

## Question 2:

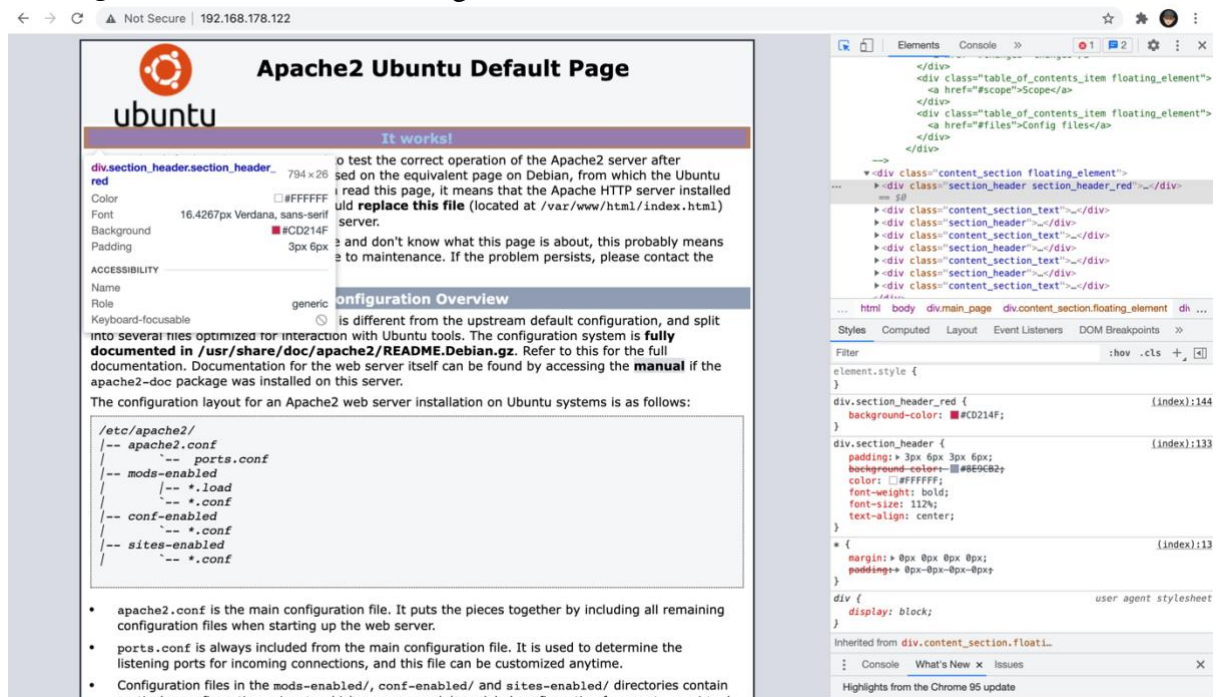
Code: Using Python on your host (windows) pc scrape the Apache 2 page you just created (or LYIT web page) and parse it minimally for later processing. For example:

2.1. What are the headings?

2.2. How many times does the word Apache2 appear ? Think of searching for an event id or user id that is in your system.

2.3. Any other item

### 1. Looking for a class name of headings.



### 2. Python Code :

```
17
18
19 def read_page_contents():
20     """ scrape web page contents """
21
22     html_doc = requests.get("http://192.168.178.122")
23     soup = BeautifulSoup(html_doc.text, 'html.parser')
24
25     """searching all headings by class : section_headers"""
26     for headers in soup.find_all(class_="section_header"):
27         headings.append(headers.text.strip())
28
29     print('This are all headings name: {}'.format(headings))
30
31     """count word Apache2"""
32     pattern = 'Apache2'
33     word = soup.find_all(text=lambda text: text and pattern in text)
34     words_count = len(word)
35
36     print("Word Apache2 appears {} times".format(words_count))
37
38     """ find what the title is"""
39     title = soup.title
40
41     print("The title of website is: {}".format(title.string))
42
43
44 if __name__ == "__main__":
45     read_page_contents()
46
```

### 3. Code Execution:

```
"/Users/luca/Library/Mobile Documents/com~apple~CloudDocs/LVIT - School/Assignments/bin/python" /Users/luca/PycharmProjects/Assignment/Question_2/q_2.py
This are all headings name: ['It works!', 'Configuration Overview', 'Document Roots', 'Reporting Problems']
Word Apache2 appears 6 times
The title of website is: Apache2 Ubuntu Default Page: It works

Process finished with exit code 0
|
```