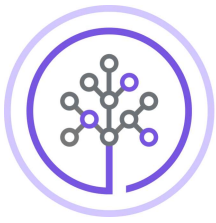


# Practice Project Overview

---



**Skills**  
Network

## Estimated Effort : 5 mins

In this practice project, you will use the skills acquired through the course and create a complete ETL pipeline for accessing data from a website and processing it to meet the requirements.

## Project Scenario

---

An international firm that is looking to expand its business in different countries across the world has recruited you. You have been hired as a junior Data Engineer and are tasked with creating an automated script that can extract the list of all countries in order of their GDPs in billion USDs (rounded to 2 decimal places), as logged by the International Monetary Fund (IMF). Since IMF releases this evaluation twice a year, this code will be used by the organization to extract the information as it is updated.

You can find the required data on [this webpage](#).

The required information needs to be made accessible as a JSON file 'Countries\_by\_GDP.json' as well as a table 'Countries\_by\_GDP' in a database file 'World\_Economies.db' with attributes 'Country' and 'GDP\_USD\_billion.'

Your boss wants you to demonstrate the success of this code by running a query on the database table to display only the entries with more than a 100 billion USD economy. Also, log the entire process of execution in a file named 'etl\_project\_log.txt'.

You must create a Python code 'etl\_project\_gdp.py' that performs all the required tasks.

## Author(s)

---

Abhishek Gagneja