**SOLUCIÓN TECNOLÓGICA PARA LA OPTIMIZACIÓN EN LA GESTIÓN DE RECURSOS HÍDRICOS EN EL PERÚ APLICANDO BUSINESS INTELLIGENCE**

**Documento de Diseño de Arquitectura**

Versión 3.0

**Control de Versiones**

**Lima, 03 de abril del 2025**

**ÍNDICE**

[**1. Introducción 3**](#_heading=)

[**2. Objetivo 3**](#_heading=)

[**3. Visión General de la Arquitectura 3**](#_heading=)

[a. Diagrama general de arquitectura 3](#_heading=)

[b. Descripción de componentes principales 4](#_heading=)

[**4. Arquitectura Física 6**](#_heading=)

[**5. Arquitectura Lógica 7**](#_heading=)

[**6. Arquitectura C4 8**](#_heading=)

[a. Diagrama de Contexto 8](#_heading=h.5inl6ecs5js8)

[b. Diagrama de Contenedores 9](#_heading=h.eetvw73bof8k)

[c. Diagrama de Componentes 10](#_heading=h.7j0ho6ihzm9s)

[**7. Fuente de Datos 11**](#_heading=)

[a. Sistemas fuente 11](#_heading=)

[**8. Procesos ETL 12**](#_heading=)

[a. Herramientas utilizadas 12](#_heading=)

[b. Flujo de trabajo del ETL 13](#_heading=h.f414qfii7y3k)

[**9. Almacenamiento de Datos 13**](#_heading=)

[a. Data Warehouse 13](#_heading=)

# **Introducción**

El presente documento describe la arquitectura técnica de la tesis titulada Solución tecnológica para la optimización en la gestión de recursos hídricos en el Perú aplicando business intelligence, cuyo propósito es transformar datos dispersos en información estructurada, accesible y útil para la toma de decisiones y modelos predictivos. Este diseño técnico establece los componentes clave, herramientas y flujos de trabajo necesarios para garantizar una solución robusta.

El enfoque se centra en integrar múltiples fuentes de datos, procesar y transformar esa información mediante procesos automatizados, almacenarla eficientemente, y finalmente presentarla a través de herramientas de análisis visual.

# **Objetivo**

Definir la arquitectura técnica que sustentará la solución de Business Intelligence, detallando los componentes tecnológicos, herramientas, procesos y buenas prácticas necesarios para capturar, transformar, almacenar, analizar y realizar modelos predictivos a partir de los datos, asegurando su calidad y disponibilidad.

# **Visión General de la Arquitectura**

En este primer módulo, se observará en detalle el flujo de trabajo del aporte.

## **Diagrama general de arquitectura**

La Figura 1 presenta el diagrama general de la arquitectura técnica propuesta. Este esquema ilustra el flujo completo de la interacción del usuario con el sistema a un nivel de Frontend, Backend, herramientas, tecnologías y su correlación entre sí.

**Figura 1.**

*Diagrama de la arquitectura del aporte.*

*Diagrama

El contenido generado por IA puede ser incorrecto.*

Elaboración propia.

## **Descripción de componentes principales**

A continuación, se describen los componentes principales que conforman la arquitectura técnica de la propuesta.

* **Frontend:** El frontend representa la interfaz de usuario con la que los usuarios finales interactúan para consultar información, visualizar datos y comunicarse con el sistema.
  + **Tecnologías empleadas:**
    - React.js: Framework principal para construir interfaces dinámicas y componentes reutilizables.
    - Vite: Herramienta moderna de desarrollo que optimiza la compilación del proyecto.
    - Node.js: Actúa como puente entre frontend y backend en tiempo de ejecución.
  + **Módulos funcionales:**
    - Página principal: Vista de bienvenida y navegación general.
    - Proceso de Gastos: Panel para visualizar indicadores presupuestales.
    - Proceso de Agua Potable y Alcantarillado: Módulo para mostrar cobertura y calidad del servicio.
    - ChatBot: Asistente conversacional integrado para consultas automatizadas y guía en el uso de la plataforma.
* **Backend:** El backend administra la lógica de negocio, procesamiento de lenguaje natural, conexiones con bases de datos y manejo de APIs.
  + Componentes clave:
    - Langchain: Orquestador de flujos conversacionales. El cual se conecta con modelos de lenguaje (LLM) y bases vectoriales.
    - OLlama + Deepseek: Modelos LLM que alimentan el chatbot para interpretar preguntas complejas. Además, ofrece comprensión semántica y contextual de las consultas.
  + ChromaDB: Base de datos vectorial que permite realizar búsquedas semánticas sobre documentos o fragmentos informativos cargados en el sistema.
  + API Manager: Interfaz de comunicación entre frontend y backend.
  + Chatbot API: Para interactuar con el asistente virtual.
  + SQL API: Para acceder al Data Warehouse y devolver resultados agregados o consultas específicas.
  + Data Warehouse: Base de datos estructurada que almacena toda la información histórica transformada. Alimentado por procesos ETL y preparado para el análisis con herramientas BI.

