

# MODELO DE ARQUITECTURA

## Empresa Farmacéutica

---

**PROYECTO:** Análisis de la Esperanza de Vida al Nacer para Factibilidad de lanzamiento de multivitamínico a nivel Global

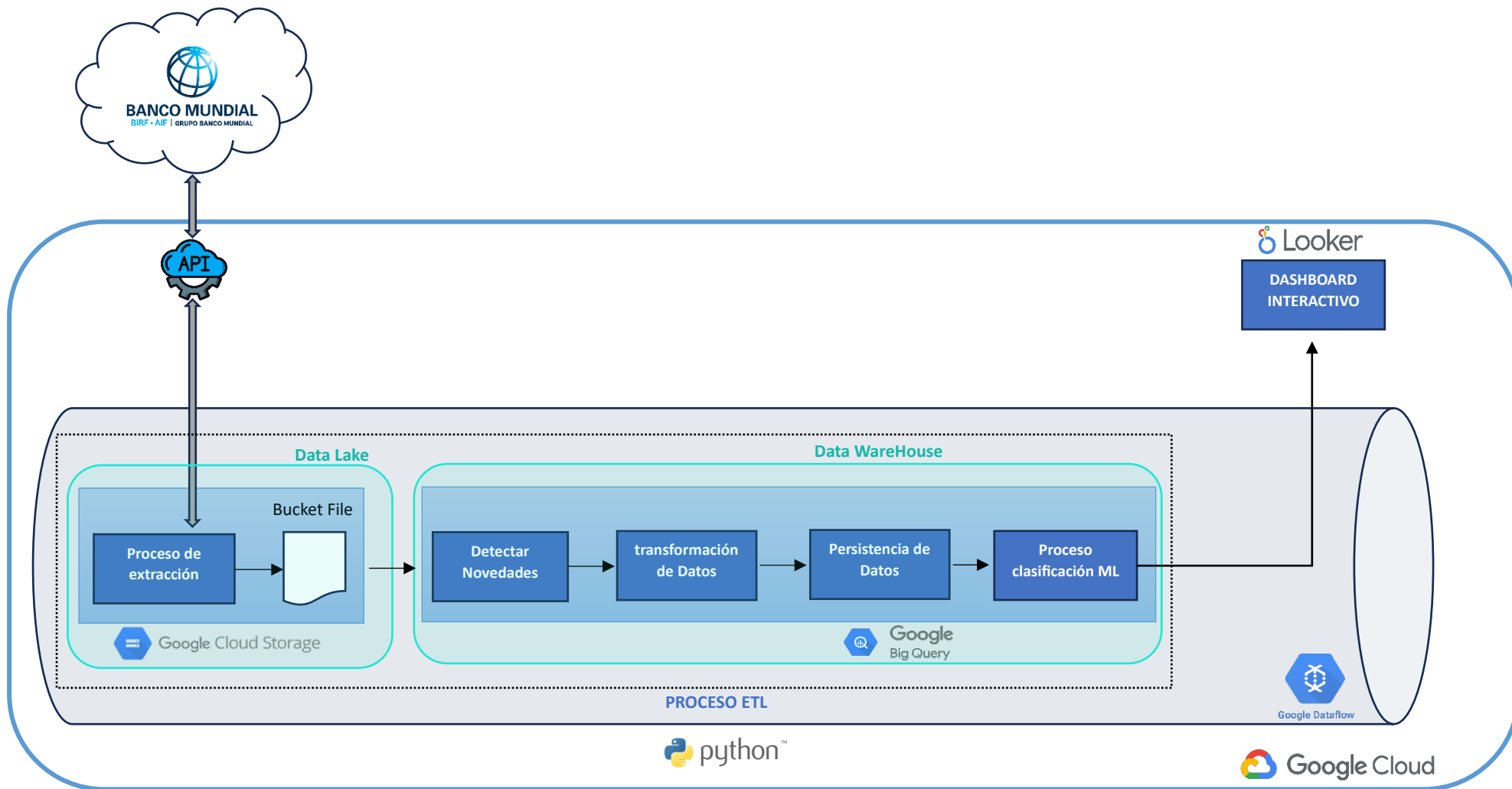
---

## 1. Introducción

Este documento describe la arquitectura tecnológica propuesta para el proyecto de análisis de esperanza de vida y factores determinantes en diferentes regiones del mundo. Este proyecto, encargado por una empresa farmacéutica líder, se centra en recopilar, procesar y analizar datos para respaldar decisiones estratégicas en el lanzamiento de productos.

La arquitectura tecnológica proporciona la infraestructura clave para gestionar datos a gran escala, realizar análisis avanzados y garantizar la disponibilidad de información relevante. Este modelo se enfoca en la eficiencia, escalabilidad y confiabilidad del proceso.

A lo largo de este documento, presentaremos los componentes clave de la arquitectura, incluyendo tecnologías, procesos y la infraestructura en la nube utilizada. El objetivo es proporcionar una plataforma sólida y escalable para la toma de decisiones informadas en la industria farmacéutica.



## 2. Detalle del Stack

1. **Proceso Consumidor de API:** Se creará un proceso que consuma la API del banco mundial y descargue los datos de los factores seleccionados. Este proceso generara unos archivos en formato Bucket para subir a **Google Cloud Storage**, que vendría a ser como nuestro **DATA LAKE**.
2. Para poder persistir los datos en Google Big Query, se ejecutarán los siguientes procesos:
  - a. **Detección de novedades:** Proceso que detectara las novedades a cargar, favoreciendo la carga incremental y evitando la carga completa de los datos.
  - b. **Transformación de Datos:** Se aplicarán las transformaciones necesarias para adecuarnos a las estructuras de datos de nuestro Warehouse.
  - c. **Persistencia en Warehouse:** Utilizaremos Google Big Query como Data Warehouse, en donde persistiremos los datos estructurados y modificados.
- Los puntos 1 y 2 constituyen la fase de ETL, que abarca desde la extracción de los datos hasta su transformación y, finalmente, su almacenamiento en el Warehouse. Sin embargo, para concluir el proceso de ETL, se llevará a cabo la ejecución de nuestro **Modelo de Machine Learning** (Modelo de Clasificación Binaria). Este modelo clasificará las distintas regiones para determinar la viabilidad del lanzamiento del producto, y se reentrenará si es necesario. Todos estos procesos serán encadenados por medio de DAGs haciendo uso de **Google DataFlow**.
3. Por último, se creará un Dashboard Interactivo en **Looker**, haciendo uso de los datos persistidos en el **WareHouse**.

De esta manera, el STACK TECNOLÓGICO a utilizar pertenece en su totalidad a **GOOGLE CLOUD PLATFORM**.

- ❖ Por otra parte, nuestro lenguaje principal de programación será **Python**.