de los parámetros que son objetivos de medición son: la temperatura, la humedad, presión atmosférica, intensidad y dirección

del viento (Uriarte, 2022). El objeto de este estudio es tener un conocimiento de la climatología y la previsión del clima, lo cual es muy útil para saber el comportamiento en corto y largo plazo de las fuentes hídricas y su afectación en las plantas de energía hidroeléctrica.

2.2.4. Herramienta software

Una herramienta o solución software está constituida por aquellos componentes y programas que en conjunto nos permiten resolver alguna idea o problema que responde a las necesidades de un usuario.

Existen diferentes tipos de soluciones software, las cuales son dependientes de donde estas estén aplicadas, como puede ser en sistemas operativos, bases de datos, administración y gestión de datos, sistemas de seguridad y respaldo, soluciones de nube, etc. Generalmente es una unión de estos conceptos para obtener un aplicativo que sirva para resolver una necesidad específica como puede ser la automatización de procesos.

De acuerdo con las necesidades de evaluación para una solución software se han categorizado algunos aspectos a tener en cuenta, los cuales son (Pearson, 2006).

-Mantenibilidad: Implica que la solución pueda ser modificada y evolucionar para así poder adaptarse a los cambios en las necesidades que puedan surgir.

-Confiablez: Hace referencia a un conjunto de características que la solución debe aportar, como lo es: funcionalidad, fiabilidad y seguridad.

Eficiencia: Su correcta ejecución evitando el desperdicio de recursos de sistema como memoria e hilos de procesamiento.

-Usabilidad: La solución debe ser lo más intuitiva posible, fácil de interactuar para el usuario y tener una interfaz apropiada para el concepto aplicativo en uso.

2.2.5. Bases de datos relacionales

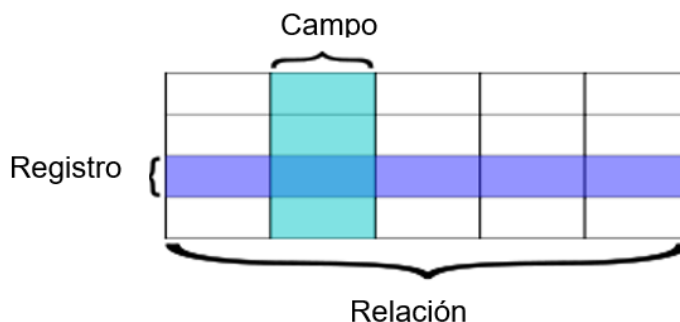
Una base de datos relacional es una recopilación de elementos de datos con relaciones predefinidas entre ellos, los cuales se organizan como un conjunto de tablas con columnas y filas. En las tablas se guarda información acerca de los objetos que se van a representar en la base de datos. En las columnas se guarda un determinado tipo de dato y un campo almacena el valor de ese dato. Las filas por su parte representan el valor de un dato relacionado a un objeto o entidad.

Los campos son los distintos tipos de datos que componen la tabla, por ejemplo, el nombre el apellido y el domicilio de una persona. En la definición del campo iría por ejemplo el nombre del campo, el tipo del campo, el ancho del campo y las posibles restricciones.

Los registros por su parte engloban la información que ira contenida en cada campo para un nuevo objeto.

Figura 3

Estructura tabla base de datos relacional



Nota. Fuente: El autor.

Una colección de datos (base de datos) contiene información acerca de una organización y el principal objetivo de un SGDB (sistema gestor de bases de datos) es almacenar y recuperar la información. La gestión de los datos implica la definición de las estructuras a usar para almacenar la información y los mecanismos para la manipulación de ella. La base de datos provisiona la

fiabilidad de acceso a la información (seguridad de acceso). Uno de los principales propósitos de la base de datos relacional es brindar una vista abstracta de los datos (Silberschatz, 2002).

Algunas de las principales características son:

- Las relaciones existen entre las tablas que componen la base de datos.
- No existen tablas con el mismo nombre.
- Cada tabla es un conjunto de columnas (campos) y filas (registros).
- Existen jerarquías para las relaciones entre las tablas, como padre e hijo, las cuales se manejan mediante las llaves (llaves primarias y foráneas)
- Se pueden crear índices, proporcionan acceso más rápido a una consulta

Llaves primarias: Son las que identifican exclusivamente a una tabla, solo puede existir un campo en una tabla que tenga esta llave.

Llaves foráneas: Es una referencia que se hace en un campo de la tabla hacia otra tabla, determinando la relación que existen entre ambas. Estas llaves no son únicas, puede existir varios campos en una tabla que tengan esta referencia.

Existen diversas restricciones que se le pueden agregar aun campo de una tabla, como puede ser su carácter único (Unique), si el valor poder ser nulo o no (Not null).

La gran ventaja que nos proporciona una base de datos relacional es el acceso fácil a la información de la base de datos, esto se hace mediante consultas SQL, el cual es un lenguaje de consultas estructurado que nos da las directrices que debemos seguir para acceder a una información específica.

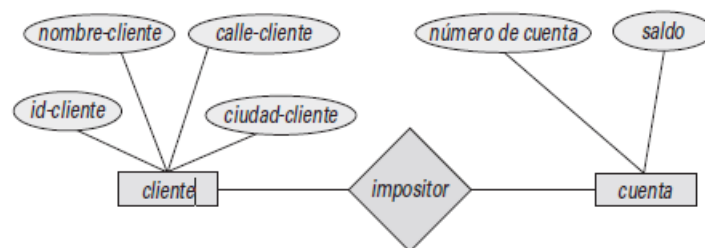
Para crear una base de datos relacional es conveniente seguir cierto procedimiento en el que planifica el tipo y las relaciones entre la información que se quiere almacenar, para ello existen

diversas maneras de representar las bases de datos que ayudan a entender y organizar la información con más claridad, estos son los principales:

- Descripción del problema: Se trata de relatar el problema que se quiere resolver o el objetivo que se busca al tener organizada la información en una base de datos, también se relata toda la información pertinente y las relaciones que existen entre ellos.
- Modelo Entidad-relación: Se trata de la creación de objetos llamados ‘entidades’ y las relaciones que entre ellos. Una entidad es un objeto indistinguible, y que son descritas por atributos. Y la relación es la asociación entre varias entidades, por ejemplo, cliente se asocia con otra en entidad país, para relacionar el cliente al país al que pertenece. La estructura básica de una base de datos se puede expresar mediante este modelo:

Figura 4

Ejemplo diagrama E-R



Nota. Fuente: El autor.

- El rectángulo representa la entidad. La elipse representa los atributos de la entidad, Los rombos las relaciones entre las entidades. Las líneas unen las entidades con sus atributos y a las entidades con las relaciones.

- Mediante este diagrama se expresa la correspondencia de cardinalidades, que me indica el número de entidades que se pueden relacionar entre sí.
- Modelo relacional: Este modelo representa la base de datos mediante un conjunto de tablas en donde se representan los datos y las relaciones entre ellos, cada tabla es una entidad y está compuesta por varias columnas que son sus atributos. Es un modelo basado en registros. Se crean tablas para cada entidad, para cada relación abstracta (muchos a muchos).

Figura 5

Ejemplo del modelo relacional.

<i>id-cliente</i>	<i>nombre-cliente</i>	<i>calle-cliente</i>	<i>ciudad-cliente</i>
19.283.746	González	Arenal	La Granja
01.928.374	Gómez	Carretas	Cerceda
67.789.901	López	Mayor	Peguerinos
18.273.609	Abril	Preciados	Valsain
32.112.312	Santos	Mayor	Peguerinos
33.666.999	Rupérez	Ramblas	León
01.928.374	Gómez	Carretas	Cerceda

(a) La tabla *cliente*

<i>número-cuenta</i>	<i>saldo</i>	<i>id-cliente</i>	<i>número-cuenta</i>
C-101	500	19.283.746	C-101
C-215	700	19.283.746	C-201
C-102	400	01.928.374	C-215
C-305	350	67.789.901	C-102
C-201	900	18.273.609	C-305
C-217	750	32.112.312	C-217
C-222	700	33.666.999	C-222
		01.928.374	C-201

(b) La tabla *cuenta*(b) La tabla *impositor*

Nota. Fuente: El autor.

Generalmente van de la mano el modelo entidad relación y el modelo relacional, ya que el modelo relacional a pesar de ser el más usado y el que da una visión más cercana a lo que es la base de datos en él no se muestran algunas de las relaciones entre las entidades. Adicionalmente cabe recalcar que al pasar del modelo entidad-relación al modelo relacional se debe hacer una normalización de las tablas para evitar redundancia en los datos.

2.2.6. Consultas a APIs

Por sus siglas en inglés, interfaz de programación de aplicaciones. Es un mecanismo provisto de definiciones y protocolos que permite la comunicación e intercambio de información entre sistemas, las cuales mediante ciertos procedimientos ofrecen una biblioteca de datos para ser usados por otro software. En el concepto de obtener datos abiertos desde la web se refiere a “API web”, las cuales son usadas para el intercambio de información entre diferentes entes de interés (Pardo, 2019).

Una API ofrece ciertas funcionalidades sobre un servidor en la web para que puedan ser usadas por aplicaciones cliente mediante el uso de ciertos procedimientos. Las API permiten la comunicación con otros sin saber su implementación, manteniendo la seguridad del servidor.

Para conectarse con las APIs y usar la información que estas proporcionan para crear aplicativos se pueden usar plataformas de integración que conecten los elementos. Para empezar a usar una API específica se requiere sobre la documentación de esta en la que se describa los detalles involucrados para su uso, como su forma de acceso y las operaciones interactivas.

La comunicación en la API, que es entre cliente y servidor se realiza mediante el intercambio de mensajes HTTP, los cuales se hacen en el rol de ‘peticiones’ (enviado por el cliente) y ‘respuestas’ (materialización de la solicitud realizada).

La petición contiene la información necesaria para que la API realice internamente la acción requerida. Las consultas en la API están diseñadas para ser accedidas por un llamado ‘método’, que es el que indica el tipo de procedimiento a realizar para recopilar la información y el cual debe ser especificado en la petición, estos métodos son: GET (recupera una representación de un recurso de datos), HEAD (recupera la cabecera de una respuesta), POST (crea un nuevo recurso de datos), PUT (actualiza o crea un recurso), PATCH (actualización parcial del recurso),

DELETE (elimina un recurso de datos existente), OPTIONS (obtiene las opciones de comunicación de cierto recurso). Normalmente una API no dispone de todos los métodos, siendo los métodos GET y POST los de uso más frecuente (Pardo, 2019).

2.3. Fuentes de información de acceso público para datos hidrológicos

2.3.1. XM

Es una empresa especializada en la gestión de sistemas en tiempo real, administradora del mercado de energía mayorista y el desarrollo de soluciones de servicio de energía e información (XM, 2022).

La gestión de sistemas está relacionada con el intercambio de información con sistemas transaccionales o plataformas tecnológicas.

XM opera el Sistema de Interconectado Nacional (SIN), el cual es la entidad que se encarga de distribuir la energía que generan todas las centrales hidroeléctricas en Colombia. También se encarga de la administración del Mercado de Energía Mayorista (MEM), por lo que realiza funciones de Centro Nacional de Despacho (CND), Administrador de Sistemas de Intercambios Comerciales (ASIC) y Liquidador de Cuentas de cargo por Uso del SIN.

Para lo que repercute a este proyecto interesa los datos que ofrece respecto a la operación del SIN, es decir lo referente a los recursos de generación eléctrica en Colombia; principalmente de las plantas hidroeléctricas, para lo cual XM dispone de un portal web (www.xm.com.co) el cual nos brinda información, que es de acceso público (histórica y actualizada) acerca de las características y propiedades que presentan los recursos hídricos (embalses, ríos, reservas) relacionados con las plantas hidroeléctricas. Algunas de esas características son, por ejemplo, el volumen útil del embalse, el volumen de energía que produce el embalse, el volumen y la energía de los vertimientos por embalse, los volúmenes de ríos que alimentan los embalses, su aporte en energía, el caudal de estos.

Para acceder a los datos ofrecidos por XM hay dos maneras, hallándose en cada una de ellas su singularidad, pues existe información que se comparte y también hay información exclusiva en cada una, aparte de que el formato de salida de los datos en cada una varia. La primera manera se basa en el ingreso al portal web y localizar la sección de interés y allí buscar la información y descargarla, todo vía web; y la segunda manera es hacer uso de la API web que ofrece XM, el cual nos brinda un repositorio en GIT con las indicaciones que se deben seguir para realizar las solicitudes a esta API.

2.3.2. UPME

La Unidad de Planeación Minero-Energética, es la unidad de Administración Especial de orden Nacional, adscrita al Ministerio de Minas y Energía de Colombia, creada en 1992. Su función es la de la planeación del desarrollo minero-energético, apoyo en la política pública.

En sus labores específicas se destaca la planeación con los agentes del sector minero energético con el fin de producir y divulgar información útil para el aprovechamiento de los recursos energéticos y garantizar el óptimo y oportuno abastecimiento de los recursos, su distribución, su utilización y sus efectos (UPME., 2022).

