

# 使用Python爬虫定时爬取天气数据并保存

## 一、实现效果



	id	ct_id	county_id	county_name	weather_code	weather_info	msg_info	gmt_create	gmt_modify	rsver_1
	主键	省会ID	城市id	城市名字	城市天气编号 (获取天气专用)	天气情况	城市情况	创建时间	数据修改时间	预留字段
<input type="checkbox"/>	1	0101	010101	北京	101010100	温度:2℃ 风力:3级 湿度:19% 风向:北风 PM2.5:17 空气质量:良 温度范围:低温...	2018/7/29 第一次添加county	2018-07-29 21:53:50	NULL	NULL
<input type="checkbox"/>	2	0101	010102	海淀	101010200	温度:0℃ 风力:2级 湿度:37% 风向:东北风 PM2.5:Null 空气质量:Null 温度范围...	2018/7/29 第一次添加county	2018-07-29 21:53:50	NULL	NULL
<input type="checkbox"/>	3	0101	010103	朝阳	101010300	温度:2℃ 风力:3级 湿度:23% 风向:北风 PM2.5:Null 空气质量:Null 温度范围...	2018/7/29 第一次添加county	2018-07-29 21:53:50	NULL	NULL
<input type="checkbox"/>	4	0101	010104	顺义	101010400	温度:2℃ 风力:3级 湿度:24% 风向:东北风 PM2.5:Null 空气质量:Null 温度范围...	2018/7/29 第一次添加county	2018-07-29 21:53:50	NULL	NULL
<input type="checkbox"/>	5	0101	010105	怀柔	101010500	温度:1℃ 风力:3级 湿度:19% 风向:东风 PM2.5:Null 空气质量:Null 温度范围...	2018/7/29 第一次添加county	2018-07-29 21:53:50	NULL	NULL
<input type="checkbox"/>	6	0101	010106	通州	101010600	温度:1℃ 风力:2级 湿度:26% 风向:北风	2018/7/29 第一次添加county	2018-07-29 21:53:50	NULL	NULL

图1 实现效果(数据来源:http://wthrcdn.etouch.cn)

## 二、运行前准备

由于腾讯云的服务器的默认python版本为python2.7, 在运行前, 需要打开**腾讯云终端**在服务器端升级pip并安装python相关包, 以下命令需要**依次运行**:

```
1 python -m pip install --upgrade pip
2 pip install request
3 pip install BeautifulSoup4
4 pip install PyMySQL
```

