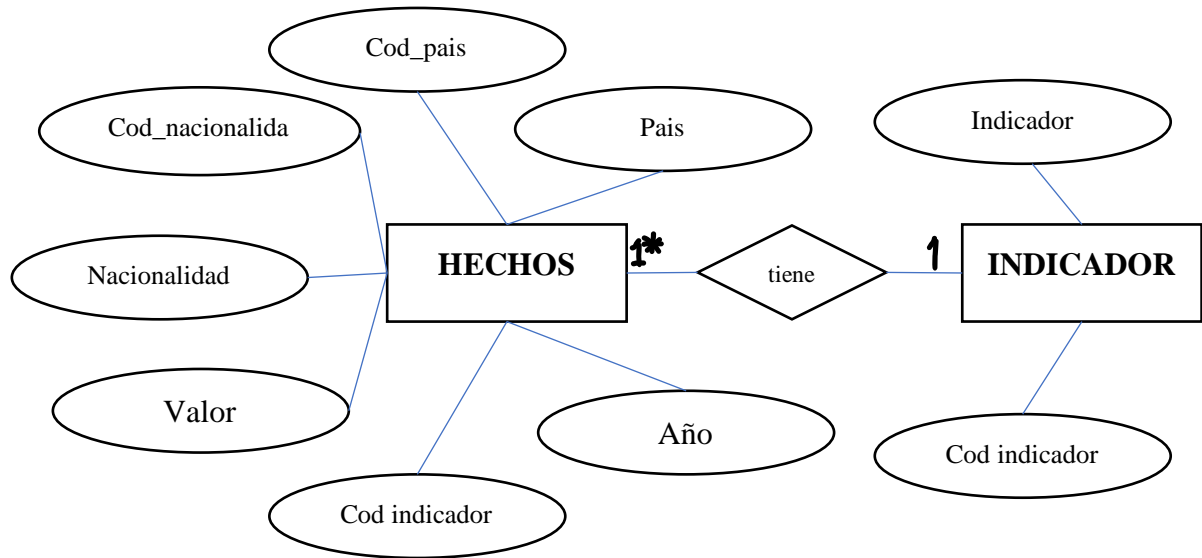
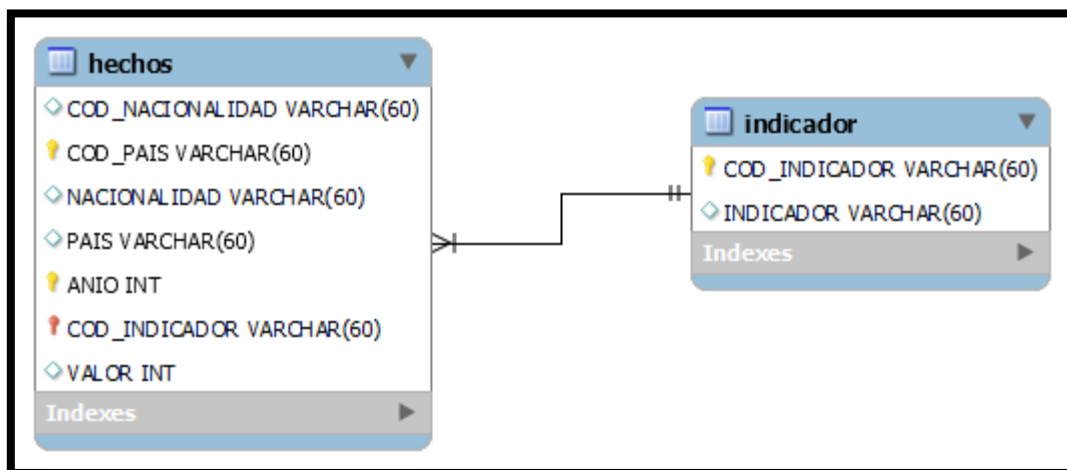


DATA ENGINEERING - SEMANA 2

DIAGRAMA ENTIDAD RELACIÓN.



MODELO ESTRELLA.