Uno de los servicios brindados por la UPME es el del Sistema de Información Minero Energético (SIMEC), el cual es la matriz de información, donde se consolidan las principales variables del sector para ser consultadas. A su vez la SIMEC se divide en SIEL (Sistema de información minero colombiano), SIAME (Sistema de información ambiental y minero energético) y el SI3EA (Sistema de información de eficiencia energética y energías alternativas).

El SIEL permite hacer consultas de generación hidroeléctrica, variables de los embalses, reservas, ríos, vertimientos; variables de generación eléctrica, consumo de combustible; todo de manera actual e histórica. Para acceder a estos datos basta con entrar al portal web del servicio

SIEL (<http://www.siel.gov.co>) y localizar la sección de interés y aportar los datos de consulta para posteriormente descargarlo con los distintos formatos que se ponen a disposición.

2.4. Tecnologías involucradas en el desarrollo del proyecto

Para conseguir los objetivos planteados en este proyecto se requiere realizar la consulta, gestión y puesta a disposición los datos de interés al usuario final. Para ello se hará uso de algunas herramientas de conocimiento general y otras más específicas. Para la gestión de datos se seleccionó ‘PostgreSQL’ como SGBD (sistemas de gestión de bases de datos) debido a su amplia capacidad de almacenamiento, escalabilidad y a su fácil uso en funciones de código en lenguajes de programación como ‘Python’, el cual es el lenguaje de programación en el que se desarrollara la implementación para la consulta, procesamiento y almacenamiento de datos, esto debido a que este cuenta con gran cantidad de funciones para ello. La generación y representación de reportes estadísticos se realizará mediante el framework ‘Streamlit’, el cual está basado en Python, mas adelante se profundizará más acerca de esta herramienta.

Adicional a las herramientas de desarrollo se usará un sistema de control de versiones, esto debido a que es importante poder tener un historial de cambios en las implementaciones y poder revertir algún cambio no deseado, para ello se seleccionó a ‘Git’, ya que este es gratuito, de código abierto y nos permite guardar toda la documentación relacionada al desarrollo.

2.4.1. Streamlit

Es un framework open-source escrito y para uso en Python para aplicaciones web, especialmente para front-end. Su uso es muy frecuente en aplicaciones de ciencia de datos y machine learning por su facilidad para integrar representaciones de los datos mediante graficas.

Streamlit es similar al framework Dash, diseñados para hacer diferentes representaciones de datos mediante tableros (dashboards) y graficos. Streamlit tiene la facilidad de que en él se crean muchos elementos del front-end estandar, sin la necesidad de gastar mucho tiempo personalizando elementos CSS y HTML. Otra ventaja es que es gratuito y creando una cuenta en su página web nos permite compartir la aplicación web creada (Carrion, 2022).

2.4.2. Herramientas de automatización

Inicialmente se realizará mediante el uso de archivos .bat y tareas programadas mediante el administrador de tareas de Windows y posteriormente se evaluará el uso de un programa externo que permita mayor personalización.

3. Metodología de desarrollo del proyecto

El desarrollo de las actividades para este proyecto se hará utilizando el modelo de prototipado evolutivo. Este modelo hace referencia a la ejecución de actividades concernientes al desarrollo de software en forma modular; esto quiere decir que se va a desarrollar el software por módulos, cada uno con su ciclo completo (Pressman, 2002).

El uso de este modelo metodológico permite un mayor entendimiento del propósito, diseño y requerimientos del sistema, obteniendo así un sistema basado en las necesidades reales del usuario.

El uso de este modelo brinda las siguientes ventajas:

- El usuario final obtiene resultados parciales de su producto.

- Los riesgos de integración del producto se mitigan en etapas tempranas porque la integración se va dando de forma progresiva conforme se va culminando cada módulo.
- El producto de software se puede ir afinando y mejorando en cada iteración.
- Se puede optar por la reutilización de componentes en cada iteración y mejorarlos.

Desventajas:

- Al no priorizarse los requisitos de manera integral desde el inicio, pueden surgir problemas posteriores en la arquitectura.
- El usuario piensa que el producto ya se encuentra terminado en las primeras entregas (Sommerville, 2002).

Los módulos propuestos para el desarrollo de este proyecto son los siguientes:

1. Identificación las necesidades de los potenciales usuarios del software para así definir los parámetros de consulta a realizar.
2. Selección de las fuentes de información (de acceso público) que se usaran. De acuerdo con las necesidades del usuario se buscarán las fuentes de información que aporten los datos a partir de los cuales se obtendrán los reportes deseados.
3. Estructuración de la base de datos: Se definirá una estructura para la base de datos en la que se almacenaran los datos que se descargue.
4. Diseño y desarrollo del back-end del software: En esta fase se realizará el desarrollo para automatizar la descarga de los datos, realizándoles tratamiento y organizándolos de acuerdo con la estructura de la base de datos planteada. Los datos por almacenar de acuerdo con su fuente usualmente vienen en diversos formatos, nombre de etiquetas y sin categorizar; por lo cual será necesario hacer el tratamiento de estos. Este aspecto es muy

importante, pues uno de los principales objetivos de este proyecto es tener la información de hidrología y generación de energía en una base de datos que este actualizada permanentemente y de manera ordenada. También se realizarán la implementación para la generación de los reportes estadísticos de las métricas de interés que se mostrarán.

5. Diseño y desarrollo del front-end del software: Se realizará la interfaz de visualización de la herramienta software con todas sus funcionalidades deseadas.
6. Pruebas: Se diseñarán y ejecutarán un set de pruebas funcionales de validación.

4. Definición de requerimientos

De acuerdo con la problemática presentada se definió que la herramienta software a crear deberá tener los siguientes requerimientos:

4.1. Requerimientos funcionales

- El sistema debe actualizar automáticamente los datos de las métricas a tratar en la base de datos.
- El sistema permite que cualquier usuario pueda acceder a todo el contenido desplegado.
- El sistema debe permitir acceder a información actualizada de las métricas al acceder a él.
- El sistema debe permitir la selección de la métrica que quiere generar los datos.
- El sistema debe permitir seleccionar el rango de fecha para la generación del reporte.
- El sistema debe permitir seleccionar los ejes que quiere mostrar en las gráficas.
- El sistema debe permitir la descarga de los datos generados para la métrica seleccionada.

4.2. Requerimientos no funcionales

- El sistema debe permitir su ejecución en cualquier navegador y sistema operativo.
- El sistema no debe tardar más de 30 segundos en generar las gráficas o la descarga de los datos.
- El sistema debe ubicar sus botones y acceso a secciones en un lugar visible.

5. Diseño

5.1. Definición de las métricas a tratar.

De acuerdo con la revisión bibliográfica de las entidades que trabajan con los datos concernientes a hidrología y generación de energía de las principales plantas y afluentes del país, se decidió seleccionar algunas métricas (las principales y más tratadas) para recolectar los datos concernientes a estas, con el fin de tenerlos a disposición de una manera organizada y categorizada en una base de datos y así poder ser consultarlos de una manera más fácil y ordenada y ser llevados a una aplicación para visualizarlos, analizarlos y descargarlos.

Las métricas que se tratan expresan un valor de un elemento de medición aplicado en un ámbito o contexto, ya sea a nivel embalse, río, recurso (planta generadora de energía) o a sistema (país), en un cierto rango de fecha y a cierta frecuencia. De acuerdo con el elemento de medición cada métrica puede estar expresada en distintas unidades de medición. Las métricas seleccionadas con su respectiva unidad y ámbito de medición son las siguientes:

Tabla 2*Métricas seleccionadas a descargar*

id	métrica	unidad	ámbito	frecuencia generación
1	Aportes Energía Río	kWh	rio	diaria
2	Volumen Útil Embalse (%)	%	embalse	diaria
3	Aportes Energía Río (%)	%	rio	diaria
4	Aportes Energía SIN	kWh	sistema	diaria
5	Generación Neta Energía Recurso	kWh	recurso	diaria
6	Volumen Útil Energía Embalse	kWh	embalse	diaria
7	Disponibilidad Real Energía Recurso	kW	recurso	diaria
8	Volumen Útil SIN Energía	kWh	sistema	diaria
9	Vertimientos Energía Embalse	kWh	embalse	diaria
10	Precio de Bolsa Nacional COL	COP/kWh	sistema	diaria

5.1.1. Descripción de las métricas a tratar.

- **Aportes Energía Río (KWh):** Caudales de energía de los ríos que aportan agua a algún embalse del SIN (Sistema Interconectado Nacional).
- **Volumen Útil Embalse (%):** Porcentaje de volumen almacenado en el embalse por encima del nivel mínimo del mismo.
- **Aportes Energía Río (%):** Porcentaje de aportes de energía de los ríos que aportan a algún embalse del SIN que están por encima de la media histórica (a término de 1 año).
- **Aportes Energía SIN (KWh):** Caudal diario de energía que todo el SIN en conjunto aporta.
- **Generación Neta Energía Recurso (KWh):** Generación diaria neta de energía de las plantas de energía nacionales en sus puntos de frontera.

- **Volumen Útil Energía Embalse (KWh):** Volumen almacenado por encima del Nivel mínimo técnico en los embalses, es reportado diariamente por los agentes operadores.
- **Disponibilidad Real Energía Recurso (KWh):** Disponibilidad promedio calculada a partir de la fecha de los eventos que modifican la disponibilidad de las unidades de generación de los generadores, así como de la disponibilidad reportada al Centro Nacional de Despacho al ocurrir el cambio de estado de una unidad.
- **Volumen Útil Energía Sistema (KWh):** Volumen almacenado por encima del Nivel mínimo técnico a nivel SIN, es calculado por la sumatoria de los aportes de los embalses pertenecientes al SIN.
- **Vertimientos Energía Embalse (KWh):** Cantidad de agua diaria que debe ser evacuada en los embalses cuando la reserva sobrepasa la capacidad máxima de almacenamiento.
- **Precio de Bolsa Nacional Colombia (COP/KWh):** Precio de oferta de la última planta flexible para atender la demanda comercial nacional, mas el delta de incremento para remunerar los costos no cubiertos de las plantas térmicas en el despacho ideal (XM, 2023).

5.2. Selección de las fuentes de información

Las fuentes de información seleccionadas para extraer y recolectar la información requerida son provenientes de XM, empresa administradora del mercado de energía mayorista de Colombia (MEN) y operadora del Sistema de Interconectado Nacional (SIN). Se selecciono esta fuente pues después de una revisión, es esta entidad la que nos brinda los datos de interés respecto a las métricas elegidas, datos de acceso público, de primera mano y de manera histórica de las operaciones que repercuten al SIN, teniendo categorizaciones tanto por embalse, ríos y plantas; y mostrándola en diferente frecuencia (horario, diario y mensual).

Para acceder a los datos ofrecidos por XM hay dos maneras, hallándose en cada una de ellas su singularidad, pues existe información que se comparte y también hay información exclusiva en cada una, aparte de que el formato de salida de los datos en cada una varia. La primera manera se basa en el ingreso al portal web y localizar la sección de interés y allí buscar la información y descargarla, todo vía web (el portal web de XM para la obtención de la información se conoce como Sinergox), lo cual se puede resumir en el ingreso a un enlace de descarga directo; y la segunda manera es hacer uso de la API web que la entidad ofrece, la cual nos brinda un repositorio en GIT con las indicaciones que se deben seguir para realizar las solicitudes a esta API.

En la Tabla N°3 podemos ver cuál es la fuente de información de acuerdo con cada métrica seleccionada.

Tabla 3

Fuentes de información de las métricas seleccionadas

id	Métrica	Fuente
1	Aportes Energía Río	API
2	Volumen Útil Embalse (%)	Enlace descarga
3	Aportes Energía Río (%)	Enlace descarga
4	Aportes Energía SIN	API
5	Generación Neta Energía Recurso	API
6	Volumen Útil Energía Embalse	API
7	Disponibilidad Real Energía Recurso	API
8	Volumen Útil SIN Energía	API
9	Vertimientos Energía Embalse	Enlace descarga
10	Precio de Bolsa Nacional COL	API

5.3. Diseño de la base de datos

Tener un actual e histórico registro en la base de datos de los datos de las métricas seleccionadas es la base de este proyecto, pues con estos datos (que se actualizan periódica y

automáticamente) es que se alimentara la herramienta software para mostrar los reportes y gráficos.

Se quiere tener un modelo de base de datos en el cual las métricas estén organizadas en tablas de acuerdo con su ámbito y en las cuales existan campos que relacionen a cada una de ellas con otras tablas que especifican el registro del ámbito en cuestión y el registro del tipo de métrica.

En base a esto se decidió realizar una base de datos relacional para el proyecto, es decir una base de datos SQL, este modelo es adecuado pues este se basa principalmente y para los propósitos mencionados: en la representación de los datos mediante tablas interrelacionadas y la atomicidad en la realización de las operaciones.

