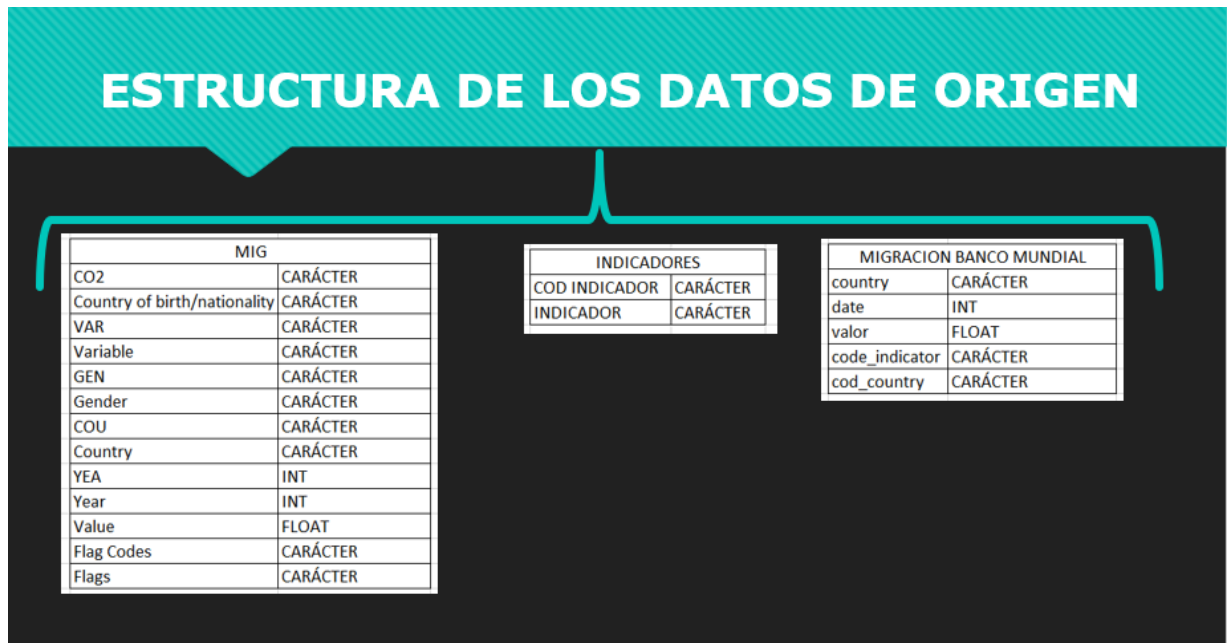


Data Engineering- Semana 2

Diagrama ER

En un principio contamos con las siguientes tres tablas.

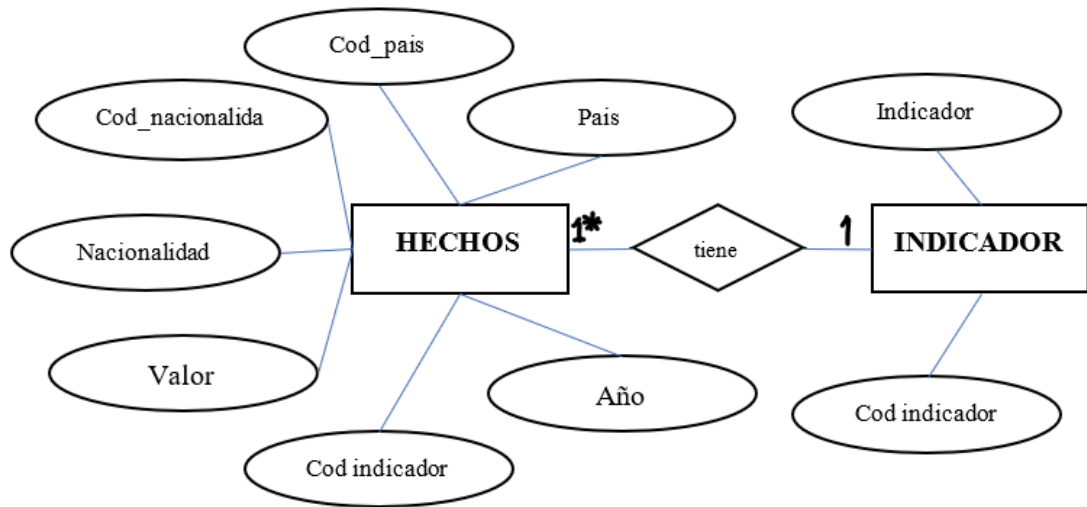


La primera tabla cuenta con los indicadores socio económicos, la segunda contiene el nombre de los indicadores y la tercera trata sobre los países y los años de los respectivos indicadores.

