Automation for data sourcing of daily S&P 500 stocks prices (Open, High, Low, Close and Volume) on AWS EC2 using Cron Jobs

Written by: Ankush Garg

In this article I will be covering how automation can be done to source daily S&P500 stock prices. I am doing automation on AWS EC2 instance sing Cron Jobs.

Step 1: Setup EC2 instance. Refer to my articles on link: Medium, LinkedIn, GitHub

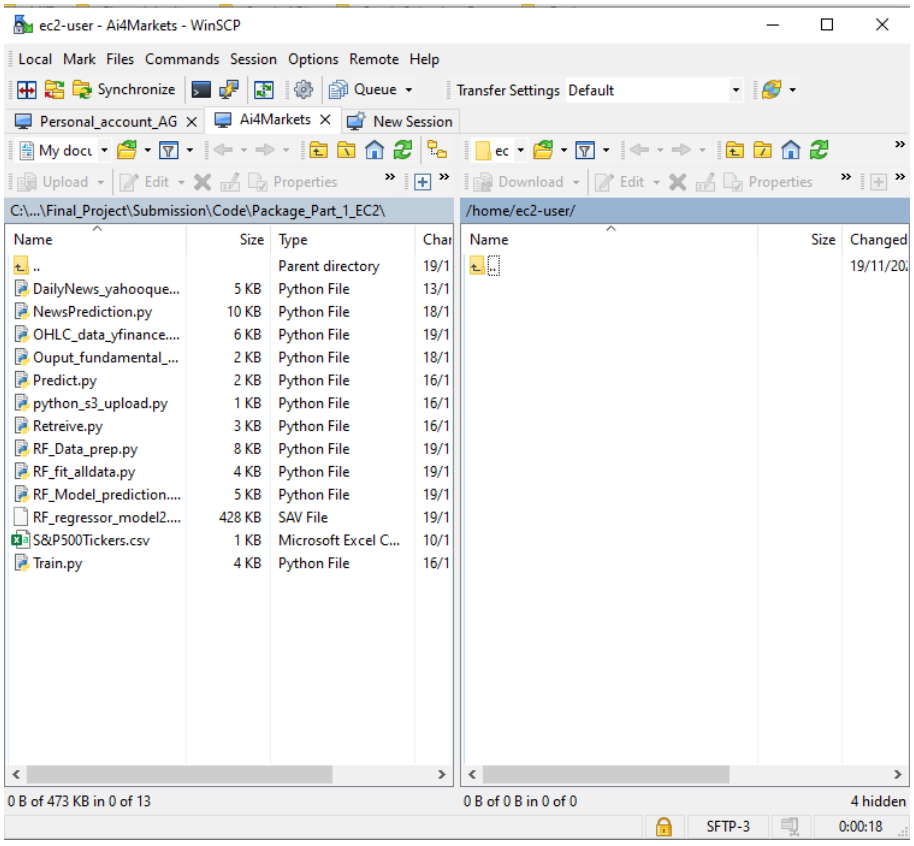
Step 2: Create python script to source daily stock prices (Open, High, Low, Close and Volume). Refer to my articles link: Medium, LinkedIn, Github

Step 3: Pull S&P500 list of stocks. Refer to article link: Medium, LinkedIn, Github

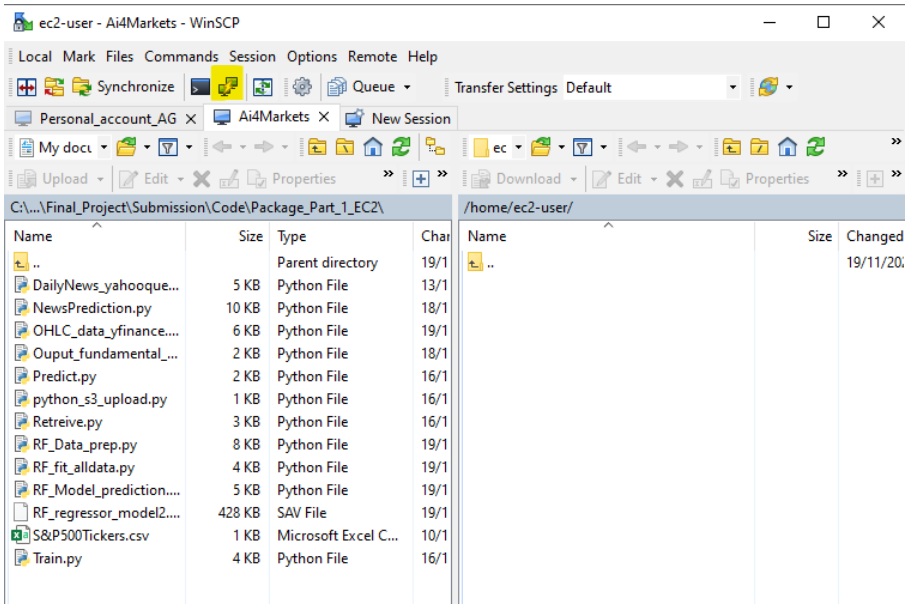
Step 4: Windows: Copy files from your local machine to EC2 Instance. Refer to my article on: Medium, LinkedIn, Github,

