

#scrapy部署工具

```
pip install scrapyd
```

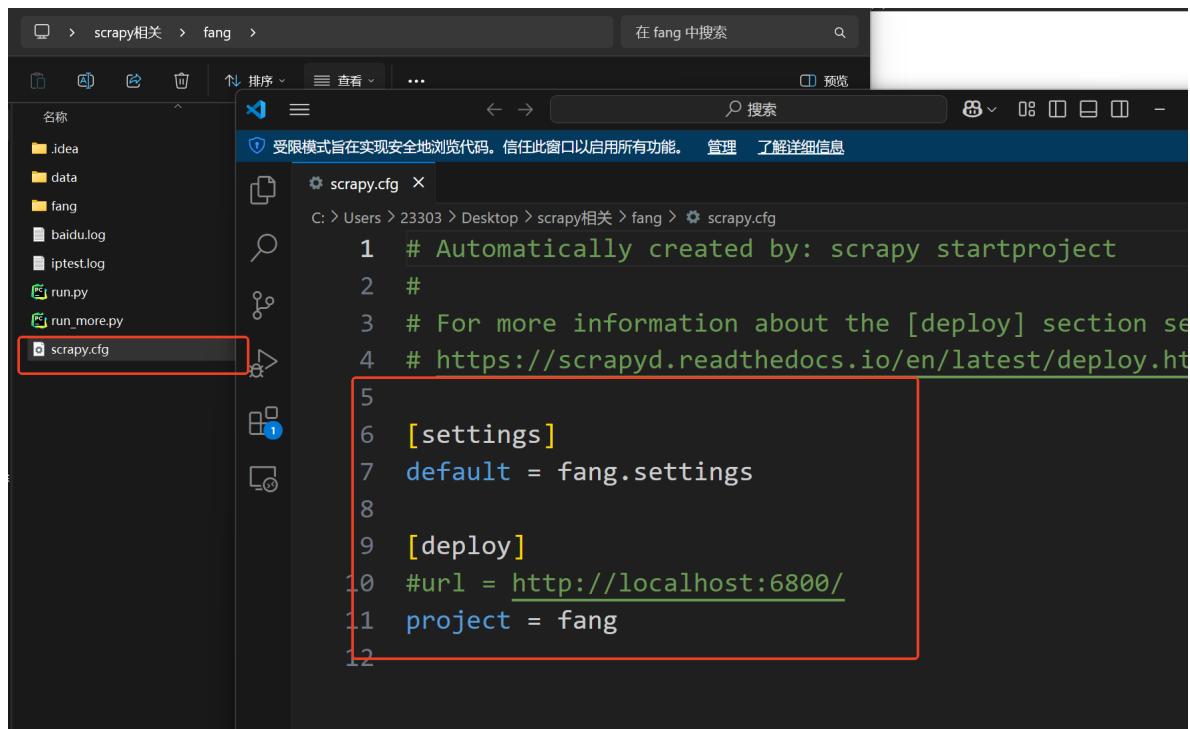
#这个是把scrapyd接口封装了，方便使用

```
pip install python-scrapyd-api
```

