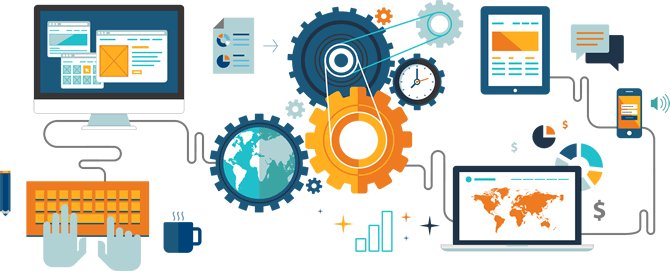
**Big Schedules Automation (Ocean Port Lanes Traffic & Schedules)**



<https://www.bigschedules.com> is a website source, which leverages big data sources, and displays live vessel information about current vessel locations in the search results. It rapidly check multiple sources of published sailing schedules at once to provide current schedule data and it makes search criteria recommendation based on past history.

Scrapping the schedules extraction from the above mentioned website source as per required parameters entered by the user in input file called **“Input.xlsx”**

**Tools & Technologies for Solution:** https://static.thenounproject.com/png/1390657-200.png

Following tools and technologies used to complete this project:

1. Anaconda Environment (Python 3.6 or Above) [[1]](#footnote-1)
2. Excel
3. Batch Script

Required Libraries:

1. Selenium (Python Wrapper) [[2]](#footnote-2) to scrap data
2. Pandas[[3]](#footnote-3) to manipulate excel data
3. Logging[[4]](#footnote-4) to log all the events

**Requirements:**

Data required following carriers:

1. Cosco
2. APL
3. Hyundai
4. Matrix
5. Maersk

**Note:** This requirement is flexible. We can use any carrier that used by website

**Setup & Configurations:**

Following steps are required to setup & configure environment to execute the script:

**Preferable:**

Download Anaconda from <https://www.anaconda.com/download/>

Version: Python 3.6 or Above

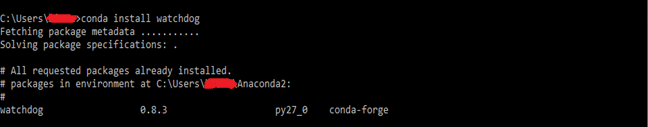
* After installation of Anaconda, few libraries needed first. Although Anaconda automatically install and remove dependencies for all the libraries by itself. Few external required libraries are as follow:

1. Selenium (Web Driver)
2. Pandas
3. Logging

To install a library:

* Open anaconda command prompt
* Type ‘conda install ‘library name’

See the example in image below



**Requirements (Where the script will reside):**

Required machine requirements are as follow:

* The Python script (Big.py), bat script (RunBigScheduleScript), an input file (Data.xlsx) and a folder named ‘Data’ must be placed in one directory
* Always use the naming convention for carrier names and for origin and destination names. (For the ease, we are placing a file in which adding origins, destinations and carriers names that websites use. So to add these parameter to input file, just copy paste names from that file)

**How to Execute Script:**

1. Double click the bat file named ‘RunBigScheduleScript.bat’

**Output:**

After executing the script, you will have the schedules of given origins-destinations in separate excel files placed in folder named **“Data”** in the same directory.

1. <https://www.anaconda.com/download/> [↑](#footnote-ref-1)
2. <http://selenium-python.readthedocs.io/> [↑](#footnote-ref-2)
3. <http://pandas.pydata.org/pandas-docs/version/0.15/tutorials.html> [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)