pip install scrapy

pip install scrapyd

pip install scrapyd-redis

pip install scrapyd-client

pip install scrapyd-api

pip3 install scrapy-splash

启动服务器的scrapyd

$ scrapyd

scrapyd默认将运行在<http://127.0.0.1:6800>

scrapyd默认绑定的地址是127.0.0.0

若想要从外网访问则更改scrapyd的配置文件，我的linux版本是CentOS，因此scrapyd配置文件在此目录下：/usr/local/lib/python3.7/site-packages/scrapyd/default\_scrapyd.conf

用vi进行编辑:

$ vi /usr/local/lib/python3.7/site-packages/scrapyd/default\_scrapyd.conf



修改配置文件后，再重新启动scrapyd，即可在本地浏览器打开<http://domian:6800>,进入scrapyd的图形管理页面。

部署爬虫：

在完成安装scrapyd-client之后，这里多讲一些，若是在安装scrapyd-client之后，在控制台输入scrapyd-deploy提示没有此命令时。先卸载scrapyd-client $ pip uninstall scrapyd-client,然后去scrapyd-client的GitHub上<https://github.com/scrapy/scrapyd-client>下载源码重回安装，进入scrapyd-client目录，输入python setup.py install进行安装即可，现在就可以正常使用scrapyd-deploy命令来部署scrapy项目了。

首先进到scrapy项目目录，在控制台输入scrapy list如果正常列出你编写的所有爬虫名字，说明工作正常。

然后开始打包scrapy项目，修改配置文件scrapy.cfg，设置部署名称，以及服务器的地址，以及项目名称。例如：

[deploy:example\_deploy]  
url = http://127.0.0.1:6800/  
project = dannywu

然后将scrapy部署到服务器, scrapyd-deploy 部署的名称 -p 项目名称， 例如：

$ scrapyd-deploy example\_deploy -p dannywu

现在你打开服务器的scrapydweb管理界面，即可看到页面上显示有Available projects: dannywu

接下来通过scrapy-api的curl来监测爬虫：

启动爬虫：curl http://domain:6800/schedule.json -d project=项目名称 -d spider=爬虫名称例如：

$ curl http://127.0.0.1:6800/schedule.json -d project=dannywu -d spider=get\_dannywu\_spider

停止爬虫：curl http://domain:6800/cancel.json -d project=项目名称 -d job=运行ID，例如：

$ curl http://127.0.0.1:6800/cancel.json -d project=dannywu -d job=2376

删除scrapy项目：注意：一般删除scrapy项目，需要先执行命令停止项目下在运行的爬虫

curl http://domain:6800/delproject.json-d project=项目名称，例如：

curl http://127.0.0.1:6800/delproject.json-d project=dannywu

查询api中一共有多少个scrapy项目：

curl <http://localhost:6800/listprojects.json>，例如：

curl http://127.0.0.1:6800/listprojects.json

查看指定项目里面的爬虫数量：

curl http://localhost:6800/listspiders.json?project=项目名称，例如：

curl http://localhost:6800/listspiders.json?project=dannywu