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.cl
```

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
import requests
import pandas as pd
```

Download the webpage at the url

In [3]:
```
url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cl
```

In [4]:
```
data = requests.get(url).text
```

Create a soup object

In [5]:
```
soup = BeautifulSoup(data,"html.parser")
```

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

In [9]:
```
df = pd.DataFrame(columns=['Language Name', 'Average Salary'])

for row in soup.find("tbody").find_all("tr"):
    col =row.find_all("td")
    language = col[1].text
    salary = col[3].text

    df = df.append({'Language Name' : language, 'Average Salary' : sal

print (df)
```

```
   Language Name         Average Salary
0       Language  Average Annual 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 [10]:
```
# your code goes here
df.to_csv('popular-languages.csv')
```

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