Se crearon dos tipos de tablas para agregar a la base de datos, las de tipo “métrica según el ámbito” que son en las cuales se consignara como su nombre lo indica el registro de las métricas según el ámbito (rio, embalse, sistema o recurso) de la medición; y las de tipo “dimensión” que son en las cuales se consignaran todos los registros que corresponden a cada ámbito de medición junto con los campos de características y relaciones que lo describen. De esta manera se asegura que el registro que se quiera almacenar dentro de una métrica en la base de datos debe estar con relacionado con un registro existente en alguna de las tablas de dimensiones.

Las tablas del tipo “métrica según el ambito” tiene un formato de asignación de nombre: “met_ambito”, donde ámbito se reemplaza por el nombre del ámbito correspondiente. Estas tablas registran datos de acuerdo con la frecuencia de carga de cada una de las métricas, consignando valores para los campos de la fecha de medición, el parámetro (indica cual es la métrica en cuestión, de la lista de tablas de métricas), el valor y registro del ámbito implicado. Por ejemplo, la tabla de las métricas que se miden por embalse será la tabla “met_embalse” y dentro de ella habrá columnas para la fecha, parámetro (llave foránea al parámetro que indica la métrica

consignada), ámbito (llave foránea al id de una de las tablas de dimensiones) y el valor medido para la métrica que indica el parámetro. Las tablas métricas definidas para usar en la base de datos son las siguientes:

- met_rio: Donde se reúnen las mediciones de todas las métricas relacionadas con ríos.
- met_sistema: Donde se reúnen las mediciones de todas las métricas relacionadas con un sistema.
- met_embalse: Donde se reúnen las mediciones de todas las métricas relacionadas con embalses.
- met_recurso: Donde se reúnen las mediciones de todas las métricas relacionadas con recursos (platas de generación)

Las tablas de tipo “dimensiones” tiene un formato de asignación de nombre: “dim_ambito”, donde ámbito se reemplaza por el nombre del ámbito correspondiente (por ejemplo, la tabla de dimensión). Estas tablas se crearon con el objetivo de tener registrados todos los registros de los ámbitos correspondientes con , esto para que estos puedan ser relacionados con los registros de las tablas de métricas en el momento de realizar su registro. Por ejemplo, la tabla de dimensiones del ámbito “río” se llama “dim_rio”, en ella se consignan todos los ríos que pertenecen al SIN y aportan a un recurso de generación, cada uno con sus campos de características y relaciones que lo describen.

Adicionalmente se creó una tabla especial de tipo dimensiones llamada “dim_parametro”, la cual es la que almacena cada una de las métricas descargadas y que serán vinculadas en las tablas de “métrica según ámbito” para especificar la métrica en cuestión y que esta categorizada en ese ámbito.

Las tablas de dimensiones las siguientes:

- dim_recurso: Tabla que especifica un listado de los recursos (plantas de generación) disponibles y a mapear.
- dim_rio: Tabla que especifica un listado de los ríos disponibles y a mapear.
- dim_embalse: Tabla que especifica un listado de los embalses disponibles y a mapear.
- dim_region_hidro: Tabla que especifica un listado de las regiones hidrologicas disponibles y a mapear.
- dim_sistema: Tabla que especifica un listado de los sistemas disponibles y a mapear.
- dim_parametros: Tabla que especifica un listado de las métricas que se medirán y la manera de acceder a ellas.

5.3.1. Diagrama entidad relación.

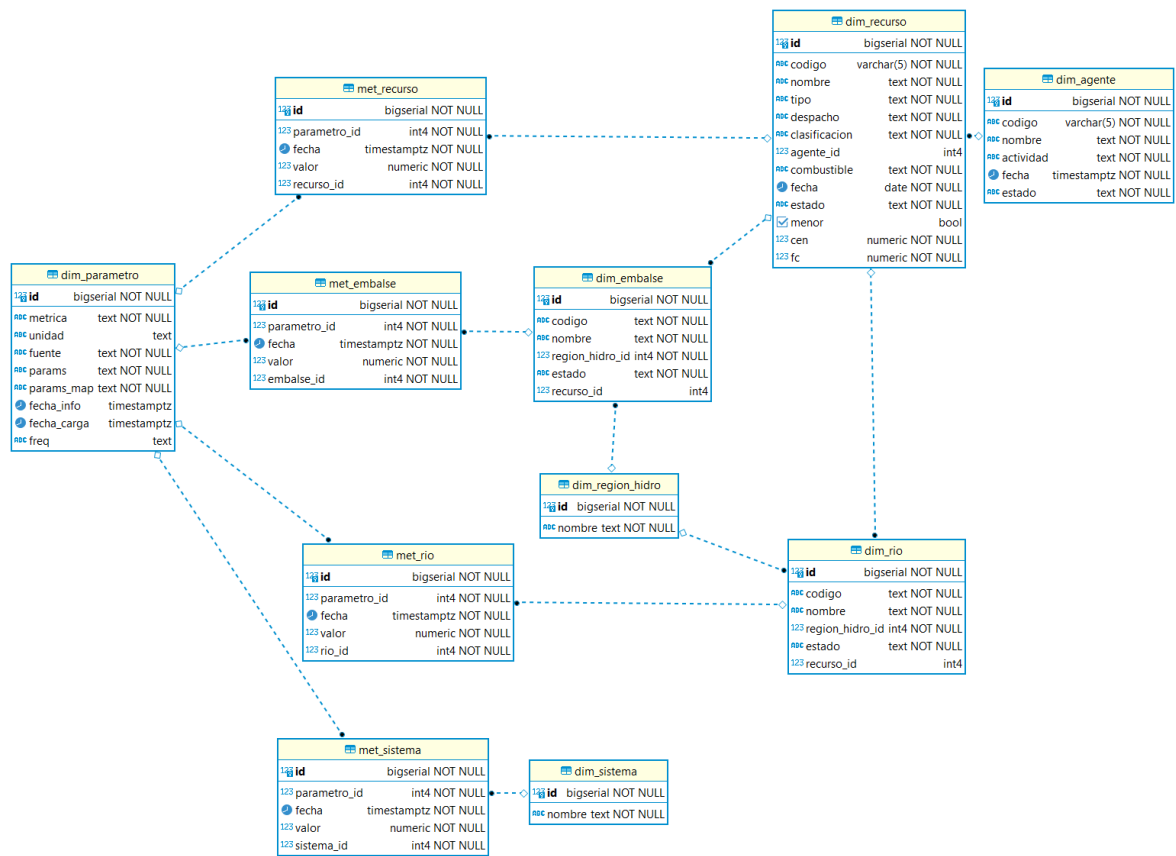
Es una representación de las entidades de una base de datos donde se muestran las características y las relaciones existente entre ellas. Se puede ejemplificar a las entidades como personas u objetos, los cuales poseen características propias que los definen y que conviven dentro de un sistema.

El diagrama se compone de entidades, relaciones y atributos; siendo los atributos aquellas características que definen a la entidad. En la descripción de cada atributo se encuentra la información sobre el tipo de dato y un valor acerca de si es dato puede ser nulo o no.

La realización de este diagrama sirve para entender más las relaciones entre las diferentes métricas usadas y así saber las dependencias entre estas, esto para tenerlo en cuenta cuando se desee realizar el almacenamiento de la información proveniente de la implementación (Lucidchart, 2022).

Figura 6

Diagrama Entidad-relación del modelo relacional de la base de datos



Nota. Fuente: El autor.

5.3.2. Entidades

La nomenclatura usada para la notación de la base de datos es la propia de del motor de base de datos usado en el proyecto: PostgreSQL

Fue necesario establecer columnas con restricción de unicidad (unique), esto con el fin de asegurar que no haya cargas repetidas en los registros añadidos.

Tabla 4*Entidad base de datos - "dim_recurso"*

Entidad: dim_recurso				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del recurso
codigo	varchar(5)		Not Null	Código del recurso
nombre	text		Not Null	Nombre del recurso
tipo	text		Not Null	Tipo del recurso
despacho	text		Not Null	Despacho del recurso
clasificación	text		Not Null	Clasificación del recurso
agente_id	integer	FK	Null	Id del agente del recurso
combustible	text		Not Null	Combustible del recurso
fecha	date		Not Null	Fecha del recurso
estado	text		Not Null	Estado del recurso
menor	bool		Null	Menor del menor del recurso
cen	numeric		Not Null	Capacidad neta del recurso
fc	numeric		Not Null	Factor de conversión del recurso
Restricciones:				
Unique (codigo)				

Nota. La tabla se refiere a los ríos que están involucrados en los recursos de generación de energía.

Tabla 5*Entidad base de datos - "dim_rio"*

Entidad: dim_rio				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del rio
codigo	varchar(5)		Not Null	Código del rio
nombre	text		Not Null	Nombre del rio
region_hidro_id	text	FK	Not Null	Id de la región hidrológica del rio
estado	text		Not Null	Estado del rio
recurso_id	text	FK	Not Null	Recurso al que pertenece el rio
Restricciones:				
Unique (código, nombre, region_hidro_id, recurso_id)				

Nota. La tabla se refiere a los embalses involucrados con los recursos de generación de energía hidroeléctrica.

Tabla 6

Entidad base de datos - "dim_embalse "

Entidad: dim_embalse				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del embalse
codigo	varchar(5)		Not Null	Código del embalse
nombre	text		Not Null	Nombre del embalse
region_hidro_id	text	FK	Not Null	Id de la región hidrológica del embalse
estado	text		Not Null	Estado del embalse
recurso_id	text	FK	Not Null	Recurso al que pertenece del embalse
Restricciones:				
Unique (código, nombre, región_hidro_id, recurso_id)				

Nota. La Tabla se refiere a regiones hidrológicas de Colombia, por ejemplo: Centro, Valle, Oriente, Antioquia, Caribe.

Tabla 7

Entidad base de datos - "dim_region_hidro"

Entidad: dim_region_hidro				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id de la región hidrológica
codigo	varchar(5)		Not Null	Código de la región hidrológica
Restricciones:				
Unique (nombre)				

Nota. La Tabla se refiere al país en el que se están generando los datos hidrológicos.

Tabla 8*Entidad base de datos - "dim_sistema"*

Entidad: dim_sistema				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del sistema
nombre	text		Not Null	Nombre del sistema
Restricciones:				
Unique (nombre)				

Nota. La Tabla es usada para almacenar los datos recolectados de las métricas que son relacionadas a ríos, es decir que miden valores a nivel rio.

Tabla 9*Entidad base de datos - "met_rio"*

Entidad: met_rio				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del valor de la métrica medido
parametro_id	text	FK	Not Null	Id de la métrica que se mide
fecha	timestamtz		Not Null	fecha de la medición
valor	numeric		Not Null	valor de la medición
rio_id	integer	FK	Not Null	Id del rio implicado en la medición
Restricciones:				
Unique (parámetro_id, fecha, rio_id)				

Nota. La Tabla es usada para almacenar los datos recolectados de las métricas que son relacionadas a sistema, es decir que miden valores a sistema o país.

Tabla 10*Entidad base de datos - "met_sistema"*

Entidad: met_sistema				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del valor de la métrica medido
parametro_id	integer	FK	Not Null	Id de la métrica que se mide

fecha	timestampz		Not Null	fecha de la medición
valor	numeric		Not Null	valor de la medición
sistema_id	integer	FK	Not Null	Id del sistema implicado en la medición
Restricciones:				
Unique (parámetro_id, fecha, sistema_id)				

Nota. La Tabla es usada para almacenar los datos recolectados de las métricas que son relacionadas a embalses, es decir que miden valores a nivel embalse.

Tabla 11

Entidad base de datos - "met_embalse"

Entidad: met_embalse				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del valor de la métrica medido
parametro_id	integer	FK	Not Null	Id de la métrica que se mide
fecha	timestampz		Not Null	fecha de la medición
valor	numeric		Not Null	valor de la medición
embalse_id	integer	FK	Not Null	Id del embalse implicado en la medición
Restricciones:				
Unique (parámetro_id, fecha, embalse_id)				

Tabla. La Tabla es usada para almacenar los datos recolectados de las métricas que son relacionadas a recurso (plantas de generación), es decir que miden valores a nivel recurso.

Tabla 12

Entidad base de datos - "met_recurso"

Entidad: met_recurso				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del valor de la métrica medido
parametro_id	integer	FK	Not Null	Id de la métrica que se mide
fecha	timestampz		Not Null	fecha de la medición
valor	numeric		Not Null	valor de la medición
recurso_id	integer	FK	Not Null	Id del recurso implicado en la medición
Restricciones:				

Unique (parámetro_id, fecha, recurso_id)

5.3.3. Carga de datos

Las tablas de la base de datos se alimentan mediante una implementación en Python que, de acuerdo con la tabla en la base de datos a cargar, hace el llamado a la API de XM o por medio de un enlace de descarga dinámico de la página web propia de XM para solicitar los datos y almacenarlos en la respectiva tabla de la base de datos. En esta implementación se mapean los datos solicitados de la métrica descargada a su respectiva tabla, columna y fila en la base de datos.

