Stock Prediction System Installation Instruction

The Stock Prediction System runs on spark platform. To implement a spark environment, we will install Docker and run spark on Docker's containers.

1. Install Docker

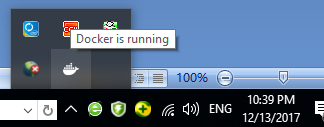
Docker can be downloaded from its official website for free:

https://www.docker.com/docker-windows

This documentation shows how to install Docker on Windows, please choose the proper version based on your operation system.

Once the download is done, run "Docker for Windows Installer.exe" to install Docker to Windows.

When Docker is running WIndows, you will find the Icon on the right buttom.



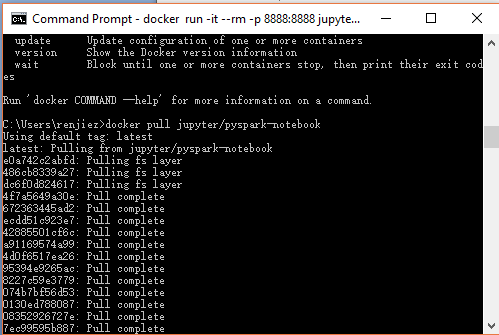
2. Install Docker container

On this step, we are going to install the necessary components into Docker's container. In order to run Stock Prediction System, we need to have Python and Spark. We are going to download Pyspark image which combines Python, Jupyter and Spark all toghter. You may find a detailed instruction here:

http://maxmelnick.com/2016/06/04/spark-docker.html

Open the Command Console or Windows Powershell, type the following command to download the image:

docker pull jupyter/pyspark-notebook



3. Run the Docker container

When the image downloading is fully complete, run this command to setup the Python and Spark environment:

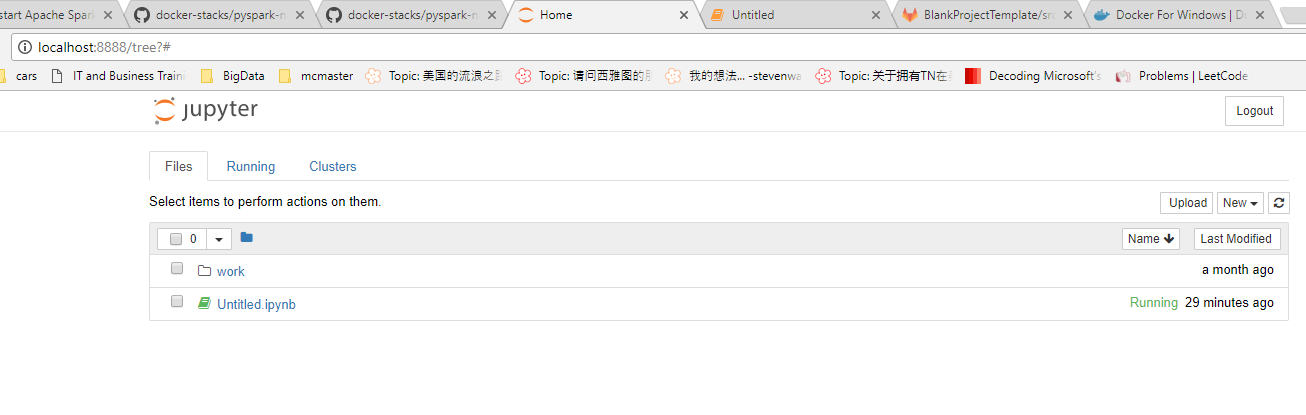
docker run -it --rm -p 8888:8888 jupyter/pyspark-notebook

Note: 8888 is the portal number, you can change it to whatever you like.

Docker set http://localhost:8888 to Jupyter Editor.

4. Run a test code

Open your web Browser and type http://localhost:8888 in the address bar, you will have the Jupyter editor page:



Click new on the top right to create a new Jupyter notebook.

Copy this part of code and paste to the editor:

import pyspark

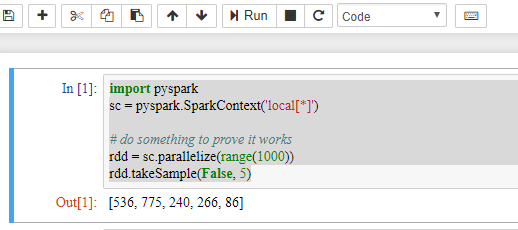
sc = pyspark.SparkContext('local[\*]')

# do something to prove it works

rdd = sc.parallelize(range(1000))

rdd.takeSample(False, 5)

Run this part of code to test the Spark.

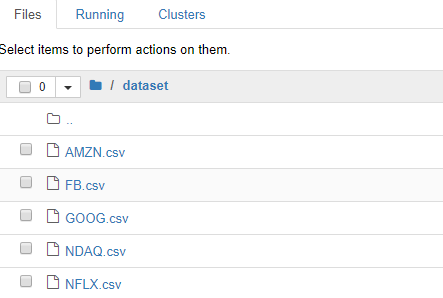


Once the output was displayed, you have completed the installation.

5. Run Stock Predict System

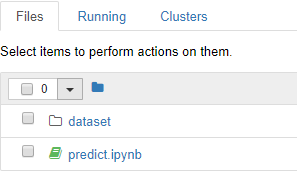
Create a new folder named dataset in the root

upload the .csv dataset files to the folder dataset.



Create a new notebook and rename it predict.ipynb.

Copy the source code from predict.py to the notebook.



Run the notebook predict.ipynb