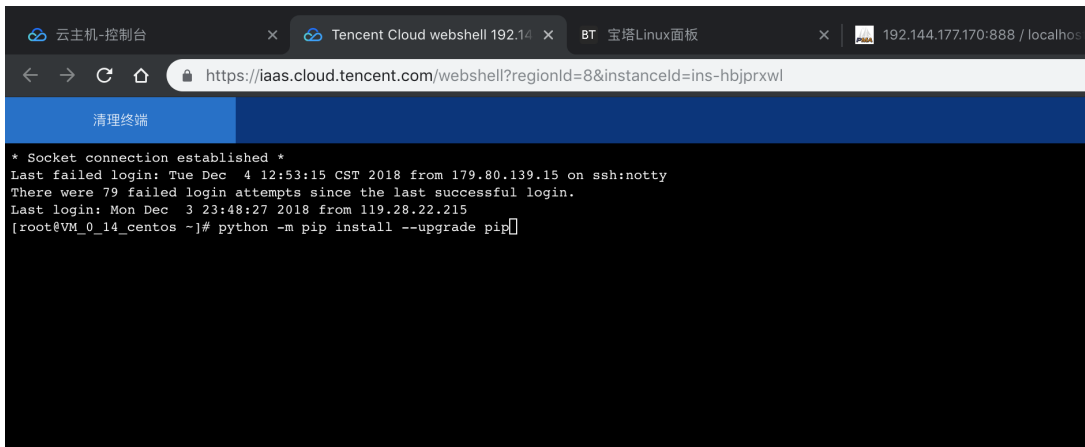


图2 腾讯云的终端界面

### 三、在Linux面板创建文件并写入

打开宝塔Linux面板，在任意地方创建一个.py文件：

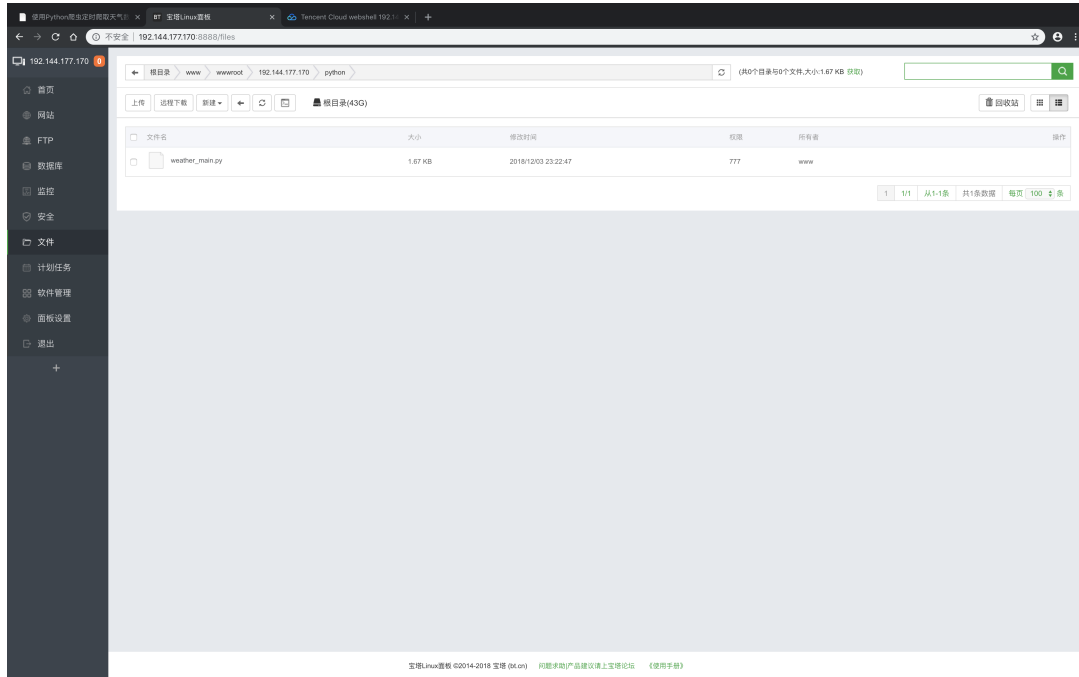


图2 创建文件

在新创建的python文件中写入：

注意在18-20行中将用户名、密码、数据库密码改为自己的

```
1 # coding=utf-8
2 """
3 2018/12/03
4
5 1801220002_陈国强
6 """
7
8 import sys
9 import requests
10 import pymysql
11 from bs4 import BeautifulSoup as bs
12
13 reload(sys)
14 sys.setdefaultencoding('utf-8')
15 # 创建连接
16 conn = pymysql.connect(host='localhost',
17                        port=3306,
18                        user='sql192_144_177_',
19                        passwd='mPreKkFEDM',
20                        db='sql192_144_177_',
21                        charset='utf8')
22 cursor = conn.cursor(cursor=pymysql.cursors.DictCursor)
23 cursor.execute('select weather_code from ins_county')
24 result = cursor.fetchall()
25 conn.commit()
26 for i in range(len(result)):
```

```

27     citycode = result[i]['weather_code']
28     url = "http://wthrcdn.etouch.cn/WeatherApi?citykey=" + citycode
29     r = requests.get(url, verify=False)
30     soup = bs(r.text, 'html.parser')
31
32     wendu = soup.find('wendu').text
33     fengli = soup.find('fengli').text
34     shidu = soup.find('shidu').text
35     fengxiang = soup.find('fengxiang').text
36     # 部分地区没有pm2.5和空气质量这个数据
37     # 极少数地区没有高低温数据
38     if soup.find('pm25'):
39         pm25 = soup.find('pm25').text
40     else:
41         pm25 = 'Null'
42     if soup.find('quality'):
43         quality = soup.find('quality').text
44     else:
45         quality = 'Null'
46     if soup.find('high_1'):
47         high = soup.find('high').text
48     else:
49         high = 'Null'
50     if soup.find('low_1'):
51         low = soup.find('low').text
52     else:
53         low = 'Null'
54     weather_content = '温度:' + wendu + '°C' + '\n风力:' + fengli + \
55                     '\n适度:' + shidu + '\n风向:' + fengxiang + \
56                     '\nPM2.5:' + pm25 + '\n空气质量:' + quality + \
57                     '\n温度范围:' + low + '-' + high + '°C'
58     sql_exe = 'UPDATE ins_county SET weather_info = \'' +
59             + weather_content + '\' WHERE weather_code = ' + citycode +
60     ' '
61
62     print(sql_exe)
63     cursor.execute(sql_exe)
64     conn.commit()
65     cursor.close()
66     conn.close()