# **Arquitectura Física**

La arquitectura física del sistema representa la distribución real de los componentes tecnológicos sobre los cuales se ejecuta la solución. Esta configuración permite visualizar cómo interactúan entre sí los distintos elementos de software, asegurando un flujo adecuado de datos y procesos desde el usuario hasta el motor analítico.

La Figura 2 muestra el Diagrama de Arquitectura Física del Sistema, que se encuentra dividido en tres bloques principales: el usuario final, el frontend y el backend.

**Figura 2.**

*Diagrama de Arquitectura Física del Sistema.*

Interfaz de usuario gráfica, Diagrama, Aplicación

El contenido generado por IA puede ser incorrecto.

Elaboración propia.

# **Arquitectura Lógica**

La arquitectura lógica del sistema define la organización estructural de los componentes funcionales que forman parte de la solución tecnológica. A diferencia de la arquitectura física, esta vista se centra en la distribución de responsabilidades y el flujo lógico de la información entre capas funcionales, asegurando una separación clara de roles y facilitando el mantenimiento y escalabilidad del sistema.

La Figura 3 presenta el Diagrama de Arquitectura Lógica, estructurado en cuatro capas principales.

**Figura 3.**

*Diagrama de Arquitectura Lógica del Sistema.*

*Diagrama

El contenido generado por IA puede ser incorrecto.*

Elaboración propia.

# **Arquitectura C4**

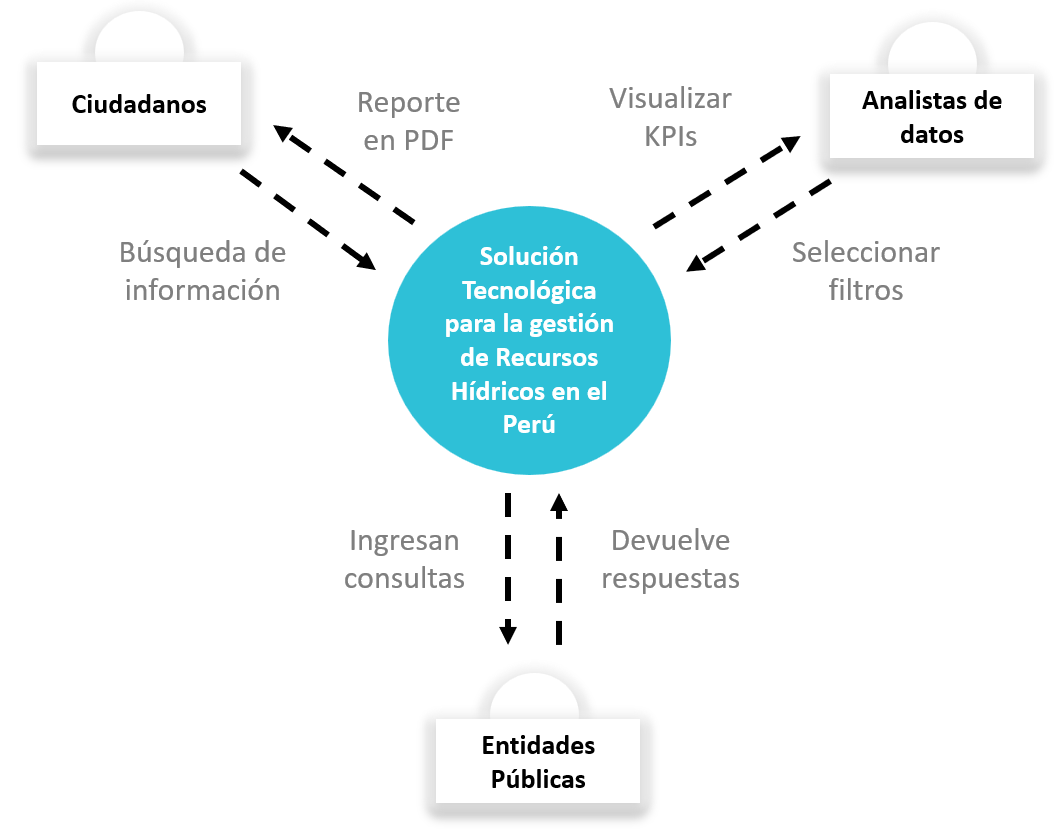
Se utilizó el modelo de arquitectura C4 para un mejor entendimiento de la estructura del sistema propuesto. En base a ello, esta arquitectura se divide en tres niveles de abstracción. Los cuales se detallan a continuación.