DICCIONARIO DE DATOS

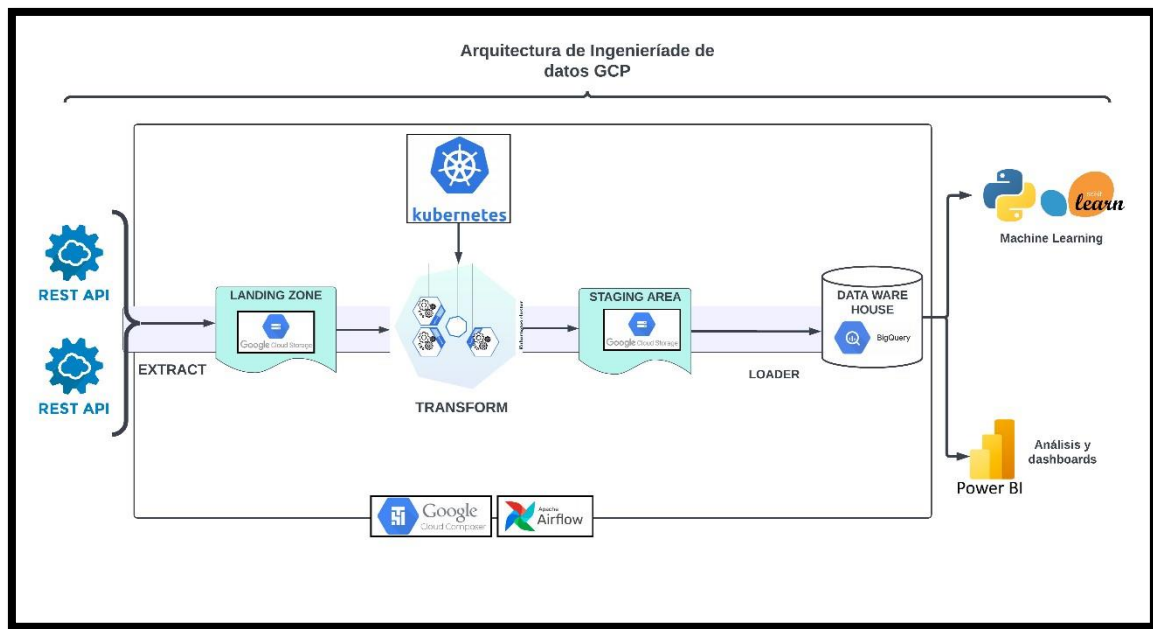
DICCIONARIO DE DATOS				
ENTIDAD	ATRIBUTO	TIPO DE DATO	PK/FK	DESCRIPCION
HECHOS	PAIS	CARÁCTER		NO NULO
	CODIGO_PAIS	CARÁCTER		NO NULO
	ANIO	INT		NO NULO
	VALOR	INT		NO NULO
	CODIGO_INDICADOR	CARÁCTER	LLAVE FORANEA	NO NULO
	CODIGO_NACIONALIDAD	CARÁCTER		
	NACIONALIDAD	CARÁCTER		
INDICADOR	COD_INDICADOR	CARÁCTER	LLAVE PRIMARIA	NO NULO
	INDICADOR	CARÁCTER		NO NULO

WORKFLOW

ARQUITECTURA.

Al tener un área de "landing zone", se puede validar la calidad de los datos recibidos antes de que entren al proceso de ETL. Esto permite detectar y corregir errores o inconsistencias en los datos antes de que se procesen y almacenen en el sistema. Asimismo, al tener un área de "staging", se puede hacer una validación adicional y preparar los datos para su transformación, lo que ayuda a mejorar aún más la calidad de los datos procesados y cargados en el sistema final.

Flexibilidad y escalabilidad: Al tener un área de "landing zone" y un área de "staging", se puede diseñar un proceso de ETL escalable y flexible. Esto permite agregar nuevas fuentes de datos y transformaciones sin interrumpir el flujo de datos existente. Además, se pueden realizar pruebas y experimentos en el área de "staging" sin afectar los datos del sistema final, lo que facilita el desarrollo y mejora de nuevos procesos de ETL.



FUENTE DE DATOS.

Los datos extraídos son de la página del Banco Mundial, y de la (OCDE) Organización para la Cooperación y el Desarrollo Económicos.

EXTRACCIÓN DE DATOS: Extraemos los datos mediante la API del banco mundial y de la OCDE

TRANSFORMACIÓN DE DATOS

Realizamos una limpieza y transformamos los datos para asegurarnos de que estén estandarizados y sean coherentes con los estándares del Data Ware house.

CARGA DE DATOS

Realizamos la carga de datos con Python y Pandas.

MANTENIMIENTO Y ACTUALIZACIÓN:

garantizamos que el data Ware House esté actualizado, automatizando el proceso con Apache Airflow, generando con Python el proceso de extracción y transformación. Con la finalidad de tener la información disponible para respaldar los procesos de toma de decisiones.