```

## 四、指定定时执行

在宝塔面板中点击“计划任务”，设置为每天执行一次，并在脚本内容中填入：

```
1 cd /www/wwwroot/192.144.177.170/python;python weather_main.py
```

脚本的意思是：cd打开python文件所在的绝对路径，然后使用python命令执行.py文件

注意将cd打开的绝对路径改为自己刚才创建的.py文件的路径

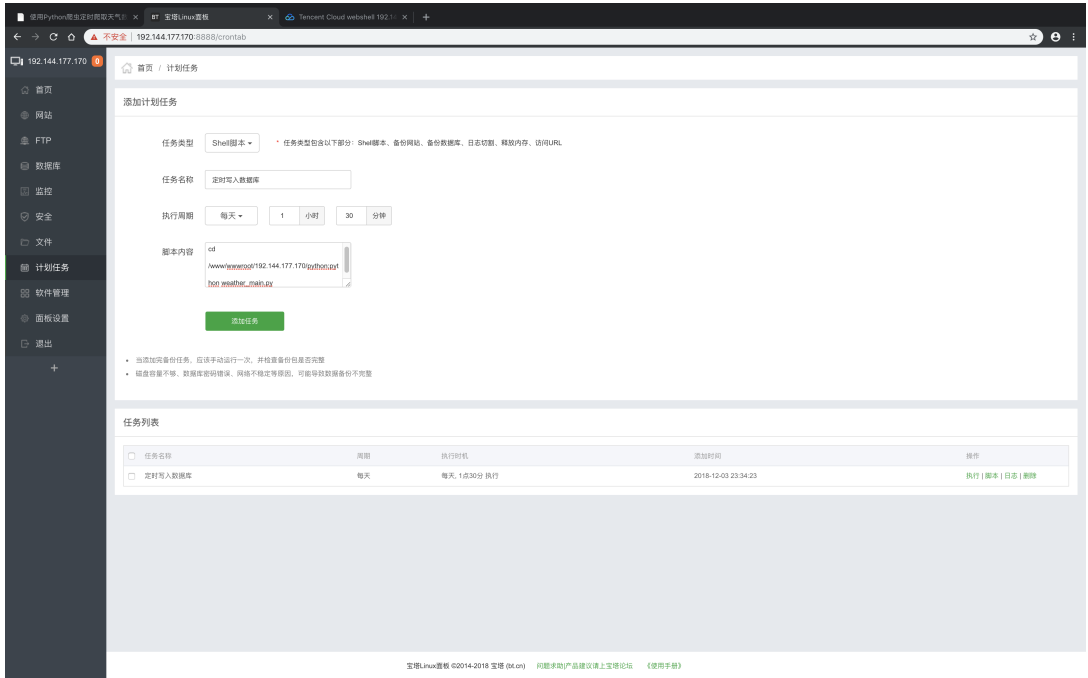


图3 创建shell脚本

可以点击下方执行按钮，但此时由于负载过大，宝塔面板可能直到程序运行完成都不会响应。