



## TRABAJO PRÁCTICO 02: Procesos ETL

# Introducción

En este trabajo práctico se abordan las cuestiones relacionadas con los procesos de ETL. Se plantean ejercicios y datasets cuyas resoluciones serán realizadas utilizando Pandas en Python y Apache Hop<sup>1</sup>, en el repositorio de la materia se encuentra esta herramienta dockerizada<sup>2</sup>.

# Consignas

Utilizando el dataset de Education Data<sup>3</sup> del World Bank Group se requiere lo siguiente.

1. **CSV de Países con la siguiente estructura:**
  - a. id auto incremental
  - b. Nombre del País
  - c. Código del país
2. **CSV de Preguntas con la siguiente estructura:**
  - a. id auto incremental
  - b. Pregunta (se requiere remover el contenido entre paréntesis)
3. **CSV de Educación primaria en años ordenado por duración de mayor a menor según el año 2023**

Resolver el ejercicio utilizando Apache Hop y realizar el mismo proceso a través de una Notebook de Colab o Jupyter. Comentar los resultados obtenidos en cada una de las aproximaciones y comparar el proceso. Para esto genere un breve informe y envíelo en formato PDF.

---

<sup>1</sup> <https://hop.apache.org/>

<sup>2</sup> <https://github.com/bdm-unlu/2025/tree/main/docker/hop>

<sup>3</sup> <https://data.worldbank.org/topic/education?view=chart>