La implementación está diseñada para gestionar la disponibilidad por fecha de los datos, los datos nulos y verificar el tipo de datos que se quiere cargar y vincular a la base de datos. También se verifican que se cumplan las relaciones y restricciones solicitadas en la base de datos antes de ingresar los datos en ella.

La implementación se divide en 3 secciones, una para la carga de los datos de las tablas de dimensiones que se relacionan con las métricas a medir, otra para la carga de los datos de las tablas de las métricas proveniente de la API y otra para la carga de las métricas provenientes de un enlace de descarga directo proveniente de una página web.

La carga de las tablas de dimensiones se hace mediante el uso de enlace directo a la página de XM, estas tablas son las primeras en cargarse a la base de datos, pues son las que nos referenciarán algunos de los campos de los registros de las tablas de las métricas. La implementación solicita la descarga y hace el mapeo de la tabla respuesta a la respectiva tabla en la base de datos.

Para realizar la carga de las tablas de las métricas, la implementación hace uso de la predefinida tabla (“dim_parametro”), la cual tiene unos campos con información que indica los datos

y parámetros necesarios para la efectuar la descarga según su fuente de origen; así como los parámetros para mapear los campos descargados a la tabla y columna deseada en la base de datos. Esta tabla se creó de acuerdo con la fuente de origen y los parámetros que se necesitan especificar en esa fuente para realizar la descarga.

Para realizar las descargas mediante la API o mediante el uso del enlace directo de la página web se creó una tabla (“dim_parametro”), la cual es la base para que la implementación tome los datos necesarios y realice la descarga y el posterior cargue a la base de datos. En la Tabla 12 se encuentran la especificación de la tabla “dim_parametro”, donde se encuentran las métricas deseadas para descargar, con su respectiva fuente de descarga (API o página web), los parámetros que se deben incluir para conseguir su descarga desde la implementación y los parámetros que se deben mapear de otras tablas al descargar la métrica. También se encuentran columnas para las fechas en las que se cargó la última vez esa métrica y las fechas de disponibilidad de la información, esto para realizar la carga en el rango de fecha adecuado. Los parámetros de descarga son diferentes dependiendo la fuente de descarga.

Tabla 13

Entidad base de datos - "dim_parametro"

Entidad: dim_parametro				
Campo	Tipo de dato	Llave	Nulidad	Descripción
id	bigserial	PK	Not Null	Id del parámetro
metrica	text		Not Null	Métrica del parámetro
unidad	text		Null	Unidad de medición de la métrica
fuelle	text		Not Null	Fuente del parámetro
params	text		Not Null	Parámetros de descarga
params_map	text		Not Null	Parámetros para mapear en la db
fecha_info	timestamtz		Null	Fecha disponible de la información
fecha_carga	timestamtz		Null	Ultima fecha en que se realizó la carga
freq	text		Null	Frecuencia con la que se carga el parámetro

Restricciones:**Unique (métrica, fuente)**

Los siguientes son las columnas de la tabla con su descripción:

- Métrica: Nombre de la métrica que se quiere descargar.
- Unidad: Unidad de medición de la métrica.
- Fuente: Fuente de descarga de la métrica (API o página web de XM)
- Params: Los parámetros que son necesarios de especificar a la fuente para realizar la descarga. Se pasan como un conjunto de datos de tipo clave valor mediante un diccionario.
- Params_map: Los parámetros que deben ser especificados para realizar la carga de la métrica descargada hacia la base de datos. Se pasan como un conjunto de datos de tipo clave valor mediante un diccionario.
- Fecha_carga: Ultima fecha en la que se realizó la carga.
- Fecha_info: Ultima fecha de la información descargada.
- Freq: Frecuencia con la que se carga la información.

Especificación de la columna “params_map”, los valores se pasan mediante clave valor en un diccionario para su uso en la implementación en python:

- Tabla_met: Especifica el nombre de la tbla en la base de datos en la que se debe cargar la métrica.
- Col_fk: Especifica la columna dentro de la tabla de la métrica en la base de datos que es llave foránea al tipo de recurso que está midiendo la métrica. Es llave foránea a una de las tablas de dimensiones. E.g. “rio_id”, “recurso_id”, “sistema_id”.

- Tabla_dim: Tabla en la que se encuentra la columna de llave foránea especificada en el parámetro “Col_fk”. Es una de las tablas de dimensiones. E.g. “dim_recurso”, “dim_sistema”, “dim_embalse”.
- Col_map_dim: Columna en la tabla de dimensiones en la que se busca el valor de la llave foránea.
- Col_map_xm: Nombre de la clave en la que se busca el valor del nombre del recurso que está midiendo la métrica en la respuesta de la API. Esto debido a que la respuesta que da la API en su archivo JSON entrega una columna con una etiqueta variable para identificar el nombre del recurso medido, el cual es propio para cada métrica, por ejemplo, para la métrica “Aportes Caudal” se indica con “Name”, mientras que para la métrica “Aporres rio” aparece como “Nombre Río”.
- Conflicttools: Especificación de las columnas en las tablas de la base de datos que tienen restricción de unicidad. Los valores se pasan como una lista.

Los parámetros solicitados en la columna “Params” son diferentes de acuerdo con el tipo de fuente de descarga y se especificaran en el detalle de cada una de ellas.

5.3.3.1. Descarga a partir del llamado a la API de XM:

Para probar el funcionamiento de la API se utilizó el software “Postman”, el cual permite realizar consultas API con diferentes configuraciones.

El llamado a la API de XM se hace mediante el uso de un URL más la configuración de la llamada que se incluyen en los llamados “header” y “body” de la llamada. El URL para la solicitud de llamada es la conformación entre el URL de la API de XM más la granularidad (periodicidad con la que viene la información solicitada) de la información que se va a solicitar.

Los parámetros que podemos configurar para la llamada y que están incluidos en el “header” son metadatos que se mandan a la API para ayudar al servidor a comprender el tipo de solicitud como por ejemplo el “content-type”, que en este caso es de tipo “x-www-form-urlencoded”, lo que me indica que los datos solicitados se encuentran en un formato de clave y valor (Manish Shivanandhan, 2022).

Los parámetros que se pasan en el body son los que indican las especificaciones de la información solicitada, estos varían de acuerdo con la API, en el caso de la API de XM los valores permitidos son los siguientes:

- MetricId: Indica el código de la métrica a solicitar (e.g. AporEner, clave para solicitar la información relacionada con los Aportes de energía). Este es un parámetro obligatorio para la realización de la llamada.
- StartDate: Fecha de inicio del intervalo de los datos solicitados. Opcional en algunas métricas.
- EndDate: Fecha de fin del intervalo de los datos solicitados. Opcional en algunas métricas.
- Entity: Me indica la entidad a la que pertenece la métrica solicitada. Se encuentran los valores:
 - Sistema (Indica que la métrica pertenece a un ámbito informativo general que no refiere a valores medidos en recursos hídricos).
 - Recurso: Plantas de energía.
 - Embalses.
 - Ríos

Para conocer cuáles son los valores aceptados como parámetros para realizar el llamado a la API, XM en su página web nos brinda un enlace que nos dirige a un repositorio en Git el cual

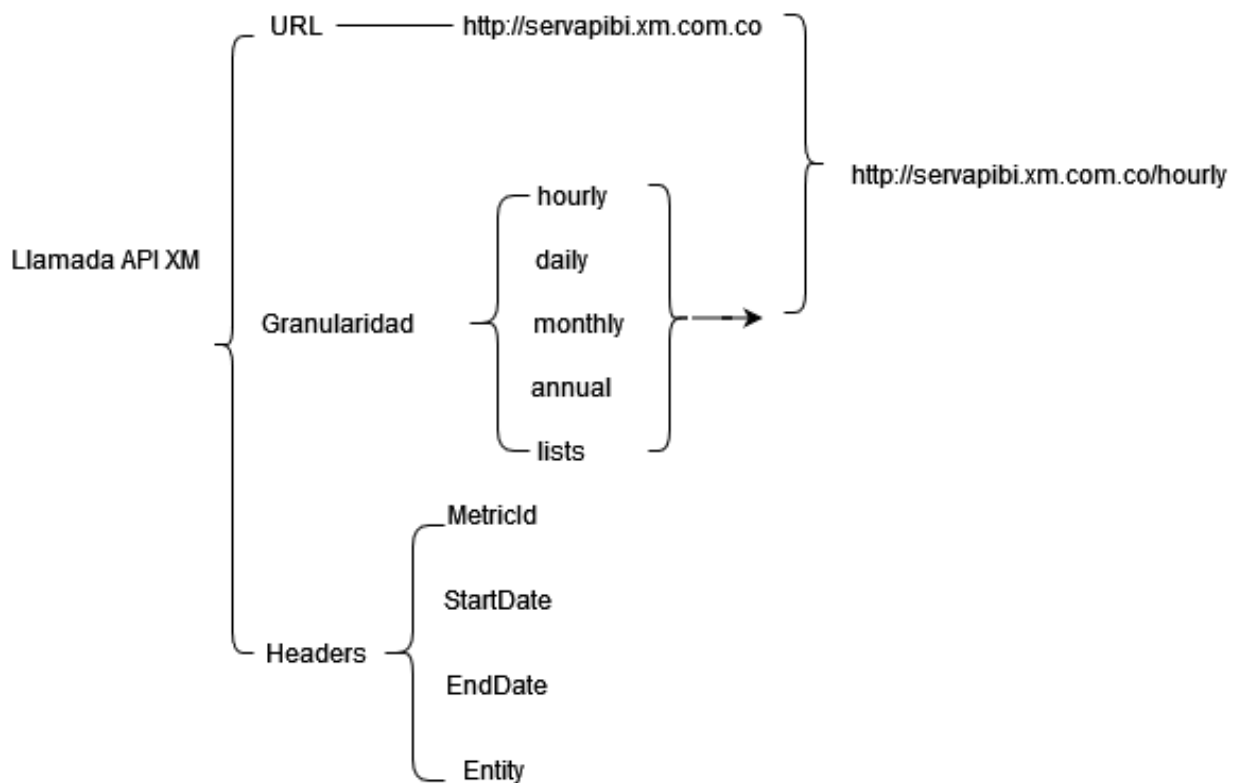
nos muestra el instructivo para poder realizar los llamados a su API dependiendo de la información que se desee consultar de acuerdo con la métrica buscada. Allí también encontramos un archivo JSON en el cual se citan todas las métricas permitidas para descargar con sus respectivos parámetros para acceder a ellas (archivo adjunto en anexos).

Toda la información solicitada vía API viene en formato JSON (Ver Figura N°8), del cual posteriormente serán mapeados los datos relevantes y de interés.

En el siguiente diagrama podemos observar cómo se configura una llamada a la API de XM:

Figura 7

Funcionamiento API XM



Nota. Fuente: El autor.

Cuando el valor de granularidad es “lists” significa que los datos solicitados son de tipo de lista de valores, como las dimensiones.

La API de XM cuenta con ciertas restricciones que se deben tener en cuenta al hacer los llamados que se refieren a la cantidad de información que se entrega como respuesta de acuerdo con la granularidad seleccionada, y son:

- Granularidad horaria o diaria entrega datos para 31 días.
- Granularidad mensual entrega datos para 731 días.
- Granularidad anual entrega datos para 366 días.

Aparte de las descargas de las métricas seleccionadas, fue necesario hacer la descarga de algunas de las tablas de dimensiones, las cuales me muestran una lista de una entidad específica. Las tablas de dimensiones para este proyecto son para : Recurso, Ríos, Embalses, Región hidrológica, Sistema. Ya que estas tablas no varían constantemente en el tiempo, estas son cargadas en un periodo de tiempo más extendido y no con una frecuencia establecida como si se cargan las tablas de dimensiones.

Los parámetros que solicita la implementación para realizar la descarga desde la API y que están especificados en la tabla de “dim_parametro” en la columna “params” son los siguientes:

- MetricId: Código de la métrica a solicitar (e.g. AporEner, clave para solicitar la información relacionada con los Aportes de energía). Este es un parámetro obligatorio para la realización de la llamada.
- Indicador: Debido a que algunas métricas cuentan con dos conjuntos de datos de acuerdo con el ámbito de recolección de datos (Ver Figura N°8), por ejemplo, la métrica “Aportes de Energía”. Debido a esto se hace una distinción de cuál es el indicador para la métrica deseada.

Figura 8

Ejemplo de una métrica con múltiple indicador

```
"AporEner": [  
  [  
    0,  
    "Aportes Energia",  
    "Sistema",  
    "Diaria"  
  ],  
  [  
    1,  
    "Aportes Energia por Rio",  
    "Rio",  
    "Diaria"  
  ]  
],
```

Nota. Fuente: El autor.

- Col: Nombre de la clave en el archivo JSON de respuesta que contiene los valores de la métrica.
- Datecol: Nombre de la clave en el archivo JSON de respuesta que contiene los valores de las fechas para métrica.

