

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

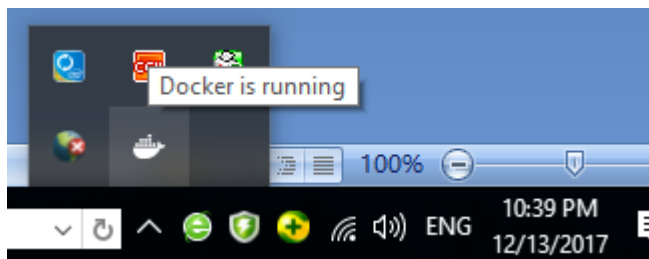
Docker can be downloaded from its official website for free:

<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 button.



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 together. 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
```

```
Command Prompt - docker run -it --rm -p 8888:8888 jupyter...
update      Update configuration of one or more containers
version     Show the Docker version information
wait        Block until one or more containers stop, then print their exit codes
es

Run 'docker COMMAND --help' for more information on a command.

C:\Users\renjiez>docker pull jupyter/pyspark-notebook
Using default tag: latest
latest: Pulling from jupyter/pyspark-notebook
e0a742c2abfd: Pulling fs layer
486cb8339a27: Pulling fs layer
dc6f0d824617: Pulling fs layer
4f7a5649a30e: Pull complete
672363445ad2: Pull complete
ecdd51c923e7: Pull complete
42885501cf6c: Pull complete
a91169574a99: Pull complete
4d0f6517ea26: Pull complete
95394e9265ac: Pull complete
8227c59e3779: Pull complete
074b7bf56d53: Pull complete
0130ed788087: Pull complete
08352926727e: Pull complete
7ec99595b887: Pull complete
```

3. Run the Docker container

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

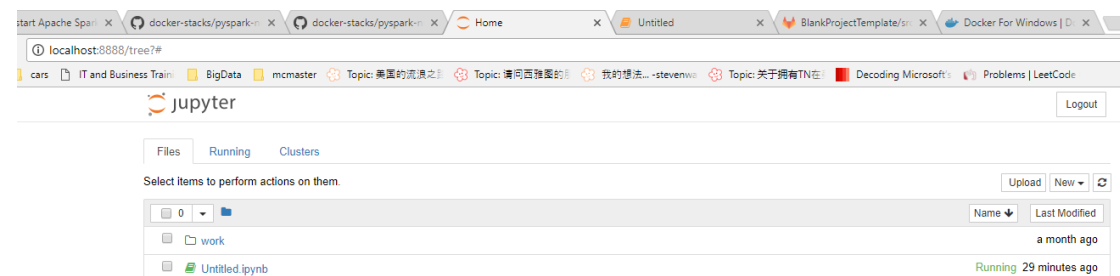
```
docker run -it --rm -p 8888:8888 jupyter/pyspark-notebook
```

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.

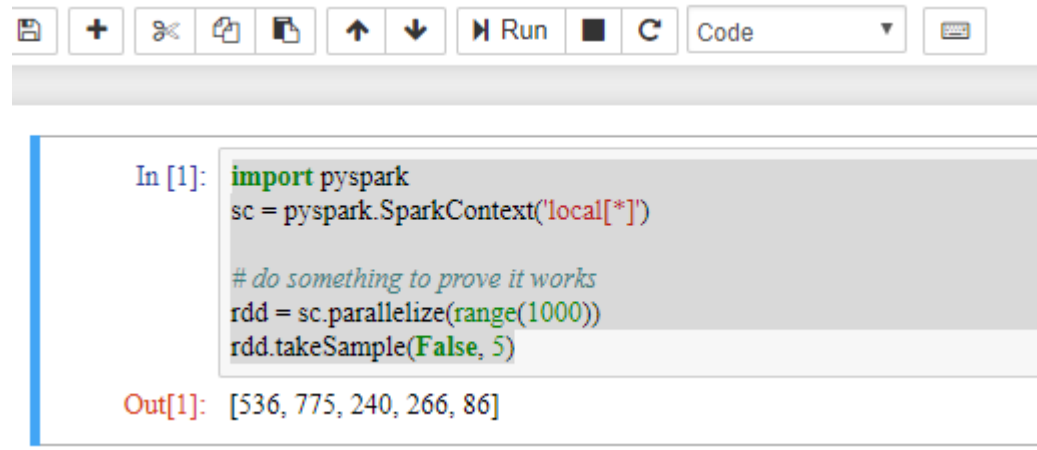
Copy this part of code and paste to the editor:

```
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.



The screenshot shows a Jupyter Notebook interface. At the top, there is a toolbar with icons for saving, adding, deleting, copying, pasting, undo, redo, and running code. Below the toolbar is a code cell with the following Python code:

```
In [1]: import pyspark
        sc = pyspark.SparkContext('local[*]')

        # do something to prove it works
        rdd = sc.parallelize(range(1000))
        rdd.takeSample(False, 5)
```

Below the code cell, the output is displayed:

```
Out[1]: [536, 775, 240, 266, 86]
```

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

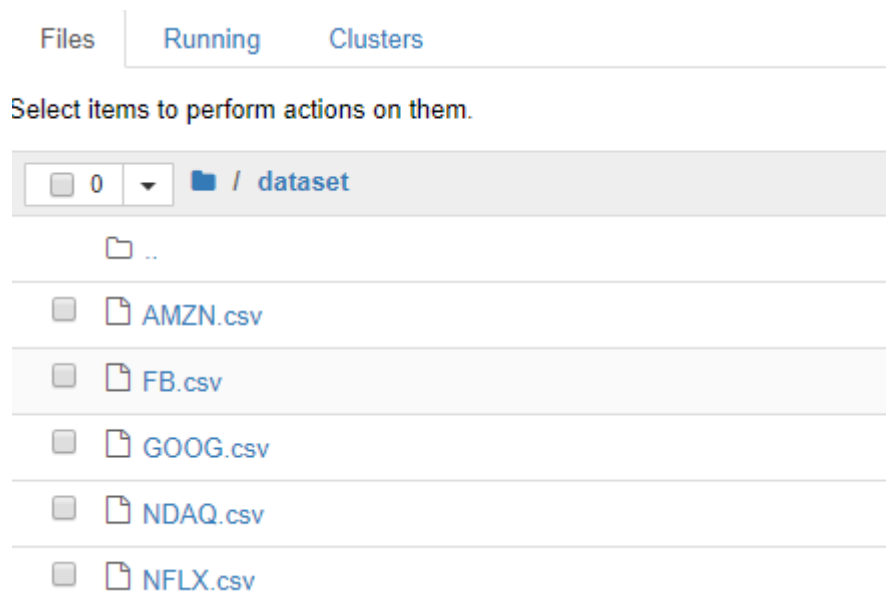
5. Run Stock Predict System

Create a new notebook and rename it predict.ipynb.

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

Create a new folder named dataset in the root

upload the .csv dataset files to the folder dataset.



The screenshot shows a file explorer interface with three tabs: 'Files', 'Running', and 'Clusters'. The 'Files' tab is selected. Below the tabs, there is a text input field with '0' and a dropdown arrow, followed by a folder icon and the text '/ dataset'. Below this, there is a list of files:

- ..
- AMZN.csv
- FB.csv
- GOOG.csv
- NDAQ.csv
- NFLX.csv