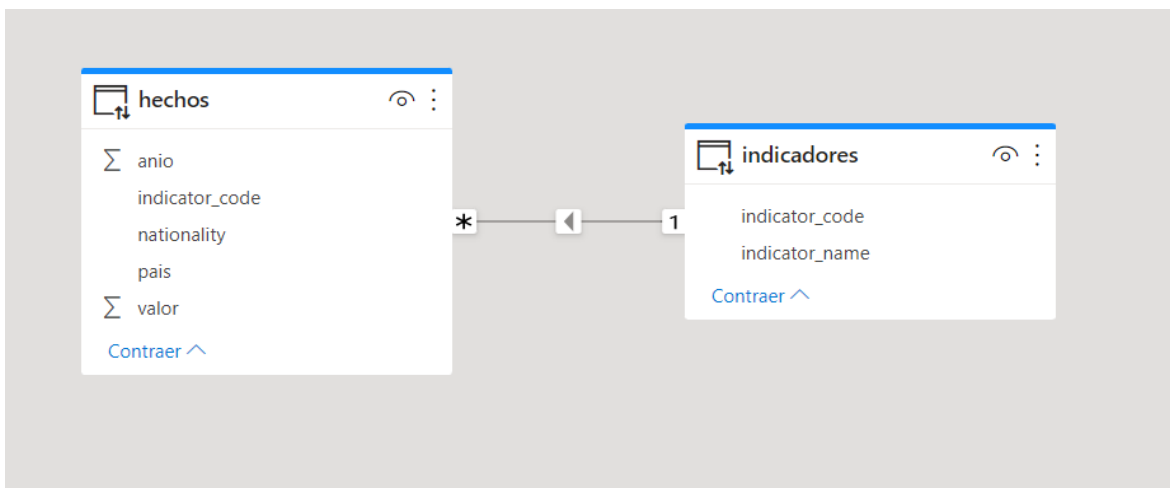


Data Engineering- Semana 2

Diagrama ER

Para el diagrama entidad-relación contamos con dos tablas, una de hechos en la cual se encuentran 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.



Diccionario de datos

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

Tabla de Hechos			
Atributo	Descripción	Condición	Tipo de dato
pais	Nombre del país		varchar(100)
anio	Año del indicador		Int(10)
valor	valor del indicador		float(15)
codigo_indicador	Código asociado al indicador	Foreing Key	varchar(100)

Tabla de dimensiones			
Atributo	Descripción	Condición	Tipo de dato
indicador_code	Código asociado al indicador	Primary key	Varchar(100)
indicador_name	nombre del indicador		Varchar(100)

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 de Ingeniería de datos GCP