En la Figura N°9 y Figura N°10 se puede ver un ejemplo de una consulta y respuesta de la API frente a la consulta de la métrica “Aportes Energía por Sistema” visto en Postman:

Figura 9

Ejemplo configuración de una consulta API en “Postman”

The image shows a screenshot of the Postman API client interface. At the top, the method is set to 'POST' and the URL is 'http://servapibi.xm.com.co/daily'. Below the URL bar, there are tabs for 'Params', 'Authorization', 'Headers (10)', 'Body', 'Pre-request Script', 'Tests', and 'Settings'. The 'Body' tab is selected and highlighted with an orange underline. Under the 'Body' tab, there are radio buttons for different body types: 'none', 'form-data', 'x-www-form-urlencoded' (which is selected), 'raw', 'binary', and 'GraphQL'. Below these options is a table with four rows, each representing a form parameter. Each row has a checked checkbox in the first column, followed by the 'KEY' and the 'VALUE'.

	KEY	VALUE
<input checked="" type="checkbox"/>	metricid	AporEner
<input checked="" type="checkbox"/>	entity	Sistema
<input checked="" type="checkbox"/>	startdate	2022-07-01
<input checked="" type="checkbox"/>	enddate	2022-07-28

Nota. Fuente: El autor.

Figura 10*Ejemplo Formato de salida respuesta API*

```
{
  "Metric": {
    "Id": "AporEner",
    "Name": "Aportes Energía por Sistema",
    "StartDate": "2022-07-01T00:00:00",
    "EndDate": "2022-07-28T00:00:00"
  },
  "Items": [
    {
      "Date": "2022-07-01",
      "DailyEntities": [
        {
          "Id": "Sistema",
          "Value": "274960500.000000"
        }
      ]
    },
    {
      "Date": "2022-07-02",
      "DailyEntities": [
        {
          "Id": "Sistema",
          "Value": "322783600.000000"
        }
      ]
    }
  ]
}
```

Nota: Fuente El autor.

5.3.3.2. Descarga a partir del enlace directo ofrecido en la página web de XM:

Debido a que no todas las métricas de interés para este proyecto se encuentran en la API de XM, pero si se encuentran en su página web y son accedidas para su descarga mediante un enlace directo, fue necesario hacer una inclusión en la implementación que permita realizar la descarga por este medio, esto a partir de la modificación automatizada del link de descarga. El archivo que se descarga en un .xls.

En la implementación que se instaurara ira modificando este enlace de acuerdo con las necesidades, ya sea por la métrica deseada o por el rango fechas.

Para realizar la descarga por medio de esta fuente se hizo una inclusión de este método en la tabla de “dim_parametro”, que es la que me indica los parámetros necesarios para realizar la descarga de cada métrica. Los parámetros para realizar la descarga y que se incluyen en la columna “params” son los siguientes:

- Filename: Nombre del archivo que se desea cargar.
- Link: Link del archivo a cargar. Se agrega el histórico tambien. [Reciente, histórico]. Esto debido a que la pagina proporciona un enlace diferente cuando la información solicitada no es la última generada.
- Last_year: Último año disponible para la métrica en la carpeta de información de a página web de XM.
- Col: Nombre de la columna que contiene los valores de la métrica en el archivo de Excel que se descarga.
- Row_headers: Numero de fila en Excel que contiene el nombre de las columnas. Esto debido a que para unas métricas los nombres de las columnas empiezan en lugares diferentes.
- Datecol: Nombre de la columna en el archivo de Excel donde esta la fecha.

5.3.3.3. Diagrama de secuencia de extracción de datos.

El diagrama de secuencia describe cómo y en qué orden funcionan un grupo objetos para cumplir con una funcionalidad específica. Estos diagramas muestran la interacción ordenada para ciertos eventos. Muestra los objetos o componentes relacionándose mediante el intercambio de mensajes y respuestas.

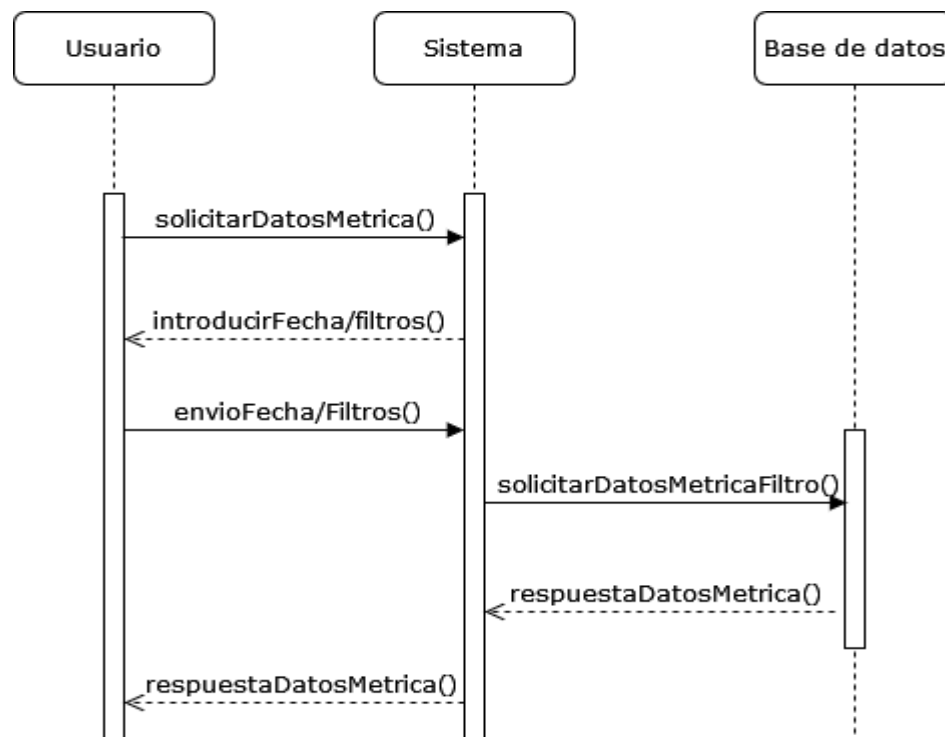
En la representación el eje vertical esquivale al tiempo y el eje horizontal a los participantes de la interacción, cada participante tiene una línea de representación en el eje x y los mensajes se muestran de acuerdo con la necesidad e interacción mediante flechas entre ellos.

Figura 11

Diagrama extracción de datos por parte de la implementación que descarga las métricas automáticamente.



Nota. Fuente: El autor

Figura 12. Diagrama de extracción de datos por parte del usuario

Nota. Fuente: El autor

5.3.4. Automatización de descarga de información a la base de datos desde las fuentes de información seleccionadas.

Para conseguir que la información en la base de datos se encuentre actualizada siempre se decidió vincular la implementación de Python que realiza la carga de datos a la base de datos con un programa de automatización de tareas, el cual se encarga de ejecutar la implementación periódicamente en un intervalo de tiempo establecido.

Para conseguir una gestión óptima a prueba de errores en la carga de la base de datos, la implementación en conjunto con el programa de automatización cuenta con una depuración de

errores y notificación vía email en caso de presentarse un error en la ejecución y posterior carga de alguna o varias métricas.

Para la carga de las tablas de dimensiones, puesto que estas tablas son las que nos brindan los datos de los ámbitos en los cuales se miden las métricas (recursos, embalses, ríos o sistema). Estas tablas se cargan con una frecuencia mucho mayor, puesto que solo necesitaran ser actualizadas cuando se generen nuevos recursos, embalses o ríos en el sistema. Para este tipo de tablas se decidió una frecuencia de carga mensual, para de esta manera ir verificando si se han generado nuevos recursos en el sistema y así agregarlos, o simplemente dejar la tabla con los valores que ya tenía.

La frecuencia de ejecución para las tablas de métricas, ya sean provenientes de la API o del enlace directo de descarga se decidió de acuerdo con periodicidad con la que aparece la métrica que debe cargarse de manera más frecuente. La implementación tomara el dato de la frecuencia de carga de cada métrica de la tabla de los parámetros de las métricas (“dim_paramatreo”) (Ver Tabla N°1), y de acuerdo con este valor y al tiempo transcurrido desde la última carga de la métrica, se permite o no realizar la petición para su descarga y actualización en las tablas.

5.4. Diseño de la herramienta software

Puesto que lo fundamental en este proyecto es la correcta y periódica consecución de los datos a almacenar en la base de datos para tenerlos disponibles para visualizar, extraer y realizar gráficos y reportes estadísticos a partir de ellos; se decido usar un framework para crear la aplicación interactiva para el usuario, la cual pueda vincular la extracción de los datos de la base de datos en tiempo real y realizar gráficos, reportes estadísticos y opciones de descarga de estos, todo esto mediante la implementación de una interfaz gráfica.

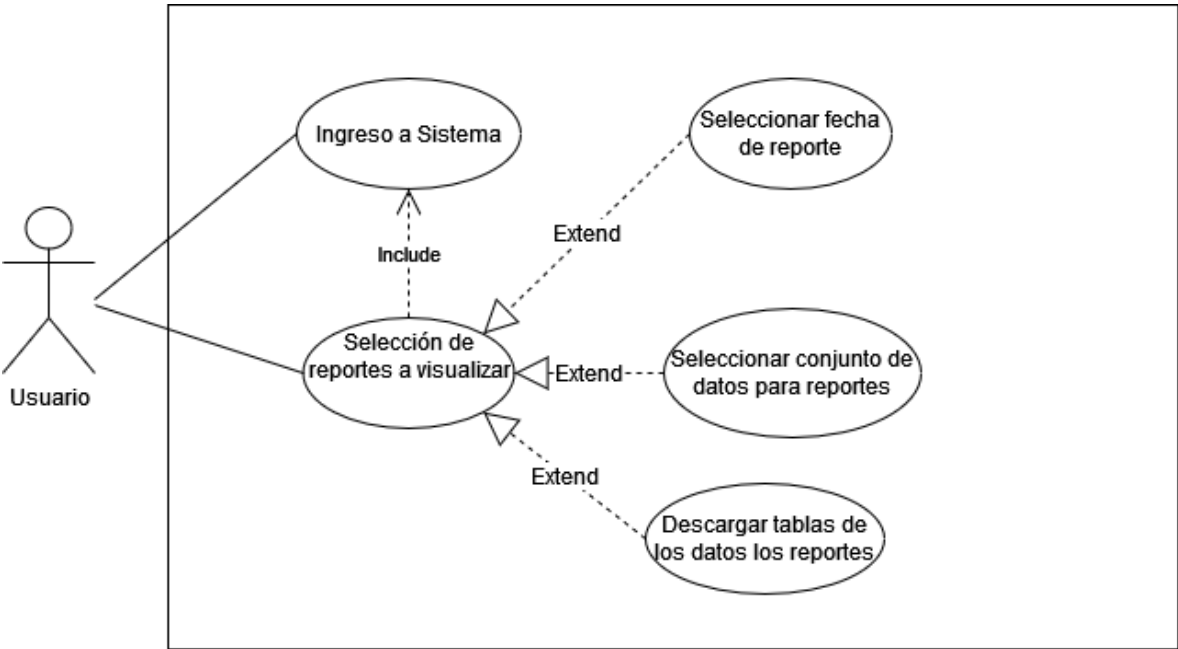
El framework que se seleccionó para los propósitos mencionados ha sido ‘streamlit’, el cual se ejecuta en Python y es ideal para la manipulación de datos y la representación de reportes y gráficos estadísticos debido al fácil acceso a librerías de tratamiento y representación de datos de Python como lo son: Pandas, Numpy, Plotly, Matplot Lib, etc.

5.4.1. Casos de Uso

Los casos de uso se usan para describir la interacción entre el usuario y el software.

Figura 13

Casos de Uso



Nota. Fuente: El autor.

Tabla 14

Caso de uso 1 – Ingreso al sistema

CASO DE USO 1	
Caso de Uso	Ingreso al sistema
Actor(es)	Usuario

Propósito	El usuario ingresa satisfactoriamente al sistema.	
Descripción	Es el caso de uso para verificar que el usuario ingresa al sistema.	
Precondición(es)	Que el software este instalado y permita su ejecución desde el navegador sin problema.	
Flujo principal	Acciones de Actor(es)	Respuesta del Sistema
	1. El usuario ingresa la dirección del servidor de ejecución de la herramienta web en el navegador y accede.	
		2. Se muestra la pantalla de inicio del software, la cual incluye una bienvenida.
Sub-flujos	A el usuario no le es permitido ingresar al sistema.	
Poscondición(es)	Apertura de la pantalla principal del sistema.	

Tabla 15

Caso de uso 2 – Selección de reportes a visualizar.

CASO DE USO 2		
Caso de Uso	Selección de reportes a visualizar.	
Actor(es)	Usuario	
Propósito	Permitir al usuario seleccionar el reporte deseado para visualizar.	
Descripción	El usuario de acuerdo con su interés puede seleccionar entre alguno de los reportes que se disponen para su interacción.	
Precondición(es)	Que el usuario haya ingresado y se encuentre en la página de inicio del sistema.	
Flujo principal	Acciones de Actor(es)	Respuesta del Sistema
	1. El usuario da clic en el apartado de los reportes meteorológicos que se encuentran en el menú de lateral.	