#这个用来打包程序方便scrapyd部署

```
pip install scrapyd-client
```

这里我们先改一下部署配置文件



scrapy.cfg

C: > Users > 23303 > Desktop > scrapy相关 > fang > scrapy.cfg

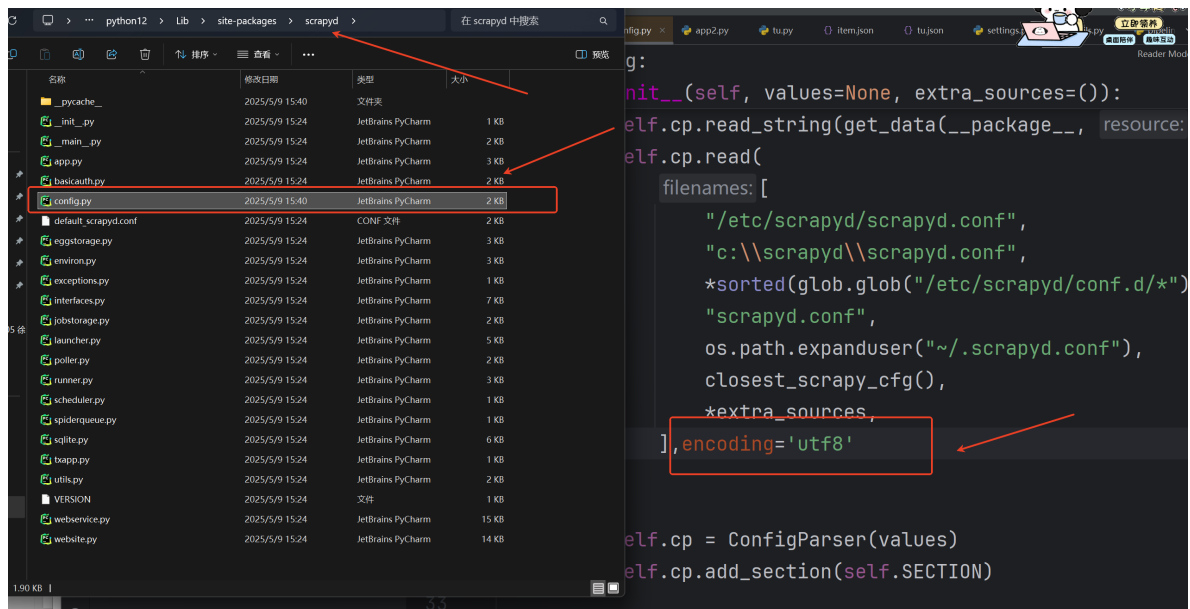
```
5
6 [settings]
7 default = fang.settings
8 #demo1为部署名，可以随意取
9 [deploy:demo1]
10 #如果是服务器就写服务器IP地址
11 url = http://localhost:6800/
12 # 这为项目名，用默认值
13 project = fang
14
```

```
[settings]
default = fang.settings
#demo1为部署名，可以随意取
[deploy:demo1]
#如果是服务器就写服务器IP地址
url = http://localhost:6800/
# 这为项目名，用默认值
project = fang
```

这里先启动scrapyd (这是个服务，不能关闭，关了就不能用了)

报错 (你不一定报错，不报错可跳过这部分)

```
PS C:\Users\23303\Desktop\scrapy相关\fang> scrapyd
Unhandled Error
Traceback (most recent call last):
  File "D:\python12\Lib\site-packages\twisted\application\app.py", line 446, in getApplication
    application = service.loadApplication(filename, style, passphrase)
  File "D:\python12\Lib\site-packages\twisted\application\service.py", line 404, in loadApplication
    application = sob.loadValueFromFile(filename, "application")
  File "D:\python12\Lib\site-packages\twisted\persisted\sob.py", line 175, in loadValueFromFile
    eval(codeObj, d, d)
  File "D:\python12\Lib\site-packages\scrapyd\txapp.py", line 4, in <module>
    application = get_application()
  File "D:\python12\Lib\site-packages\scrapyd\__init__.py", line 14, in get_application
    config = Config()
  File "D:\python12\Lib\site-packages\scrapyd\config.py", line 19, in __init__
    self.cp.read(
  File "D:\python12\Lib\configparser.py", line 684, in read
    self._read(fp, filename)
  File "D:\python12\Lib\configparser.py", line 999, in _read
    for lineno, line in enumerate(fp, start=1):
builtins.UnicodeDecodeError: 'gbk' codec can't decode byte 0xaf in position 266: illegal multibyte sequence
```



