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