	2. Se muestra la pantalla del apartado de los reportes meteorológicos en la que se incluye una breve descripción y submenús para acceder al reporte deseado.
3. El usuario da clic en uno de los reportes que se muestran a disposición como submenú de la pantalla anterior	
	4. Se muestra el reporte seleccionado (para la fecha predeterminada: fecha actual). El reporte incluye la tabla con los valores, la respectiva gráfica y las herramientas de selección de datos.
Sub-flujos	El reporte no se puede cargar de forma correcta.
Poscondición(es)	Visualización del reporte deseado.

Tabla 16*Caso de uso 3 - Selección Fecha de reporte*

CASO DE USO 3		
Caso de Uso	Seleccionar fecha de reporte.	
Actor(es)	Usuario	
Propósito	Permitir que el usuario seleccione la fecha a la cual desea la generación del reporte.	
Descripción	Los reportes se generan de acuerdo con la fecha elegida, sino se elige fecha, el reporte se generará a partir de la fecha actual de ingreso.	
Precondición(es)	Que el usuario haya ingresado al sistema.	
Flujo principal	Acciones de Actor(es)	Respuesta del Sistema

	1. El usuario interactúa con el calendario que se encuentra en el menú lateral para elegir la fecha deseada.
	2. Se muestra la fecha elegida debajo del calendario y las pantallas de selección de reportes.
Sub-flujos	Si el usuario no selecciona ninguna fecha, la fecha predeterminada será la actual del ingreso al sistema.
Poscondición(es)	Se cambian las fechas de generación de los reportes y graficas asociadas.

Tabla 17*Caso de uso 4 - Selección de conjunto de datos para el reporte*

CASO DE USO 4	
Caso de Uso	Selección de conjuntos de datos para reportes.
Actor(es)	Usuario
Propósito	Permitir que el usuario sesgue o adicione más campos para la generación de reportes.
Descripción	Algunos reportes cuentan con filtros que permiten al usuario agregar o quitar campos para la generación de los reportes deseados.
Precondición(es)	El usuario ha ingresado a la sección de elegir un reporte para mostrar ubicada en la sección de reportes meteorológicos.
Flujo principal	Acciones de Actor(es)
	Respuesta del Sistema
	1. El usuario hace clic en uno de los filtros que se encuentran en el reporte abierto.
	2. Se muestran las opciones para elegir en ese filtro.
	3. El usuario según sea el caso, selecciona uno o más de las opciones mostradas.

	4. El reporte que se muestra en la página cambia para mostrar en adición o sustracción los filtros elegidos.
Sub-flujos	Si el reporte elegido no tiene filtros para usar o si el usuario no selecciona ninguno, se mostrará el reporte con sus campos predeterminados.
Poscondición(es)	Visualización del reporte con su respectiva grafica asociada con la adicción o sustracción de campos realizada.

Tabla 18

Caso de uso 5 – Descargar tablas de los datos de los reportes

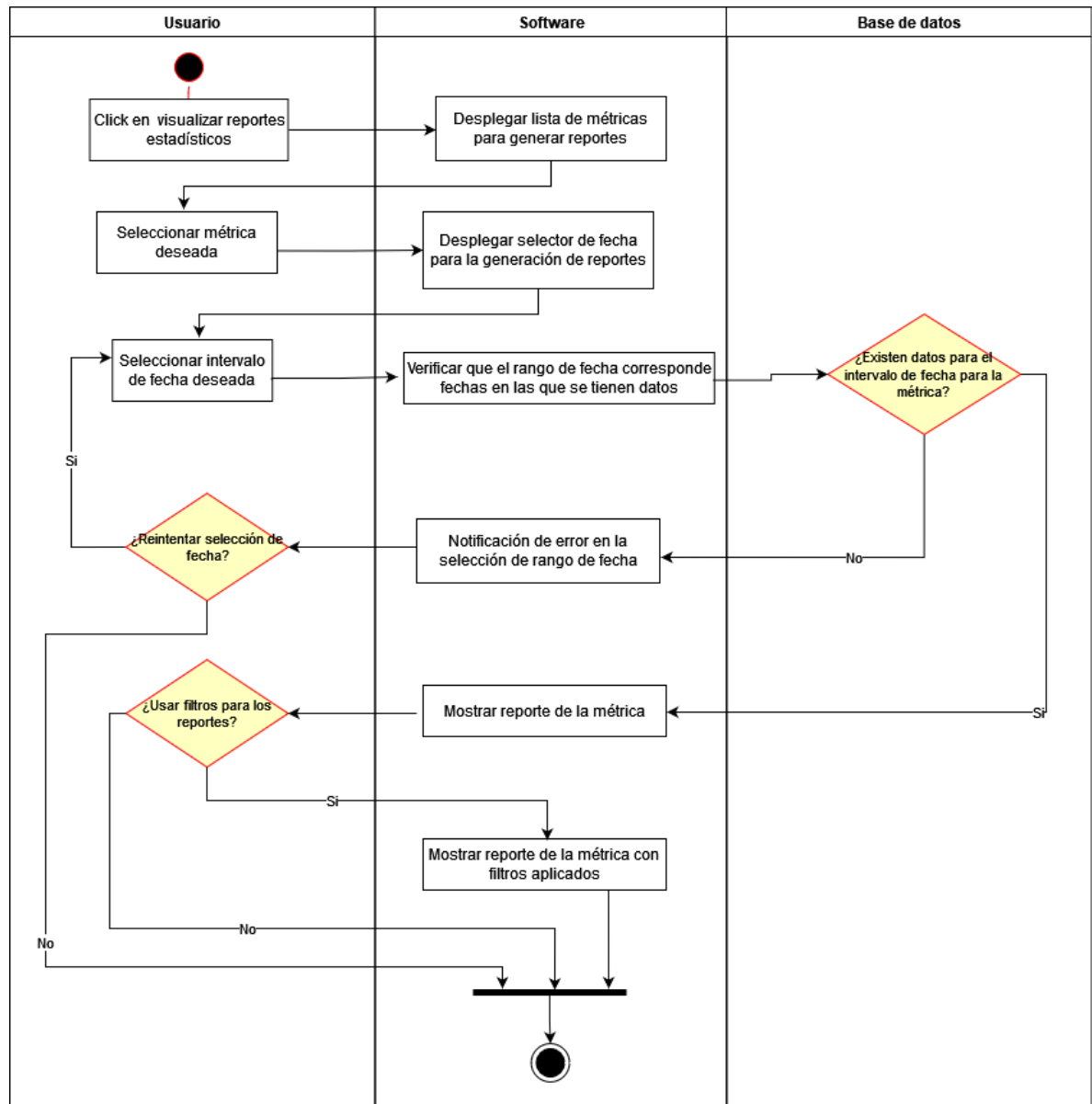
CASO DE USO 5	
Caso de Uso	Descargar tablas de los datos de los reportes.
Actor(es)	Usuario
Propósito	Permitir al usuario la descarga de los reportes seleccionados y generados.
Descripción	Este caso se inicia cuando el usuario descarga un reporte después de su generación.
Precondición(es)	El usuario ha seleccionado un reporte, con o sin filtros. El usuario se encuentra en la pantalla de visualización del reporte.
Flujo principal	Acciones de Actor(es)
	Respuesta del Sistema
	1. El usuario da clic en el botón de descargar.
	2. El sistema envía un archivo .xls con los datos del reporte elegido, de acuerdo con la fecha elegida o predeterminada.
Sub-flujos	El usuario no oprime el botón de descarga.
Poscondición(es)	El usuario recibe un archivo .xls con los datos de los reportes.

5.4.2. Diagrama de actividades

En el diagrama de actividades ayudan a ilustrar el comportamiento y el flujo de tareas ejecutadas por el sistema. Los diagramas de actividades muestran las posibles acciones que realiza y solicita el sistema de acuerdo con la interacción del usuario con él, considerándose desde el punto que ha iniciado la visualización del sistema (Lucidchart, 2022).

Figura 14.

Diagrama de actividades para la generación de reportes estadístico de acuerdo con una métrica seleccionada



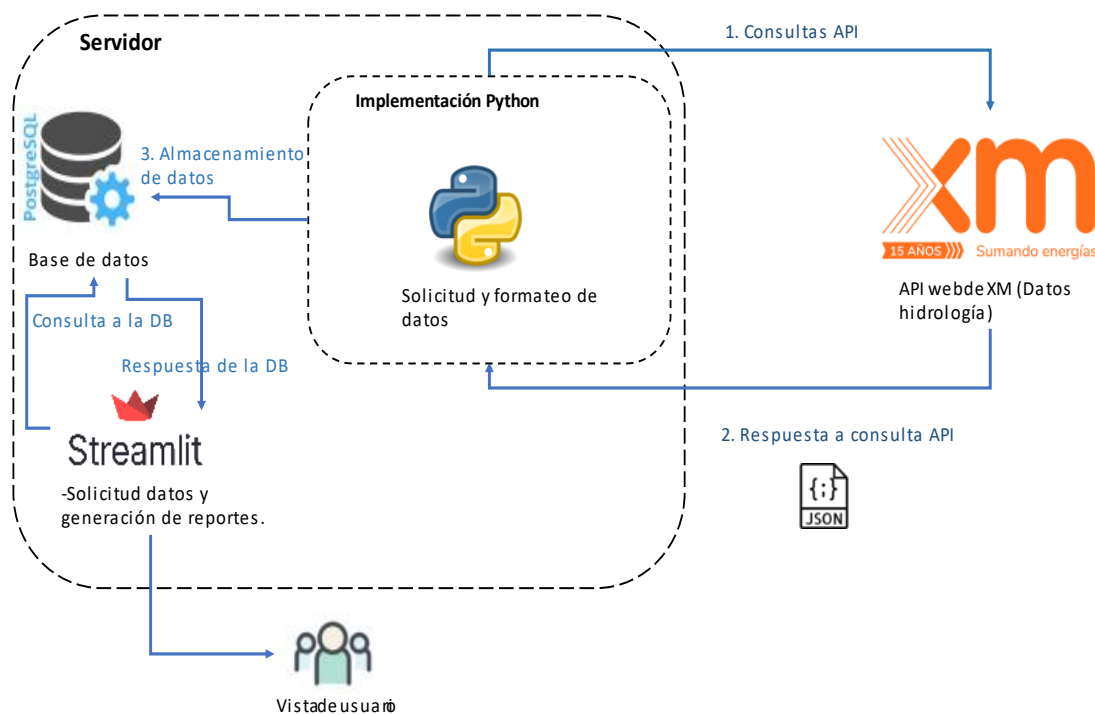
Nota. Fuente: El autor

5.5. Diagrama de arquitectura

El diagrama de arquitectura es muy útil para representar visualmente los componentes de un sistema y las interacciones que tienen entre sí para realizar ciertos propósitos, en él se muestra el resultado de ensamblar ciertos elementos arquitectónicos para satisfacer una funcionalidad y ejecución de un requisito del sistema (Amazon, 2023). El diagrama ilustra las transacciones de datos entre los diferentes componentes, de esta manera mostrando también como el software interactúa con medios exteriores (Wolf, 1992).

Figura 15

Diagrama de arquitectura - extracción de datos, generación y visualización de reportes



Nota. Fuente: El autor

6. Implementación

El desarrollo de la interfaz interactiva para el usuario está diseñado para mostrarle reportes estadísticos, los cuales incluyen gráficos interactivos acerca de las principales métricas relacionadas a los datos de hidrología y de generación de energía hidroeléctrica del país.

Las funciones principales que se presentan en el desarrollo son:

- Acceso a la información actualizada de los reportes relacionados a métricas seleccionadas.
- Selección de la métrica deseada para la generación y visualización del reporte.
- Permitir seleccionar el rango de fechas para la generación del reporte.
- Permitir seleccionar entre los ejes que quiere mostrar en las gráficas.
- Permitir la descarga de los datos generados por la métrica seleccionada en el intervalo de tiempo elegido.

Para poder implementar las anteriores funciones el sistema de manera automática y en segundo plano debe realizar la consulta y almacenamiento de los datos de las métricas de acuerdo con la fuente de información seleccionada para ello.

El desarrollo concerniente a la generación de reportes mediante la consulta a los datos de las métricas almacenados en la base de datos y la interfaz gráfica interactiva para el usuario de estos se realizó con el uso del framework de código abierto basado en Python “Streamlit”. Streamlit fue diseñado especialmente para la creación de aplicaciones web, orientado a la ciencia de datos, enfocándose en la presentación y análisis de datos. Streamlit permite fácilmente la consulta de datos a bases de datos y la creación de tableros estadísticos automatizados y tablas a partir de cualquier conjunto de datos, esto gracias a la interacción con librerías especializadas para ello. Streamlit proporciona una biblioteca de componentes preconstruidos para la creación de las diferentes interfaces interactivas para el usuario, las operaciones con los datos, y la generación de los gráficos estadísticos se hacen con bibliotecas y herramientas de visualización como Padas,

Numpy, Matplotlib, Plotly. Este framework es completamente personalizable en sus elementos HTML y es posible agregarle personalización mediante código HTML, CSS, JavaScript (Keepcoding, 2023).