找到图里的位置加一个`encoding='utf8'`即可

再次启动成功

```
C:\Users\23303>scrapy
2025-05-09T15:40:50+0800 [-] Loading D:\python12\Lib\site-packages\scrapy\txapp.py...
2025-05-09T15:40:50+0800 [scrapy.basicauth#info] Basic authentication disabled as either 'username' or 'password' is un
set
2025-05-09T15:40:50+0800 [scrapy.app#info] Scrapy web console available at http://127.0.0.1:6800/
2025-05-09T15:40:50+0800 [-] Loaded.
2025-05-09T15:40:50+0800 [twisted.application.app.AppLogger#info] twisted 24.11.0 (D:\python12\python.exe 3.12.5) startin
g up.
2025-05-09T15:40:50+0800 [twisted.application.app.AppLogger#info] reactor class: twisted.internet.selectreactor.SelectRe
actor.
2025-05-09T15:40:50+0800 [-] Site starting on 6800
2025-05-09T15:40:50+0800 [twisted.web.server.Site#info] Starting factory <twisted.web.server.Site object at 0x0000024F7E
A7C530>
2025-05-09T15:40:50+0800 [Launcher] Scrapy 1.5.0 started: max_proc=64, runner='scrapy.runner'
2025-05-09T15:41:08+0800 [twisted.python.log#info] "127.0.0.1" -- [09/May/2025:07:41:07 +0000] "GET / HTTP/1.1" 200 950
 "-" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/136.0.0.0 Safari/537.36 Ed
g/136.0.0.0"
2025-05-09T15:41:08+0800 [twisted.python.log#info] "127.0.0.1" -- [09/May/2025:07:41:07 +0000] "GET /favicon.ico HTTP/1
.1" 404 167 "http://127.0.0.1:6800/" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) C
hrome/136.0.0.0 Safari/537.36 Edg/136.0.0.0"
2025-05-09T15:42:08+0800 [twisted.web.http.HTTPChannel#info] Timing out client: IPv4Address(type='TCP', host='127.0.0.1'
, port=54029)
2025-05-09T15:42:08+0800 [twisted.web.http.HTTPChannel#info] Timing out client: IPv4Address(type='TCP', host='127.0.0.1'
, port=54028)
```

本地浏览器网文可以看到



## Scrapy

- [Jobs](#)
- [Logs](#)
- [Documentation](#)

No Scrapy projects yet.

This web UI is for monitoring only. To upload projects and schedule crawls, use the API. For example, using [curl](#):

```
curl http://localhost:6800/schedule.json -d project=default -d spider=somespider
```

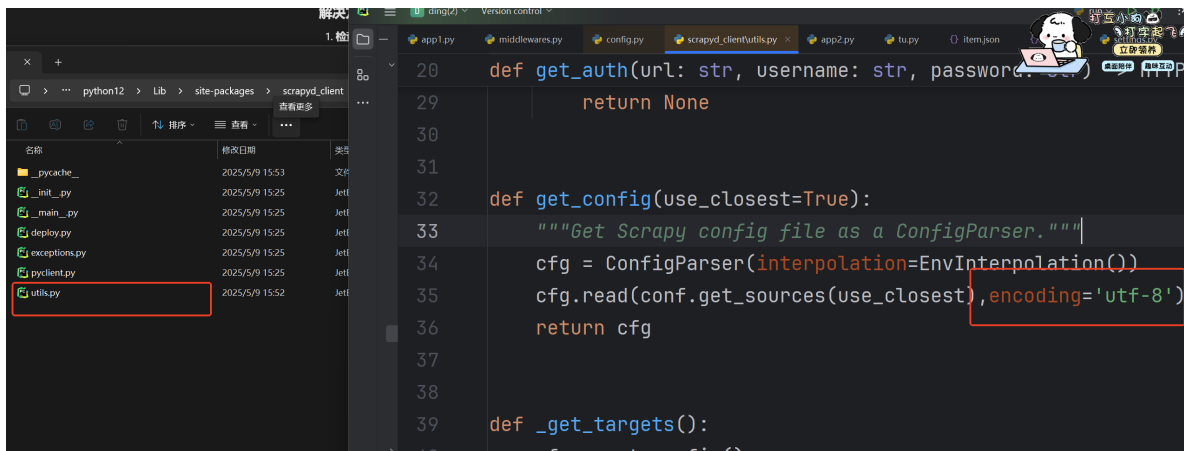
See the [Scrapy documentation](#) for details.

然后试验一下部署代码

```
scrapyd-deploy
```

又炸了

```
PS C:\Users\23303\Desktop\scrapy相关\fang> scrapyd-deploy
Traceback (most recent call last):
  File "<frozen runpy>", line 198, in _run_module_as_main
  File "<frozen runpy>", line 88, in _run_code
  File "D:\python12\Scripts\scrapyd-deploy.exe\__main__.py", line 7, in <module>
  File "D:\python12\Lib\site-packages\scrapyd_client\deploy.py", line 77, in main
    target = _get_targets()[opts.target]
             ^^^^^^^^^^^^^^^^^
  File "D:\python12\Lib\site-packages\scrapyd_client\utils.py", line 40, in _get_targets
    cfg = get_config()
          ^^^^^^^^^^^
  File "D:\python12\Lib\site-packages\scrapyd_client\utils.py", line 35, in get_config
    cfg.read(conf.get_sources(use_closest))
  File "D:\python12\Lib\configparser.py", line 684, in read
    self._read(fp, filename)
  File "D:\python12\Lib\configparser.py", line 999, in _read
    for lineno, line in enumerate(fp, start=1):
                       ^^^^^^^^^^^^^^^^^^^^^
UnicodeDecodeError: 'gbk' codec can't decode byte 0xaf in position 266: illegal multibyte sequence
PS C:\Users\23303\Desktop\scrapy相关\fang> |
```