## ***Diagrama de Contexto***

La elaboración de este primer nivel se apoyó en la relación de los diferentes usuarios con el programa y funciones principales (ver Figura 4).

**Figura 4.**

*Diagrama de Contexto de la solución tecnológica propuesta.*

**

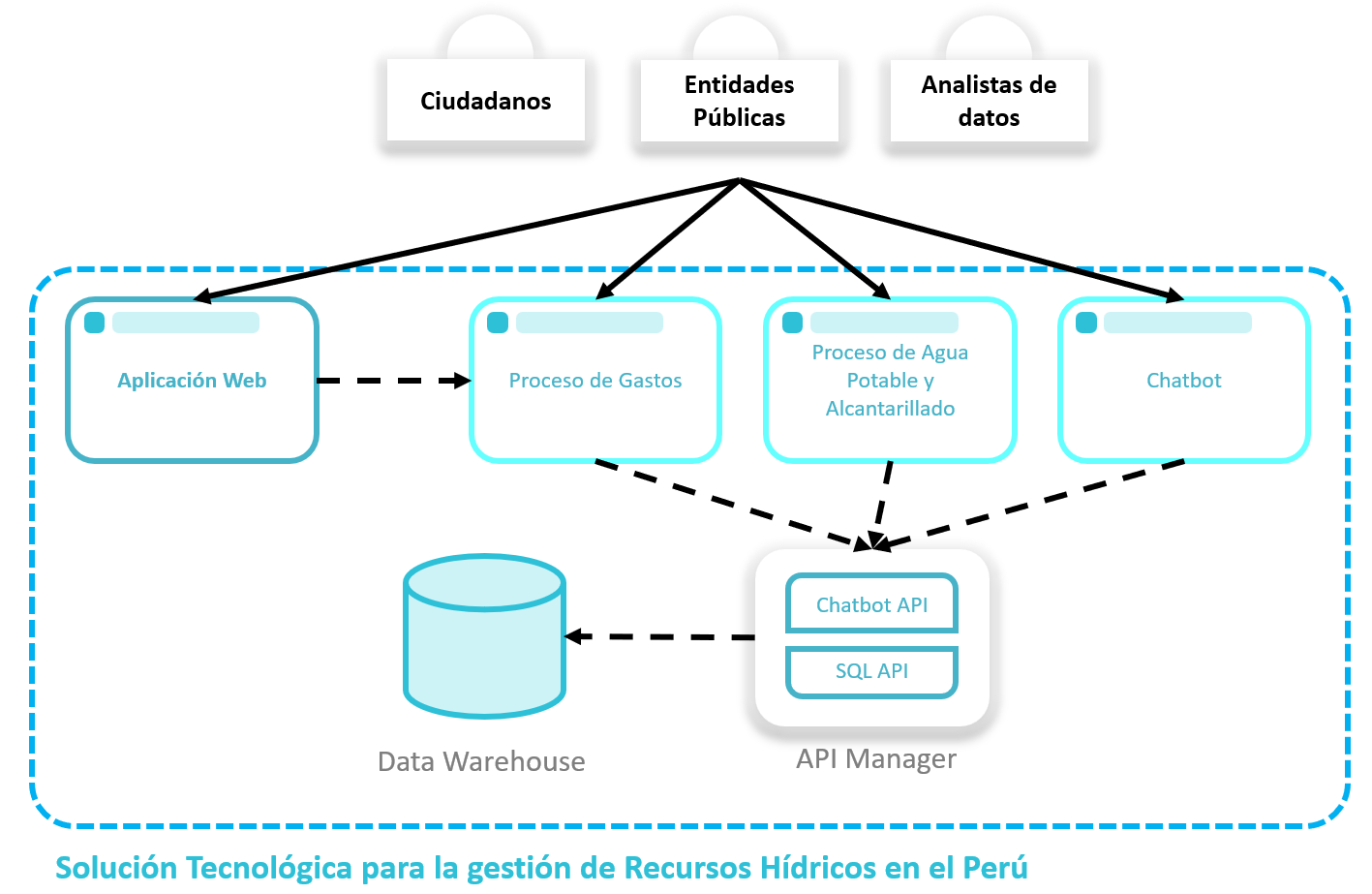
Elaboración propia.

