

1.在 xshell 输入命令 mongod 将服务启动（端口号 27017）
2.再新开个终端，输入 sudo mongo 进入 shell 环境，（输入 show dbs 即可查看所有数据库，use XXX 即可进入 XXX 数据库，show collections 查看所有表，db.YYY.find()查看 YYY 表里的数据，db.dropDatabase()删除当前数据库）

3.打开 settings.py 写入以下信息：

```
#MONGODB 主机名
MONGODB_HOST = "127.0.0.1"
#MONGODB 端口号
MONGODB_PORT = 27017
#数据库名称
MONGODB_DBNAME = "Douban"
#存放数据的表名称
MONGODB_SHEETNAME = "doubanmovies"
```

4.在爬虫项目的 pipelines.py 文件 import pymongo 还有 from scrapy.conf import settings
在 class XXXPipeline 的初始化方法中

```
host=settings["MONGODB_HOST"]
port=settings["MONGODB_PORT"]
dbname=settings["MONGODB_DBNAME"]
sheetname=settings["MONGODB_SHEETNAME"]
# 创建 MONGODB 数据库链接
client = pymongo.MongoClient(host=host, port=port)
# 指定数据库
mydb = client[dbname]
# 存放数据的数据库表名
self.sheet = mydb[sheetname]
```

```
def process_item(self, item, spider):
    data = dict(item)
    self.sheet.insert(data)
    return item
```

5.scrapy crawl XXX 运行爬虫即可

6.在 xshell 退出再重新运行 mongo, show dbs , use XXX , show collections , db.YYY.find()就出来了

7.打开本地客户端 Robomongo 也能看到

8.如果因为之前多试了几次导致数据库多插了数据,可以 db.dropDatabase()再运行爬虫项目即可