Mac/ Linux link: Medium, LinkedIn, Github

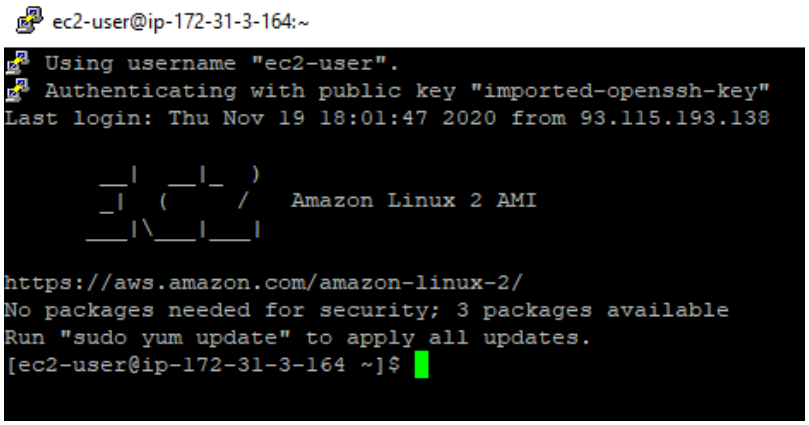
Step 5: Copy to files: a) S&P500Tickers.csv – file containing list of stocks in S&P500, b) marketdata_yfinance.py – python script to source the daily stock prices



Step 6: By default new EC2 instance comes with python package 2.7. We have to upgrad python to 3.7. Open Putty: Login to EC2 using WinSCP or Putty. Click on the putty icon highlighted in yellow in below screenshot

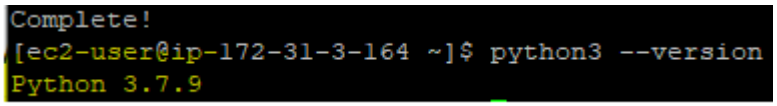


You will get the screen like below:



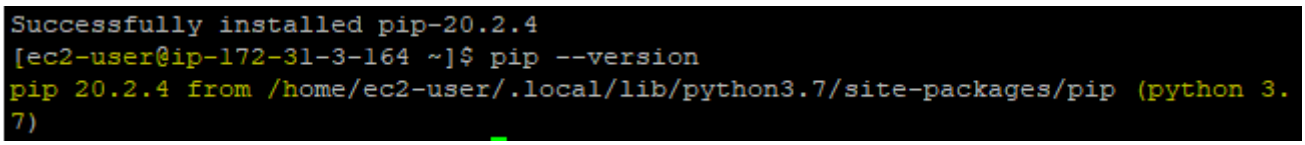
Step 7: Run the command: sudo yum install python37. You will be prompted to enter y/n. Enter y

Python should get upgraded to 3.7. Check it using python3 –version



Step 8: Next must install pip. Run the command: python3 -m pip install --user --upgrade pip. This should successfully install pip.

Step 9: Check the version of pip installed. Pip –version. Pip must get upgraded to version 20.2 on python 3.7



Step 10: Install yfinance. Command: pip install yfinance

Step 11: run command ‘ls’ and ensure both Ensure both : ‘S&P500Tickers.csv’ and ‘marketdata_yfinance.py’ are present. If not, please copy them. Details on creating these files are present in step 2 and 3 in above article

Step 12: Run the command ‘python3 marketdata_yfinance.py’. Check if script has run successfully. If yes, then file 'OHLC_yfinance_data.csv' is created. In case some errors are thrown then it could be because of some python packages not present in your EC2 instance. Install the packages required using ‘pip install command’

Step 13: Set-up cronjob for auto-update. I will be setting up job to run this script daily at 4:00 am. Run the two commands mentioned below to start the cronjob on Linux 2 EC2 instance:

sudo systemctl start crond

sudo systemctl enable crond

Step 14: Edit the cronfile to start the daily automated runs of the scripts that we ran manually above. Run the command: crontab -e

Step 15: Press 'i' to edit

Step 16: Enter: 0 4 * * * python3 /home/ec2-user/marketdata_yfinance.py

This will run the crontab daily at 4:00 am machine time. If you want to change the time and frequency of run of this file then that has to be done using different values to * in corn. There are I total 5 stars, I have put 0 for the first star and 4 for the 2nd star

5 stars are explained below:

1st star: Minute (ranges from 0-59)
2nd star: Hour (ranges from 0-23)
3rd star: Day (ranges from 1-31)
4th star: Month(ranges from 1-12)
5th star: Day-of-week (0-7. 0 & 7 is Sun. 1-Mon, 2-Tue...etc)

CONGRATULATIONS!!! You have successfully completed automation of daily run of market data sourcing for stocks in S&P500 list.