6.1. Diseño de Interfaces

De acuerdo con los requerimientos del sistema, en conjunto con las funcionalidades ofrecidas por el framework “Streamlit”, se decidió realizar las siguientes interfaces para el acceso al usuario:

Interfaz 1- Inicial: Es la primera pantalla que nos muestra la aplicación al ingresar en ella, ahí se muestra en el menú lateral las secciones a las que se puede acceder, como también se muestra una sección central con un mensaje de bienvenida. Las secciones a las que se tiene acceso incluyen la sección de visualización de la lista de reportes estadísticos a seleccionar para visualizar referentes a la sección respectiva. El menú lateral se puede esconder/visualizar, esto para tener una mejor visualización de la página.

Figura 16

Interfaz 1 - inicio (Pantalla de bienvenida) de la herramienta software



Nota. Fuente: El autor

En la tabla N°19. se indican las secciones de esta interfaz.

Tabla 19

Especificación secciones interfaz 1

Sección	Contenido
1	Título del aplicativo.
2	Resumen del del tipo de contenido e interacción posible del usuario que se encontrar dentro del aplicativo.
3	Descripción del contenido de las secciones que se podrán encontrar dentro del aplicativo.
4	Menú de las secciones a las que es posible acceder.
5	Imagen/logo del sitio.
6	botón cerrar menú de secciones

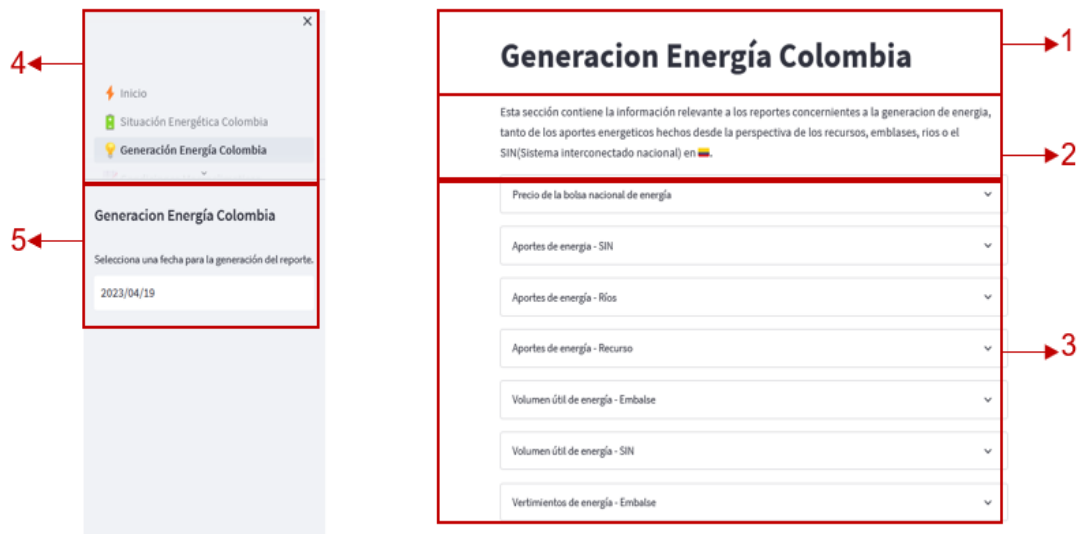
Interfaz 2 - de selección de reporte estadístico a visualizar: Es la pantalla a la que se accede después de dar clic en el botón de una de las secciones del menú lateral desde la interfaz inicial. En esta pantalla aún tenemos el menú lateral de la pantalla de inicio para volver a esa

pantalla o para cambiar a otra sección, como también se tiene el componente del calendario para seleccionar el rango de fecha deseado para la presentación de los reportes.

En esta pantalla central se nos muestra una breve descripción de los tipos de reportes que se encontrarán y una lista de los reportes estadísticos disponibles para seleccionar y visualizar de acuerdo con la sección accedida.

Figura 17

Interfaz 2 - selección de reporte estadístico a visualizar.



Nota. Fuente: El autor

En la Tabla N°20 se listan las secciones de esta interfaz:

Tabla 20

Especificación secciones interfaz 2

Sección	Contenido
1	Título de la sección.
2	Resumen del contenido que se encontrara en la sección.
3	Lista de los reportes estadísticos disponibles para ser accedidos.
4	Menú de las secciones a las que es posible acceder.
5	Calendario para selección de la fecha desde la que se desea generar el reporte.

Al hacer clic en la casilla lateral de la fecha se desglosa un cuadro para seleccionar la fecha deseada desde la cual se desea mostrar el reporte hasta la fecha actual.

Interfaz 3 - de selección de fecha para la generación del reporte estadístico a visualizar: Es la interfaz que se accede cuando se decide seleccionar una fecha (fecha inicial para la generación del reporte estadístico) en el widget calendario, al seleccionar esa fecha el reporte estadístico se genera desde esa fecha seleccionada hasta la fecha actual en el calendario (Si no se selecciona fecha la fecha será la del primer día del presente año).

Figura 18

Interfaz 3 - selección de fecha a generar reporte.



The screenshot displays the 'Generacion Energía Colombia' interface. On the left, a calendar for March 2023 is shown, with the date 14th highlighted in red. Below the calendar, a text box contains the date '2023/03/14'. On the right, the title 'Generacion Energía Colombia' is followed by a descriptive paragraph: 'Esta sección contiene información relevante a los reportes concernientes a generacion de energía, tanto de los aportes energeticos hechos desde la perspectiva de los recursos, embalses, rios o el SIN(Sistema interconectado nacional) en Colombia.' Below this, there is a list of seven energy generation metrics, each with a dropdown arrow: 1. Aportes de Energía Acumulado - Ríos (GWh), 2. Volumen útil de energía - Embalse (%), 3. Aportes de Energía - Ríos (%), 4. Aportes de energía - SIN (GWh), 5. Generación Neta Energía - Recurso (KWh), 6. Volumen útil de energía - Embalse (KWh), and 7. Disponibilidad real energía - Recurso (KWh).

Nota. Fuente: El autor.

Interfaz 4 - Visualización de reporte: Se accede a esta al hacer clic sobre un reporte de los disponibles en una de las secciones accedidas en la pantalla de inicio. Al hacer clic sobre el reporte se abre una sección desplazable que nos mostrara el reporte seleccionado. En esta pantalla vemos el reporte estadístico seleccionado de acuerdo con el rango de fecha elegido, aquí se muestra la gráfica generada para el reporte con la selección de los posibles filtros

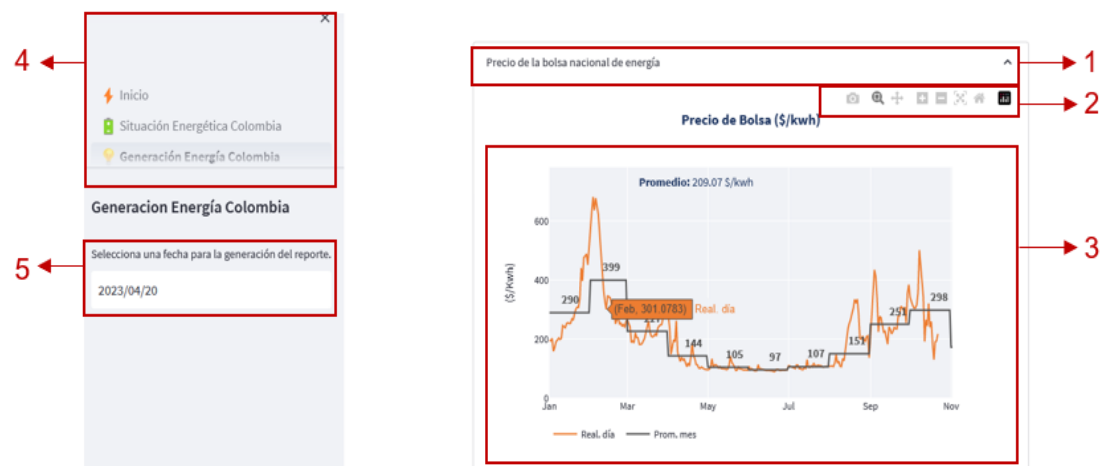
a aplicar, una tabla que muestra los datos que se representaron en la gráfica y un botón al finalizar la sección para realizar la descarga del reporte en cuestión.

En la gráfica, al pasar el mouse sobre la gráfica encontramos se visualizan en una mini ventana los valores para ese punto, en la esquina superior derecha encontramos herramientas para generar la descarga de la imagen en formato .png, aplicación de zoom y gestión de los ejes.

En esta pantalla se conservan los menús laterales para cambio de sección y el componente de selección de rango de fecha, el cual al hacer interacción con el actualiza en tiempo real la visualización de las gráficas de acuerdo con la fecha seleccionada. Se muestra la gráfica generada con las herramientas de esta para su descarga y la interacción del mouse sobre esta.

Figura 19

Interfaz 4 – Visualización de reporte



Nota. Fuente: El autor.

Tabla 21

Especificación secciones interfaz 4

Sección	Contenido
1	Título del reporte
2	Herramientas de interacción del grafico del reporte.
3	Lista de los reportes estadísticos disponibles para ser accedidos.

4	Menú de las secciones a las que es posible acceder.
5	Calendario para selección de la fecha desde la que se desea generar el reporte.

Figura 20

Interfaz 5 - Visualización de tabla de datos del grafico del reporte y enlace a descarga de estos

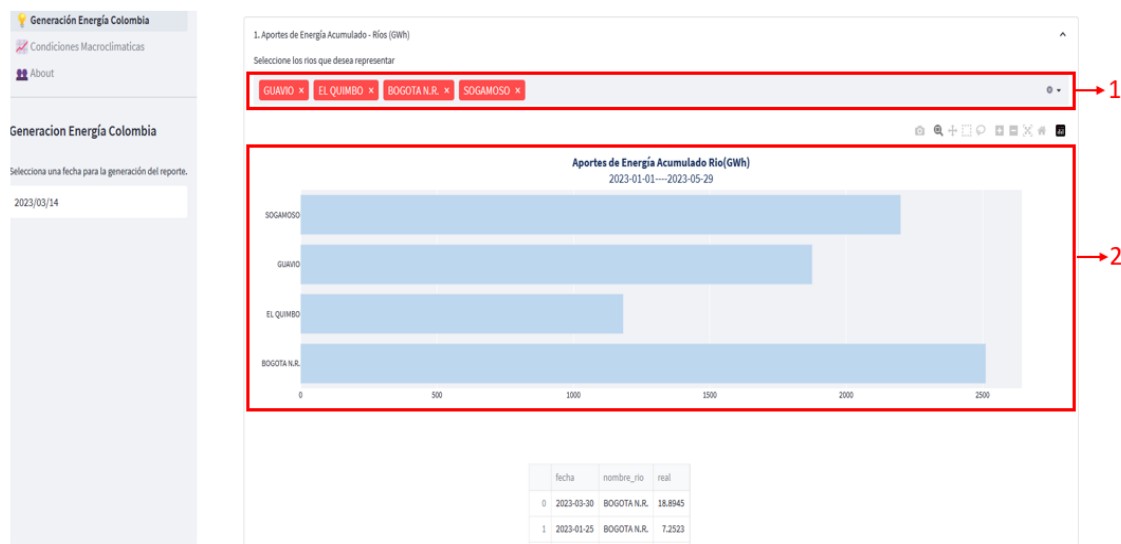


Nota. Fuente: El autor.

Interfaz 6 - Visualización de la selección de filtros para aplicar: Se accede a esta al hacer clic sobre un reporte de los disponibles en una de las secciones accedidas en la pantalla de inicio. Al hacer clic sobre el reporte se abre una sección desplazable que nos mostrara el reporte seleccionado. Una vez dentro del reporte, dependiendo del reporte este nos permitirá filtrar entre los diferentes ejes (ríos, embalses, recursos) que se quieran agregar en el reporte, esto afectará el gráfico generado el dataframe disponible para descargar.

Figura 21

Interfaz 6 – Visualización de la selección de filtros para aplicar.



Nota. Fuente: El autor.

Tabla 22

Especificación secciones interfaz 6

Sección	Contenido
1	Selector de valores (ejes) a incluir en el reporte.
2	Herramientas de interacción del grafico del reporte.

7. Pruebas funcionales de validación

Las pruebas de funcionalidad o usabilidad se realizan para validar la calidad del software e identificar las posibles falencias que presenta en su ejecución. Estas son realizadas haciendo la validación de los supuestos asignados en el diseño y requerimientos del software, verificando así

todas sus que todas las funciones asignadas sean ejecutas como se esperan y con una adecuada implementación.

En la tabla N°23 se puede ver las características del equipo en el que se llevó a cabo las pruebas.

