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

**Documento de Diseño de Arquitectura**

Versión 1.0

**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. Fuente de Datos 5**](#_heading=)

[a. Sistemas fuente 5](#_heading=)

[**5. Procesos ETL 7**](#_heading=)

[a. Herramientas utilizadas 7](#_heading=)

[b. Flujo de trabajo del ETL 7](#_heading=h.f414qfii7y3k)

[**6. Almacenamiento de Datos 8**](#_heading=)

[a. Data Warehouse 8](#_heading=)

[b. Cubo Tabular 8](#_heading=h.v2ybnmz1jfwh)

# **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**

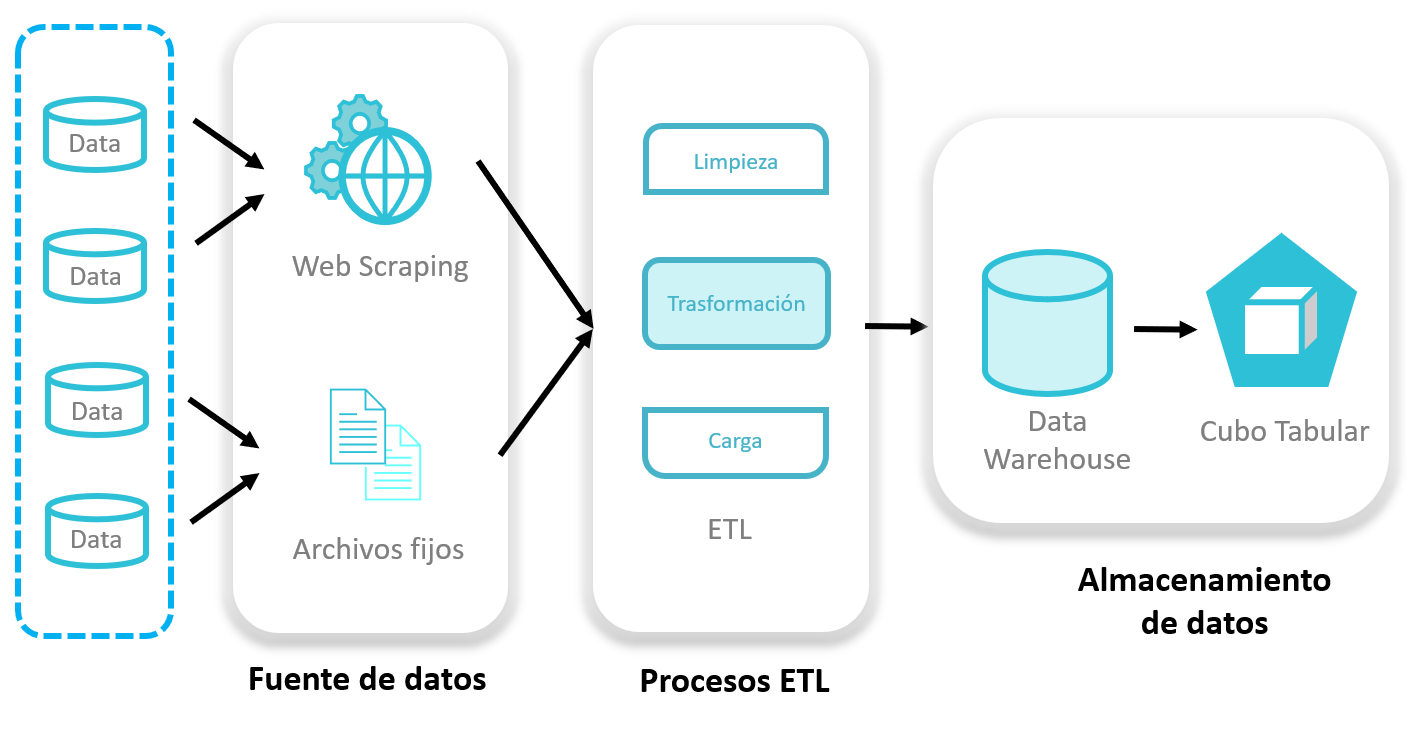
En este primer módulo, se observará en detalle el flujo de trabajo del aporte.

## **Diagrama general de arquitectura**

La Figura 1 presenta el diagrama general de la arquitectura técnica propuesta. Este esquema ilustra el flujo completo de los datos, desde su origen (mediante técnicas de web scraping y archivos fijos), luego procesados a través de un flujo ETL que permite su limpieza, transformación y carga en un Data Warehouse centralizado. Finalmente, la información estructurada es modelada mediante un cubo tabular, que facilitará posteriores análisis multidimensionales.

**Figura 1.**

*Diagrama de la arquitectura del aporte.*



Elaboración propia.

## **Descripción de componentes principales**

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

* **Fuentes de Datos:** Incluyen orígenes heterogéneos como datos extraídos mediante técnicas de web scraping desde sitios web, y archivos fijos estructurados (CSV, Excel, etc.) proporcionados de organizaciones gubernamentales peruanas.
* **Data Warehouse:** Empezando con el proceso de ETL, encargado de extracción, transformación y carga de los datos. Asimismo, el data warehouse representa el repositorio centralizado de datos estructurados. Su diseño se basa en un modelo dimensional que facilita la navegación y análisis de los datos.
* **Cubo Tabular:** La información almacenada en el Data Warehouse es modelada en un cubo tabular, lo que permite la agregación eficiente de datos y su análisis desde múltiples dimensiones.

# **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. |
| Distritos declarados en estado de emergencia por riesgo de desastre asociados al Recurso Hídrico - [Superintendencia Nacional de Servicios de Saneamiento – SUNASS] | Este conjunto de datos monitorea y caracteriza los Decretos Supremos emitidos declarados en estado de emergencia por riesgo de desastre de origen natural relacionado al recurso hídrico entre los años 2017 - 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 | Autoridad Nacional del Agua (ANA) | Distribución y Calidad del agua | Esta metadata sirve para verificar la distribución del agua por departamento y su categorización a nivel de calidad ICARHS entre los años 2015 - 2024. |
| - | UBIGEO | El conjunto de datos permite identificar la ubicación de un lugar en Perú. |

Elaboración propia.

Como se evidencia en la Tabla 1, dos de los archivos son entregados de manera interna, por ende se tomará como archivos fijos.Sumado a los archivos Distritos declarados en estado de emergencia por riesgo de desastre asociados al Recurso Hídrico - [Superintendencia Nacional de Servicios de Saneamiento – SUNASS] y Sistema de Registro de Información de Área Técnica Municipal sobre los prestadores del ámbito rural - [Superintendencia Nacional de Servicios de Saneamiento - SUNASS]. Mientras que los demás, pasarán por un flujo de Web Scraping para su uso.

# **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:

* **Python:** Utilizado para el web scraping y conexión con la base de datos.
* **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.

## **Flujo de trabajo del ETL**

* **Extracción**
  + Lectura de archivos fijos (CSV, Excel).
  + Datos obtenidos mediante web scraping exportados previamente a archivos estructurados.
* **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.

## **Cubo Tabular**

Sobre el Data Warehouse, se construirá un cubo tabular, modelado mediante Analysis Services (SSAS Tabular) en SQL Server. Este cubo permitirá:

* Análisis multidimensional eficiente, organizando los datos en medidas y dimensiones.
* Mejor rendimiento en consultas, ya que los datos son precalculados y optimizados en memoria.
* Flexibilidad en exploración de datos, facilitando la generación de reportes y dashboards en herramientas como Power BI y Excel.
* Seguridad y control de acceso, administrando los permisos a nivel de usuario o rol.