## **Diagrama de Contenedores**

La realización de este segundo nivel de arquitectura C4, se basó en explicar de manera lógica y técnica como interactúan los componentes principales del sistema (ver Figura 5).

**Figura 5.**

*Diagrama de Contenedores de la solución tecnológica propuesta.*



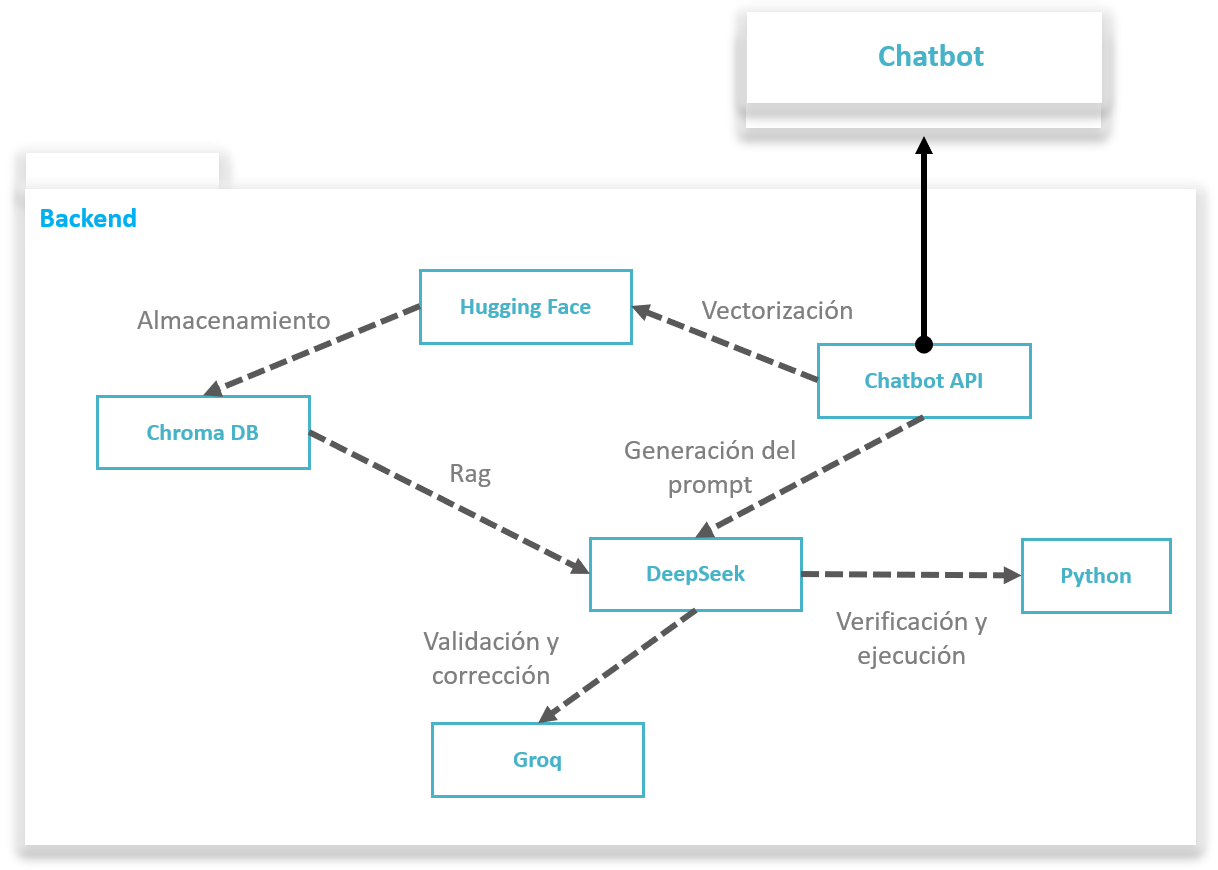
Elaboración propia.