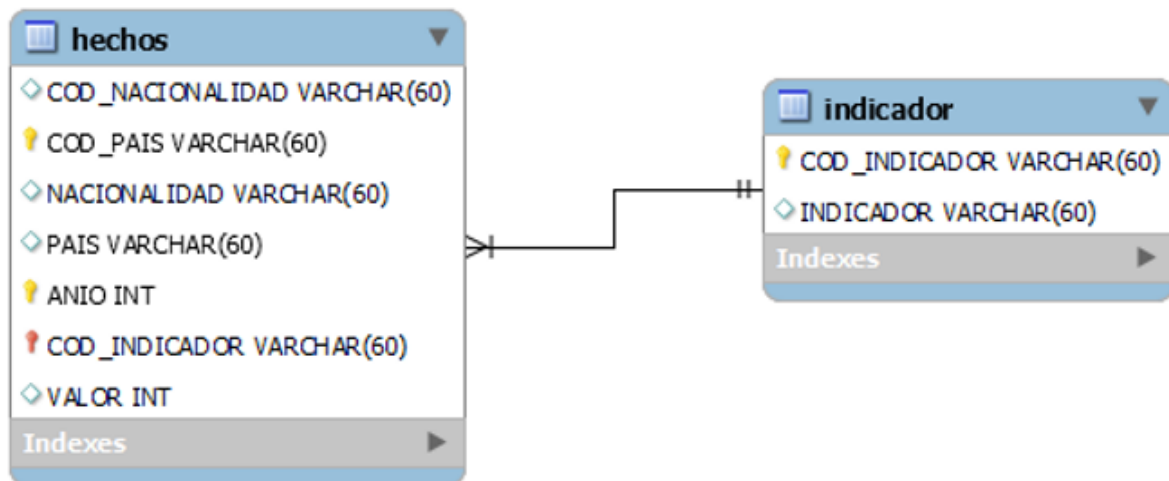
Para el diagrama entidad-relación unimos las tablas y solo contamos con dos tablas, una de hechos en la cual se encuentra los principales indicadores a analizar con sus respectivos valores a lo largo del tiempo y la segunda tabla que es de dimensiones en la cual se encuentran los nombres de dichos indicadores.

La tabla hechos y la tabla indicadores se conectan mediante la columna 'COD_INDICATOR' común en las dos tablas, la cual indica el código del indicador socio económico.

DIAGRAMA ENTIDAD RELACIÓN.



MODELO ESTRELLA



Diccionario de datos

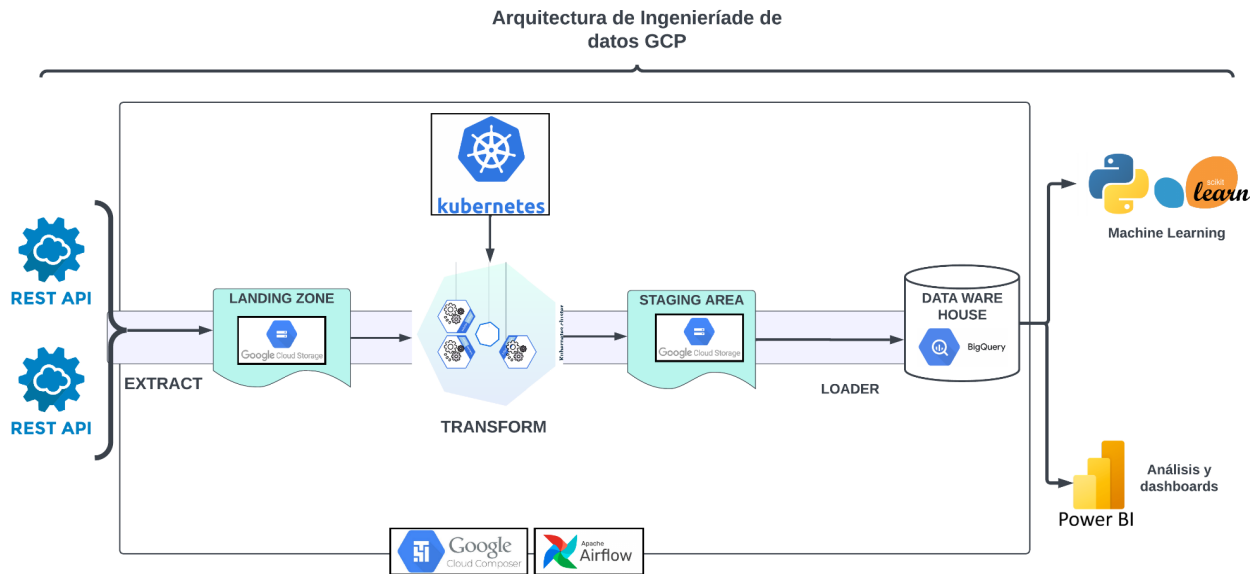
A continuación se presentan las tablas a trabajar y sus respectivas descripciones.

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 tecnologías

- Identificación de fuentes 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
- Transformación de datos: Realizamos una limpieza y transformamos los datos para asegurarnos de que estén estandarizados y sean coherentes con los estándares del datawarehouse.
- Carga de datos: Realizaremos la carga de datos con Python y Pandas.
- Análisis y reportes: Utilizaremos Power Bi para realizar análisis y generar informes para respaldar la toma de decisiones
- Mantenimiento y actualización: garantizamos que el data warehouse 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.

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.