Tabla 23

Especificaciones equipo de pruebas

Sistema Operativo	Windows 11 Home
Tipo de Sistema	64 bits
Procesador	Intel Core i5-8300H @ 2.30GHz
RAM	8,00 GB
Navegador	Mozilla Firefox v.112.0.1

En la Tabla N°24 podemos ver los datos de las pruebas realizada para la consulta y extracción de datos de la API para su posterior mapeo y almacenamiento en la base de datos. Esta prueba se basa en el método de caja blanca, el cual se enfoca en validar las funciones internas del software¹.

Tabla 24

Caso de prueba 1 - Consulta y almacenamiento en base de datos de una métrica deseada a partir de consulta a API.

Código	CP 001
Caso de prueba	Consulta y almacenamiento en base de datos de una métrica deseada a partir de consulta a API.
Descripción	Se probará la correcta descarga y posterior almacenamiento en base de datos para una métrica seleccionada a partir de la solicitud a la API de XM.
Prerrequisitos	<ul style="list-style-type: none">- Pausar la descarga automatizada de la métrica durante al menos los últimos 3 días.- Verificar que en la base de datos no haya datos para la métrica deseada en los últimos 3 días

¹ (Wikipedia.org, 2023)

Pasos	<ul style="list-style-type: none"> - Elegir una de las métricas donde su fuente de descarga sea por vía API. - Verificar que no haya datos para esta métrica en la base de datos para al menos los últimos 3 días. - Ingresar a la implementación y ejecutarla solo para la métrica seleccionada. - Verificar en la base de datos que se haya cargado datos para la métrica elegida hasta la fecha actual.
Resultado esperado	- La tabla de la métrica en la base de datos almaceno los datos solicitados a la API correctamente para hasta la fecha actual.
Resultado obtenido	Ok

En la Tabla N°25 podemos ver los datos de las pruebas realizada para la consulta y extracción de datos el enlace de descarga directo de la página para su posterior mapeo y almacenamiento en la base de datos.

Tabla 25

Caso de prueba 2 - Consulta y almacenamiento en base de datos de una métrica deseada a partir de un enlace dinámico de descarga.

Código	CP 002
Caso de prueba	Consulta y almacenamiento en base de datos de una métrica deseada a partir de un enlace dinámico de descarga.
Descripción	Se probará la correcta descarga y posterior almacenamiento en base de datos para una métrica seleccionada a partir de un enlace de descarga dinámico de la página web de XM.
Prerrequisitos	<ul style="list-style-type: none"> - Pausar la descarga automatizada de la métrica durante al menos los últimos 3 días. - Verificar que en la base de datos no haya datos para la métrica deseada en los últimos 3 días
Pasos	<ul style="list-style-type: none"> - Elegir una de las métricas donde su fuente de descarga sea por vía enlace de descarga directo. - Verificar que no haya datos para esta métrica en la base de datos para al menos los últimos 3 días. - Ingresar a la implementación y ejecutarla solo para la métrica seleccionada. - Verificar en la base de datos que se haya cargado datos para la métrica elegida hasta la fecha actual.

Resultado esperado	- La tabla de la métrica en la base de datos almaceno los datos solicitados a la API correctamente para hasta la fecha actual.
Resultado obtenido	Ok

A continuación, podemos ver los datos correspondientes al plan de pruebas para la interacción del usuario con el sistema, para generar las visualizaciones de los reportes de las métricas y realizar la descarga de los datos. Este tipo de prueba es de caja negra pues está evaluando la funcionalidad exterior del software. Este tipo de prueba es de caja negra pues está evaluando la funcionalidad exterior del software.

Tabla 26

Plan de pruebas - Interacción del usuario con la herramienta

Plan de pruebas: Interacción del usuario con la herramienta			
Actividad	Usuario	Software	Resultados
1. Desplegar la sección de Reportes de “Generación Energía Colombia”	Clic en sección	Desplegar lista de reportes seleccionables.	Visualización de la lista de los reportes disponibles para generar.
2. Desplegar el contenido de un reporte	Clic en un reporte de la lista	Genera reporte seleccionado con un rango de fecha predeterminado	Visualización de los filtros, la gráfica, la tabla de datos y la opción de descarga del reporte en cuestión.
3. Desplegar contenido de un reporte para una fecha deseada	Clic en reporte de la lista	Genera reporte seleccionado con un rango la fecha establecida por la fecha inicial seleccionada. (El rango de fecha será: fecha seleccionada hasta el presente día).	Visualización de los filtros, la gráfica, la tabla de datos y la opción de descarga del reporte seleccionado para el rango de fecha elegido.
	Clic en el widget del calendario “seleccione fecha para la generación del reporte”		
	Interacción con el widget del calendario y selección de		

	fecha inicial para la generación del reporte.		
4. Obtener descarga de los datos generados	Clic en botón descarga en el reporte	Exporta datos	Se obtiene un archivo .csv de los datos a partir de los cuales el reporte se generó.

A continuación, se describen algunos casos de prueba para la interacción del usuario con las diferentes funcionalidades exteriores del software:

Tabla 27

Caso de uso 3 - Generación de un reporte estadístico de una métrica seleccionada para el año presente.

Código	CP 003
Caso de prueba	Generación de un reporte estadístico de una métrica seleccionada para el año presente.
Descripción	Se probará la correcta generación de un reporte elegido de acuerdo con la fecha actual en la que se solicite (01-01-2023).
Prerrequisitos	- El usuario ha accedido al sistema correctamente.
Pasos	- Ingresar a la sección de Reportes de Generación de Energía - Seleccionar un reporte de la lista generada. - Seleccionar en el componente calendario la fecha deseada (01-01-2023). O se puede no seleccionar ninguna fecha, ya que la fecha predeterminada será el primer día del año del año vigente.
Resultado esperado	El reporte solicitado se muestra correctamente de acuerdo con las unidades que este representa, está en el rango de fecha esperado, los ejes que representa tienen valores válidos y consistente.
Resultado obtenido	Ok

Tabla 28

Caso de uso 4 - Generación de un reporte estadístico de una métrica seleccionada desde una fecha específica al presente

Código	CP 004
Caso de prueba	Generación de un reporte estadístico de una métrica seleccionada para desde fecha específica al presente.
Descripción	Se probará la correcta generación de un reporte elegido de acuerdo con una fecha seleccionada por el usuario en el sistema.
Prerrequisitos	- El usuario ha accedido al sistema correctamente.
Pasos	<ul style="list-style-type: none"> - Ingresar a la sección de Reportes de Generación de Energía - Seleccionar un reporte de la lista generada. - Seleccionar una fecha en el widget del calendario en el menú lateral. - Verificar que el reporte solicitado se muestra para un intervalo de tiempo, desde la fecha seleccionada hasta el día presente (puede ser el día anterior, de acuerdo con la hora de consulta)
Resultado esperado	El reporte solicitado se muestra correctamente de acuerdo con las unidades que este representa, está en el rango de fecha esperado, los ejes que representa tienen valores válidos y consistente.
Resultado obtenido	Ok

Tabla 29

Caso de uso 5 - Generación de un archivo de descarga a partir de la generación de un reporte estadístico de una métrica seleccionada

Código	CP 005
Caso de prueba	Generación de un archivo de descarga a partir de la generación de un reporte estadístico de una métrica seleccionada.
Descripción	Se probará la correcta generación de un archivo de descarga para un reporte elegido de acuerdo con una fecha seleccionada por el usuario en el sistema.
Prerrequisitos	- El usuario ha accedido al sistema correctamente.
Pasos	<ul style="list-style-type: none"> - Ingresar a la sección de Reportes de Generación de Energía - Seleccionar un reporte de la lista generada. - Seleccionar una fecha en el widget del calendario en el menú lateral (fecha inicial del reporte).

	<ul style="list-style-type: none">- Verificar que el reporte solicitado se muestra en el intervalo de fecha deseado.- Solicitar la descarga al hacer clic en el botón “Descargar Dataframe” ubicado al final del reporte.- Verificar que se obtengo un archivo .csv con los datos del reporte generado en el software.
Resultado esperado	El reporte solicitado se muestra correctamente de acuerdo con las unidades que este representa, está en el rango de fecha esperado, los ejes que representa tienen valores válidos y consistente.
Resultado obtenido	Ok

8. Conclusiones

Los objetivos proyectados se han cumplido, se ha creado una herramienta software que permite la consulta, visualización y descarga de reportes actualizados sobre métricas hidrológicas de distintos recursos hídricos en Colombia.

La metodología de prototipado evolutivo para realización del proyecto permitió realizar cada fase con su respectiva corrección y retroalimentación para así ir alineándose a los requerimientos previstos para llegar posteriormente al producto final.

Se logro diseñar un modelo de datos capaz de soportar la descarga automatizada de datos desde distintas fuentes modos y fuentes de información como lo es las consultas API y la descarga desde enlaces web dinámicos.

El sistema mantiene la base de datos actualizada gracias a la implementación de la automatización periódica de descarga de los datos.

Con el desarrollo de la implementación se le facilita al usuario final el entendimiento del comportamiento de las principales variables involucradas en la generación de energía hidroeléctrica en Colombia gracias a la fácil generación de reportes de las métricas de su preferencia.

Las pruebas realizadas permitieron verificar y modificar el software, para así tener un sistema eficiente que cumpla cumpliera con los requerimientos previstos.

9. Recomendaciones

Implementar un mayor número de métricas para la generación de los reportes para tener un contexto aún más detallado de las variables que se desean analizar.

Usar un programa especializado para la ejecución programada de la implementación de recolección de datos, esto con el fin de poder parametrizar los ítems de ejecución y poder gestionar de una mejor manera los posibles errores que se puedan generar.

En el aplicativo de usuario adicionar más categorías y subcategorías que permitan al usuario elegir el reporte deseado en base a estas, por ejemplo la selección de reportes de acuerdo con la zona del país, región hidrológica, etc.

Referencias Bibliográficas

Apache Airflow. (17 de Julio de 2022). Introduction to Apache Airflow.
<https://airflow.apache.org/>.

Campos Aranda, D.F. (1998). Procesos del ciclo hidrológico. Universidad Autónoma de San Luis de Potosí. Mexico.

Carrion, J. (18 de Julio de 2022). Streamlit vs Dash: Como crear tu propio dashboard interactivo con Python. Visiondefunnel-blog.
<https://visiondefunnel.com/2021/01/17/dash-streamlit-python/>

Dash (17 de Julio de 2022). Introduction to Dash.. <https://dash.plotly.com/introduction>.

Enel-Emgesa. (14 de Julio de 2018). Centrales de generación.
<https://www.enel.com.co/es/conoce-enel/enel-generacion/centrales-electrica>.

García, C. (2016). Guía práctica para la publicación de datos abiertos usando APIs. En datos.gob.es.

García Candil, A. (2018). Git – Como gestionar y cuidar nuestro código. [Emiliofunciona.io](https://emiliofunciona.io).

Ionox (1 de Julio de 2022). PostgreSQL: EL gestor de bases de datos a fondo.
<https://www.ionos.mx/digitalguide/servidores/know-how/postgresql>

LUCA. (17 de Julio de 2020). ¿Qué es Python?.
<https://web.archive.org/web/20200224120525/https://luca-d3.com/es/data-speaks/diccionario-tecnologico/python-lenguaje>.

Mahiques Zaragoza, J.L. (2013). The meteorology: Basic concepts to all. 3 ciencias, Revista de investigación. España.

- Palacios Sierra, R.A. (2013). Inventario documentado de represas de Colombia. Universidad Militar Nueva Granada. Bogotá D.C.
- Pardo, D. (2019). ¿Para qué sirve una API? Pandora FMS.
<https://web.archive.org/web/20190215165536/https://blog.pandorafms.org/es/para-que-sirve-una-api/>.
- Pearson Addisson, W. (2006). Ingeniería del software. Séptima edición. Sommerville. España.
- PostgreSQL. (17 de Julio de 2022). PostgreSQL Global Development Group. Recuperado el 17 de Julio de 2022 de: <https://www.postgresql.org/>.
- Pressman, R. (2002) Ingeniería de software. Un enfoque practico, quinta edición, España.
- Sommerville, I. (2002) ingeniería de software, Addison-Wesley, 6ta edición. México.
- Silberschatz, A. y Korth, F.H y Sudarshan, S. (2002). Fundamentos de bases de datos. Cuarta edición. Revista técnica Luis Grau Fernandez, Universidad Nacional de Educación a distancia. McGRaw-Hill España.
- UPME. (10 de Agosto de 2022). ¿Quiénes somos?.
<https://www1.upme.gov.co/Entornoinstitucional/Paginas/quienes-somos.aspx>.
- Uriarte, J.M. (2020). Meteorología. En Caracteristicas.co.
<https://www.caracteristicas.co/meteorologia/>.
- Veal, W. R. y Cohen, R.A. (2011). Project Earth Science: Metereology, Revised 2nd Edition. Arlington, VA: Ntional Science Teachers Association.
- XM. (14 de Julio de 2022). Hidrología, Reservas energía de los principales embalses de Colombia. <http://ido.xm.com.co/ido/SitePages/hidrologia.aspx?q=reservas>.
- XM. (10 de Agosto de 2022). Nuestra empresa. <https://www.xm.com.co/nuestra-empresa/nosotros/quienes-somos>