Skills
Network

Hands-on Lab : Web Scraping

Estimated time needed: **30 to 45** minutes

Objectives

In this lab you will perform the following:

- Extract information from a given web site
- Write the scraped data into a csv file.

Extract information from the given web site

You will extract the data from the below web site:

In [1]:
```
#this url contains the data you need to scrape
url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/datasets/Programmin
```

The data you need to scrape is the **name of the programming language** and **average annual salary**.
It is a good idea to open the url in your web broswer and study the contents of the web page before you start to scrape.

Import the required libraries

In [2]:
```
from bs4 import BeautifulSoup # module for web scrapping.
import requests  # module for downloading a web page
import pandas as pd # module for dataframes
```

Download the webpage at the url

In [3]:
```
data  = requests.get(url).text
```

Create a soup object

In [4]:
```
soup = BeautifulSoup(data,"html.parser")
```

Scrape the `Language name` and `annual average salary`.

In [24]:
```
#create an empty data frame
my_data = pd.DataFrame(columns=["Language Name", "Annual Average Salary"])

#isolate the body of the table, then loop through each row and find all the column values for each row
for row in soup.find("tbody").find_all("tr"):
    cols = row.find_all('td') # in html, a column is represented by the tag <td>
    language_name = cols[1].string # store the value in column 1 as language_name
    avg_salary = cols[3].string # store the value in column 3 as salary

    #append the data of each row to the table
    my_data = my_data.append({"Language Name":language_name, "Annual Average Salary":avg_salary}, ignore_index=True)

# drop the first row (headers)
my_data=my_data.iloc[1:, :]

my_data
```

Out[24]:

	Language Name	Annual Average Salary
1	Python	$114,383
2	Java	$101,013
3	R	$92,037
4	Javascript	$110,981
5	Swift	$130,801
6	C++	$113,865
7	C#	$88,726
8	PHP	$84,727
9	SQL	$84,793
10	Go	$94,082

Save the scrapped data into a file named *popular-languages.csv*

In [25]:

```
import csv
my_data.to_csv('popular-languages.csv', index=False, header=True)
```

Authors

Ramesh Sannareddy

Other Contributors

Rav Ahuja

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-10-17	0.1	Ramesh Sannareddy	Created initial version of the lab