继续改encoding='utf8'

个人觉得可能是scrapy的配置文件里加了中文注释的原因导致的，或许不写中文注释不会出这些问题

```
scrapyd-deploy demo1
```

```
Unknown target: default
PS C:\Users\23303\Desktop\scrapy相关\fang> scrapyd-deploy demo1
Packing version 1746777439
Deploying to project "fang" in http://localhost:6800/addversion.json
Server response (200):
{"project": "fang", "version": "1746777439", "spiders": 6, "status": "ok", "node_name": "DESKTOP-NON0MQM"}
PS C:\Users\23303\Desktop\scrapy相关\fang> |
```

这个时候就OK了

这时候看scrapyd这边

## Scrapy

- [Jobs](#)
- [Logs](#)
- [Documentation](#)

Scrapy projects:

- fang

This web UI is for monitoring only. To upload projects and schedule crawls, use the API. For example, using [curl](#):

```
curl http://localhost:6800/schedule.json -d project=default -d spider=somespider
```

See the [Scrapy documentation](#) for details.

Gerapy是一个可视化的爬虫工具

使用gerapy之前要保证scrapy 是可用的

先安装

```
pip install gerapy
```

```
gerapy init
```

```
PS C:\Users\23303\Desktop\scrapy相关\myprpject> gerapy init
Initialized workspace gerapy
```

```
cd gerapy
gerapy migrate
```

```
PS C:\Users\23303\Desktop\scrapy相关\myprpject> cd gerapy
PS C:\Users\23303\Desktop\scrapy相关\myprpject\gerapy> gerapy migrate
Operations to perform:
  Apply all migrations: admin, auth, authtoken, contenttypes, core, django_apscheduler, sessions
Running migrations:
  Applying contenttypes.0001_initial... OK
  Applying auth.0001_initial... OK
  Applying admin.0001_initial... OK
  Applying admin.0002_logentry_remove_auto_add... OK
  Applying admin.0003_logentry_add_action_flag_choices... OK
  Applying contenttypes.0002_remove_content_type_name... OK
  Applying auth.0002_alter_permission_name_max_length... OK
  Applying auth.0003_alter_user_email_max_length... OK
  Applying auth.0004_alter_user_username_opts... OK
  Applying auth.0005_alter_user_last_login_null... OK
  Applying auth.0006_require_contenttypes_0002... OK
  Applying auth.0007_alter_validators_add_error_messages... OK
  Applying auth.0008_alter_user_username_max_length... OK
  Applying auth.0009_alter_user_last_name_max_length... OK
```

```
gerapy createsuperuser
```

根据提示创建超级用户

```

PS C:\Users\23303\Desktop\scrapy相关\myprpject\gerapy> gerapy createsuperuser
Username (leave blank to use '23303'): wang
Email address:
Password:
Password (again):
This password is too short. It must contain at least 8 characters.
This password is too common.
This password is entirely numeric.
Bypass password validation and create user anyway? [y/N]: y
Superuser created successfully.
PS C:\Users\23303\Desktop\scrapy相关\myprpject\gerapy> |

```

#启动项目，IP和端口号可以改  
gerapy runserver 127.0.0.1:8008

```

PS C:\Users\23303\Desktop\scrapy相关\myprpject\gerapy> gerapy runserver 127.0.0.1:8008
Watching for file changes with StatReloader
Performing system checks...

INFO - 2025-05-09 16:09:49,912 - process: 57960 - scheduler.py - gerapy.server.core.scheduler - 105 - s
sfully synced task with jobs with force
System check identified no issues (0 silenced).
May 09, 2025 - 16:09:49
Django version 2.2.28, using settings 'gerapy.server.server.settings'
Starting development server at http://127.0.0.1:8008/
Quit the server with CTRL-BREAK.

```