## **Diagrama de Componentes**

La preparación de este tercer nivel se fundó en el detalle del componente de LLM (ver Figura 6).

**Figura 6.**

*Diagrama de Componentes de la solución tecnológica propuesta.*

**

Elaboración propia.

# **Fuente de Datos**

Este módulo, se enfoca en detallar y enlistar las distintas fuentes de datos el cual hace uso el aporte.

## **Sistemas fuente**

A continuación, se presentan los distintos sistemas fuente.

**Tabla 1.**

*Categorización de sistemas fuente.*

| **Tipo de extracción** | **Nombre del Sistema** | **Nombre del Archivo** | **Descripción** |
| --- | --- | --- | --- |
| Externa | Plataforma Nacional de Datos Abiertos | Gasto Presupuestal de las Entidades de Tratamiento Empresarial - ETES | Información de 2015 al 2023 del Gasto de las Entidades de Tratamiento Empresarial - ETES. |
| Encuesta Nacional de Programas Presupuestales (ENAPRES) - [Instituto Nacional de Estadística e Informática - INEI] | Determina la cobertura de servicios básicos de los diferentes Programas Presupuestales entre los años 2010 - 2023. |
| Sistema de Registro de Información de Área Técnica Municipal sobre los prestadores del ámbito rural - [Superintendencia Nacional de Servicios de Saneamiento - SUNASS] | El sistema permite el registro de información de las Organizaciones Comunales, la forma de organización de la prestación del servicio, la facturación e ingresos por los servicios brindados por el prestador entre los años 2019 - 2023. |
| Interna | - | UBIGEO | El conjunto de datos permite identificar la ubicación de un lugar en Perú. |

Elaboración propia.

# **Procesos ETL**

Este módulo se basa en la implementación de un flujo de trabajo que permite la extracción, transformación y carga de datos provenientes de distintas fuentes hacia el Data Warehouse.

## **Herramientas utilizadas**

El proceso ETL se desarrollará utilizando las siguientes herramientas:

* **SQL Server Management Studio (SSMS):** Utilizado para la administración de la base de datos, ejecución de consultas, monitoreo de procesos y validación de la carga de datos.
* **Microsoft Visual Studio (con Integration Services - SSIS):** Plataforma principal para el desarrollo de paquetes ETL y conexión a múltiples fuentes de datos.
* **Python:** Usado para el tratamiento y unificación de los datos con el fin de obtener un conjunto de datos más estructurado.

## **Flujo de trabajo del ETL**

* **Extracción**
  + Lectura de archivos fijos (CSV, Excel).
* **Transformación**
  + Limpieza de datos (eliminación de nulos, normalización).
  + Conversión de tipos de datos, mapeo de campos, creación de columnas calculadas.
* **Carga**
  + Inserción de los datos transformados en tablas del Data Warehouse alojado en SQL Server.
  + Ejecución controlada mediante paquetes SSIS.

# **Almacenamiento de Datos**

Esta capa representa el núcleo central donde se almacena la información ya procesada, organizada y lista para su análisis.

## **Data Warehouse**

Se implementará utilizando SQL Server, aprovechando sus capacidades de almacenamiento relacional, consultas eficientes y compatibilidad con herramientas analíticas de Microsoft. Este almacén de datos contendrá información histórica consolidada, normalizada y optimizada para consultas analíticas.